LECTURE NOTES ON DYNAMICAL SYSTEMS: FROM ADIC TRANSFORMATIONS TO GIBBS STATES (PRELIMINARY VERSION)

ALBERT M. FISHER

Abstract.

Contents

Date: August 8, 2023.

Key words and phrases. Perron-Frobenius Theorem, subshift of finite type, Shannon-Parry measure, Gibbs state.

1. Introduction

TO DO:
– invt. measures, fun'l analysis; Choquet and erg. decomp
– Kolmogorov and regularity and stationarity
– Hausdorff measure
– Kak equiv
– Osceledec

– CLT proof; log log proof; Martingale and Gordin proof; Chung-Erd proof –???

NOTE: "???" in the text is a reminder to me that this part needs completion!

1.1. **Purpose of the notes.** These very preliminary and incomplete notes have been developing and expanding gradually since 1993, most of this material having been presented in graduate courses in dynamics and ergodic theory, or in seminars, at the State University of New York, Stony Brook, Instituto de Mathemática, UFRGS Porto Alegre, the Univerity of Marseilles, and especially at the Universidade de São Paulo.

There are many excellent texts in dynamical systems, ergodic theory and probability. Some of my personal favorites are: [Bil78], [Shu87], [Bow75], [Fur81],[Shi73], [Lam66],[GS92], [Nit71], [Irw80], both because of the material covered and because they are such beautiful books.

Other invaluable texts are [Wal82], [Sin76],[Bow77], [PdM82], [PP90], [KH95], each of these being magnificent in its own way.

The last of these is a remarkable achievement in that it brings together so many topics in one place, with up-to-date results and methods of proof.

Nevertheless the field has grown so much that many interesting topics are not included there.

The list should go on to include texts in the closely related areas of fractal geometry, complex dyamics, Kleinian groups, low dimensional topology, combinatorial group theory.... After all, divisions in mathematics are entirely artificial, and if we go deep enough, it seems we find that everything is interesting and of beauty.

I think of these notes as serving three possible purposes: as a central text around which to build a dynamics course, as supplementary material to accompany a course based for instance on one or more of the books mentioned above, and as an introduction to some parts of the research literature.

The material included here is here because I needed to write it down in this way to have a basis for my own teaching of this material; other topics were already so well treated elsewhere that I haven't felt called to write my own version.

When I have taught this material, I have sometimes included entropy theory, basing my approach on some combination of [Bil78], [Wal82], and [PP90]; circle diffeomorphisms, using [CFS82] and [KH95], parts of probability theory, using [Lam66], [GS92], as well as research papers, in each case. In future versions of the course, I intend to present stable manifolds and related matters, [Shu87], [KH95], [PdM82]; Markov partitions, using [Shu87], [Bow75]; Morse-Smale diffeomorphisms, using [Irw80], [PdM82]; future dreams are to delve into: complex dynamics? Kleinian groups? low dimensional topology? combinatorial group theory? And these notes may expand along the way.

When I have taught the course, about half the time has been spent discussing concrete examples, and the blackboard was often full of diagrams and pictures. These are not (yet?) written down; most of that material forms part of the folklore of the subject. Anyone teaching this material would do well to view it a bit as a cooking project: these are some of the ingredients; you need to add and stir in a lot of your own!

Corrections, comments, additions and suggestions are welcome.

1.2. **Background.** A basic knowledge of general topology and measure theory at the level of, say, [Roy68] will be assumed. My own favorite texts in addition to Royden are: advanced calculus, "baby analysis": metric spaces etc.: [Mar74]; set theory [Hal74], measure theory [Bar66], [Roy68], [Rud87], [Oxt80], and parts of [Hal76], functional analysis [Rud73], [RS72], parts of [DS57], general topology [Kel75], algebraic topology [Arm83], complex analysis [Ahl66], [Kno45], [Lan99], [MH87].

1.3. **Why dynamics?** *Dynamical systems* can be defined as the study of group or semigroup actions on a space. The group action defines the way in which the system changes, and in the most classical setting, with actions of $\mathbb{Z}, \mathbb{N}$ or $\mathbb{R}$, we think of the action as giving the evolution of time, as observed either discretely (like frames of a movie) or continuously.

So dynamics is inherently fascinating because it can be used to study anything which is evolving or changing in time.

One can think of snowflakes drifting down out of a winter sky, of water rippling across a lake, of the eddies and vortices of a mountain stream, clouds coalescing in the heavens, the moon tracing its slow trajectory against a background of stars, a smile flickering across the face of a child...or the oscillations of prices on the stock market!

Some of the questions and examples originate in physics (indeed, the term "orbit" for $\{T^n(x) : n \in \mathbb{Z}\}$, the itinerary of the point x over all time, suggests the orbit of a planet around the sun, recalling the origins of the subject in problems of celestial mechanics, such as the question as to whether the solar system is actually stable- or whether at some future time the planets and the sun might not interact with each other in such a way that one planet gets thrown out of the system!) while other examples come from pure mathematics: from number theory, algebra, differential geometry, complex analysis. Two other main sources of examples, problems, ideas and methods are probability theory and information theory.

Dynamics is a rich subject partly because so many different points of view may be brought to bear, and because the source of examples is so varied.

INSERT: diagram with arrows-pointing both ways: prob th, ino th, number th, alg top, functional complex harmonic analysis, diff geom, algebra, ODE, PDE...

FIGURE: torus; free goup graph; Riemann surface

Not only can many different fields of mathematics prove useful within dynamics, but the dynamical point of view can help pose interesting problems, and often lead to solutions, in these other areas.

When I chose dynamics as a field, I recall that one of my motivations was that this choice would give me an excuse to study so many different parts of mathematics! Thus the study of dynamics can be both very general, and very specific, which nearly guarantees that it will always be interesting.

1.4. **Remarks on Differential Equations.** In §35.1 we develop some material on Linear Algebra which will be referred to throughout these Notes. This "Minicourse" includes an introduction to Lie Groups and Lie Algebras. This material is collected

here as basic material needed for the approach we take in the section on Ordinary Differential Equations. Then in §37 we present a minicourse on ODE.

The study of differential equations played a key role in the beginnings of the study of Dynamical Systems, but the subjects have largely gone their own ways since then, occasionally coming back together depending on the interests of authors or the needs of applications. As a result, books on Dynamical Systems are more or less divided between some that have avoided differential equations entirely, some that emphasize differential equations while ignoring "more modern" dynamical systems like ergodic theory, while others have tried to bridge that gap at least to some degree. The study of ordinary differential equations itself has divided into the study of specific examples, usually in one dimension, and the qualitative theory making use of a more topological point of view.

Our basic emphasis is much more on the ergodic theory side of things. But the differential equations world is both beautiful and fascinating, so we provide this small introduction. And as in all of mathematics, learning one part is sure to help with the others, no matter how far apart they may initially appear!

For this material, in which we are not at all expert, we have benefited especially from the accounts in [HK03], [HS74], [Lan02], [Lan01], and course notes of Marina Talet, and references on Lie groups cited in that section.

The emphasis in the minicourse is on the local theory, developed in linear spaces ($\mathbb{R}^n$ and also Banach spaces as in Lang's treatments), trying to show the key role of the exponential map, first in the linear case, from linear vector fields (elements of the Lie algebra of the matrix group) to that group, sending a line $\{tA\}$ to a linear flow. For a general vector field this flow is constructed by Picard's contraction mapping. Thus a vector field is a tangent vector to a flow, a one-parameter subgroup of the collection of diffeomorphisms. In the main part of the course we focus on the global theory, both of flows and maps, but here rather than beginning with a vector field these dynamical systems are constructed more directly, by algebraic geometric or probabilistic methods, and studied by smooth, topological and measure theoretic tools, with much of the focus on the measure theory which leads to the beautiful links with randomness discovered by the development of ergodic theory.

We also review material on linear algebra, needed throughout these Notes, and also on vector calculus, up to Stokes' Theorem, with the emphasis on getting across some of the key ideas rather than full proofs, and primarily to aid with this part on ODE.

Please Note: there are also Lecture Notes on Vector Calculus on my Webpage, with some additional material on ODEs. The material her may be moved to there eventually.

I have tried to cite references, so a lot of the material without citations is original at least in its presentation. Any orriginal material including figures should be considered copyrighted by the author and used only with attribution, please!

I owe infinitely much to wonderful teachers and mentors and colleagues and reference works. In a later version I will be more thorough with acknowledgements.

Many "unfinished" parts are indicated here; we shall see if they actually get finished up (or not!)

Any and all readers are very welcome to write in with corrections or comments.

2. Basic concepts

2.1. **Transformations, flows and group actions.**

Definition 2.1. Given a set X, a **transformation** T of X is a function $T : X \to X$. Note that for $n \geq 1$, writing $T^n = T \circ \cdots \circ T (n \text{ times})$, then $T^{n+m} = T^n \circ T^m$. We define T^0 to be the identity map $\text{id} : X \to X$, and if T is invertible, we write $T^{-n} = (T^{-1})^n$. A **flow** on X is a collection of transformations $\{\tau_t : t \in \mathbb{R}\}$, such that $\tau_0 = \text{id}$ and $\tau_{t+s} = \tau_t \circ \tau_s$. The transformation τ_t is called the **time-t map** of the flow; note that each τ_t is invertible with inverse τ_{-t}, and that for $T = \tau_1$, then $T^n = \tau_n$ for all $n \in \mathbb{Z}$.

Given a group G with identity element e and a set X, and given a function $\Phi : G \times X \to X$, we write

$$g(x) = \Phi(g, x).$$

If this satisfies:
(i) $e(x) = x$ and
(ii) $g_1(g_2(x)) = (g_1 g_2)(x)$ for all $x \in X$
then we call this an **action of the group** G (on the left) on the set X.

For example, an invertible transformation gives an action of the integers $\mathbb{Z}$, while a flow corresponds to an action of the additive group of the real numbers $\mathbb{R}$.

Note that each element of the group separately defines a transformation of the space, and these transformations work together so as to be consistent with multiplication in the group. For example, the transformation defined by g is a bijection of X, as combining properties (i) and (ii) we see that it has an inverse $x \mapsto \widetilde{g}(x)$ where $\widetilde{g} = g^{-1}$.

We make the similar definition for an action of a semigroup S. A possibly non-invertible transformation defines an action of the semigroup $\mathbb{N} = \{0, 1, 2, \dots\}$, with $e(x) = x$ and $n(x) = T^n(x)$.

The **orbit** of a point x for a group action is $\mathcal{O}(x) = \{g(x) : g \in G\}$. Defining two points x, y in S to be equivalent $x \sim_G y$ iff $y \in \mathcal{O}(x)$, the set S is partitioned into orbits, defining the **orbit equivalence relation** on S; that this is indeed an equivalence relation follows from (i) and (ii) above plus the fact that each element of a group has an inverse.

Remark 2.1. Traditionally, the idea of an abstract group could be introduced by way of a *transformation group*: a collection of transformations of a space which forms a group under composition. Thus, the rotations of a triangle form a group, isomorphic to $G = \mathbb{Z}_3 = Z/3Z$. But from the clearer, more modern viewpoint the rotations of the triangle are a particular action of G. The reason we say "clearer" is because the any invertible transformation of a space (which is not periodic) gives a transformation group which is isomorphic to Z, as a group; yet as an action, i.e. as a dynamical system, these can be very different. The point is that the two notions of isomorphism-of group, or of action, are not at all the same.

Group actions enter into dynamics in two quite different ways. First, the collection of orbit equivalence classes, equipped with the quotient topology and measure, defines the **orbit space** of the action; this construction can provide a convenient way of

defining a topological space or manifold on which our dynamical system, given by a flow, transformation, or (more generally) a further group or semigroup, will then act.

The most basic example is the torus $\mathbb{T}^d$ of dimension d. To define this space, we begin with $(\mathbb{R}^d, +)$, the additive group of $d-$dimensional Euclidean space, equipped with the usual topology and metric. The subgroup $\mathbb{Z}^d$, known as the **integer lattice**, acts on $\mathbb{R}^d$ by translation. Given $x \in \mathbb{R}^d$, then $x+\mathbb{Z}^d = \{x+y : y \in \mathbb{Z}^d\}$ is a **left coset** for the subgroup; since $\mathbb{R}^d$ is abelian, the collection of cosets itself defines a group, $\mathbb{R}^d/\mathbb{Z}^d$. To see why this is indeed the torus, we note that the product of d half-open intervals, $[0,1) \times \cdots \times [0,1)$, is a **fundamental domain** for the action of $\mathbb{Z}^d$ on $\mathbb{R}^d$: it contains exactly one representative from each coset. Now for $d = 1$, $\mathbb{R}/\mathbb{Z}$ is therefore homeomorphic to the closed interval $[0,1]$ modulo the relation $0 \sim 1$; that is the endpoints are identified to give the circle, $\mathbb{T}$. Similarly $\mathbb{R}^d/\mathbb{Z}^d$ is $[0,1] \times \cdots \times [0,1]$ with 0 identified with 1, hence is the product of d circles. We can make these identifications before or after taking the product of the closed intervals, so e.g. for $d = 2$ we have the square with opposite sides identified, and for $d = 3$ the cube with opposite sides identified.

A similar procedure produces a surface of any **genus** (this is just the number of "holes" of a surface, so the usual torus $\mathbb{T}^2$ has genus one), and with any number of **cusps**: these are pointed parts that go out to infinity. For this, beginning with $\mathbb{H}$, the upper half space with its hyperbolic metric, in the place of $\mathbb{R}^2$, we exchange the lattice $\mathbb{Z}^2$ for a discrete group Γ of isometries of $\mathbb{H}$ (*discrete* meaning here that the points of the subgroup are separated from each other in the continuous group G of all isometries) and $\mathbb{H}/\Gamma$ will be our surface (known as a **Riemann surface**). See Fig. 65, depicting the *punctured torus*, a torus with one cusp at infinity.

What happens for the topologically nice actions just described is that since the covering space is continuous while the orbits of the subgroup form a discrete subset, the geometry of the factor space is locally like that of the original, covering, space. One calls this a **homogeneous space** since the geometry is everywhere the same. This is true in the covering space since a neigborhood of the identity can be translated to any other point g (by g itself), and that property passes on to the quotient since the projection is a local homeomorphism and isometry.

Orbits for the next type of action, for instance given by a transformation or flow acting on one of the nice quotient spaces just described, have the opposite nature: they may wrap around in the space in a complicated, even dense, way. If so, the quotient space is no longer nice topologically, as it will be a non-Hausdorff space (Exercise 4.2)! This quotient space is not at all nice from the measure theory point of view as well: any attempt to find a fundamental domain for the action will produce a nonmeasurable set (see §10.7).

To investigate these further actions we need new tools beyond simple topology, precisely since the orbit space is no longer well-behaved. These interesting complications lead to new questions, concepts and methods, to the study of the long-term behavior of the system, and the relation to probabilistic and physical ideas like independence and entropy.

In any case, the space X on which the dynamics acts will come with some additional structure, perhaps inherited as above from a covering space (it may be a measure

space, a topological space, or a smooth manifold), with this structure preserved by the dynamics. Thus we may be studying a smooth map or flow on a manifold, or continuous dynamics on a compact topological space, or a measure-preserving map on a probability measure space (this means the total mass is finite, and has been normalized to equal one).

For this most basic case, we have a map or flow, where the parameter n or t can be thought of as time; so as we iterate the transformation we are watching the evolution of time, as the point moves along its orbit.

This is the usual setting for ergodic theory and dynamical systems theory, but as we have indicated there are interesting extensions. One is to the action of more general groups or semigroups, or even to pseudo-groups (sets of partially defined transformations) or equivalence relations (for example, foliations!). Furthermore, in all these cases the invariant measure can be infinite; there are interesting, naturally occuring examples. Beyond this, one can consider transformations which may not preserve any measure but which do preserve sets of measure zero (so-called measure-class preserving transformations; this area is known as *nonsingular* ergodic theory). There is also an extension is to nonautonomous dynamics (a sequence of maps on a space) and more generally to *nonstationary dynamics*: a sequence of maps along a sequence of spaces [AF05]. Natural examples here come from random dynamical systems, from renormalization theory, and from the study of time-varying vector fields.

And indeed, even when we want to focus on transformations and flows, these more general types of dynamics inevitably will come into play, as will be seen throughout these notes. But the foundation of the theory, and the source of the motivational intuition, lies in the study of transformations and flows; it is there that the most basic examples are found. It is that framework we shall emphasize. Indeed, even when considering actions of more general groups, we bring in a variety of ways the discussion back to transformations and flows, since all these are but different aspects of one unified vision.

For much more regarding group actions see...., and regarding nonsingular maps, see....

The structure of transformations: conjugacy, isomorphism and classification. In developing a mathematical theory, there are usually several ingredients, studied on an abstract level in category theory. When we have a transformation $T : X \to X$, we refer to it by the pair (X, T). Given two transformations (X, T) and (Y, S), and given an onto map $\pi : X \to Y$ such that $\pi \circ T = S \circ \pi$, so that the following diagram commutes, we say that π is a **homomorphism** from (X, T) into (Y, S).

$$\begin{array}{ccc} X & \xrightarrow{T} & X \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\ Y & \xrightarrow{S} & Y \end{array}$$

If in addition π is *surjective* then we call this a **factor map**, and say that (Y, S) is a **homomorphic image** or **factor** of (X, T), and conversely that (X, T) is an **extension** of (Y, S). We then also say that π is a **semiconjugacy** from (X, T) to (Y, S).

If π is invertible, then π is a **conjugacy** or **isomorphism**, and is written $T \cong S$ (or more properly, as $(X, T) \cong (Y, S)$).

Thus for example each of these collections of dynamical systems forms a category, with the objects the pairs (X, T) and the arrows the homomorphisms π:

-sets and functions from a set to itself

-topological spaces and continuous (self-)maps

-manifolds and differentiable maps

-metric spaces and Lipschitz maps

-groups and group translations, π is a homomorphism

-groups and group endomorphisms

-measure spaces and measure-preserving maps

-measure spaces and measure-class-preserving maps.

Given a type of conjugacy, we have the isomorphism problem: to classify a collection of dynamical systems up to the equivalence relation generated by this notion of conjugacy. In spirit, this is much like other classification problems in mathematics, such as the classifications in algebra of vector spaces or commutative rings or finite groups, or in topology of compact surfaces or $3-$manifolds, and in the same way one expects the proof of a classification theorem to be of greatly varying difficulty depending on the category chosen.

If we are given a class of maps with a good deal of structure, e.g. $\mathbb{C}^k$ maps on the circle, one could study these up to $\mathbb{C}^k$ conjugacy, or up to $\mathbb{C}^l$ conjugacy for any $l < k$, including $\mathbb{C}^0$ (topological) conjugacy. And, if these maps are in addition supplied with a natural invariant measure, we can study measure theoretic conjugacy. A great deal of the richness of dynamical systems theory comes from the interplay of these different structures. We will see explicit examples of this later on.

So one theme in dynamical systems theory (as in other parts of mathematics) is the study of conjugacies, and the corresponding attempts to classify collections of maps or flows up to that notion of equivalence.

Here are some examples of classification theorems:

-Ornstein's theorem [Orn73], [Shi73]: two Bernoulli shifts are measure-theoretically isomorphic if and only if they have the same entropy. This is valid also for infinite entropy. A measure-preserving transformation of a Lebesgue probability space is Bernoulli (isomorphic to a Bernoulli shift) if and only if it satisfies the property of *very weak Bernoulli*. These statements also holds for flows.

-Shub's theorem (Topological classification of expanding maps of the circle). An expanding map f of the circle with degree d is topologically conjugate to the linear map $x \mapsto dx (\mathrm{mod}\, 1)$. See §26.

-Denjoy's theorem (1932): a C^1 orientation-preserving diffeomorphism f of the circle with irrational rotation number θ and such that the derivative Df has bounded variation is topologically conjugate to the rotation R_θ.

-Herman's theorem (and related results of Herman, Yoccoz, Khanin-Sinai, Katznelson-Ornstein and others) with f as above, but at least C^3, then for a measure-one set of θ, the above conjugacy is $C^{2-\varepsilon}-$smooth.
-Stuctural stability for hyperbolic diffeomorphisms

-Franks' theorem [Fra69], [Fra70], [Man74]: (Classification of Anosov maps of the torus) An Anosov map f of the $d-$ torus is topologically conjugate to the linear map given by its action on homology.

-Kakutani equivalence (theorems of Katok, Feldman, Rudolph)

-orbit equivalence:
(Dye's theorem)(see [Zim84]) Any two ergodic measure-preserving transformations of a finite-measure Lebesgue space are orbit-equivalent.
(Krieger's theorem) there are two more equivalence classes for orbit equivalence: $\sigma-$finite infinite measure and measure-class preserving but with no equivalent invariant measure. These ideas are fundamental in C^*-algebra theory.

General theory. Another theme is the development of the general theory, within a particular category. Thus we have theorems about topological or measure theoretic recurrence, mixing, and entropy. This can be viewed as a different aspect of the study of the structure of transformations, and relates to isomorphism theory for instance by providing constants or properties which are preserved by various types of congugacy.

The theme of recurrence, which begins with a theorem of Poincaré, is remarkably rich, even for very general systems; this circle of questions has been explored especially by Furstenberg, who developed for this purpose two fundamental structure theorems, one in the topological and one in the probability measure-preserving categories.

In the measure theoretic setting there are profound links with probability theory, with the ergodic theorem of dynamics strengthening the strong law of large numbers, and with other limit laws such as the central limit theorem also coming into play. Thus there are applications of probabilistic ideas and methods to dynamics, and there is also the application of dynamical ideas within probability theory itself. See for example [FT15], [**?**].

There is also a fascinating and varied assortment of applications of dynamical ideas to number theory. For one especially remarkable example of this we mention Furstenberg's proof of Szemeredi's theorem. In 1975 Szemeredi had answered this long-standing (since 1936) conjecture of Erdös and Turán, that within any subset of the integers with positive upper density, one can find arithmetic sequences of arbitrary length. The ergodic theoretic translation of this is a statement about multiple recurrence; Furstenberg's method of proof (in 1977) was to first make this link, then prove the recurrence result, for which a key element was the structure theorem just

mentioned. Furstenberg's dynamical point of view together with Szemeredi's combinatorial approach led to a solution of a famous conjecture of Erdös by Green and Tao. Erdös had guessed that the primes also contain arbitrarily long arithmetic sequences. Now by the prime number theorem the number $\pi(n)$ of primes $\leq n$ is asymptotically $n/\log n$, whence $\lim \pi(n)/n = 0$ and they have density zero; however the primes are still a large set in that their integer Hausdorff dimension, defined as in [BF92] to be $d = \lim(\log \pi(n)/\log n)$, is one. We mention that Erdös' conjecture on arithmetic sequences in the primes is a special case of a still open conjecture of his from 1976: that if for $A \subseteq \mathbb{N}$, $\sum_{n\in\mathbb{N}} \frac{1}{n} = \infty$, then A contains arithmetic sequences of arbitrary length.

Examples. A third theme is the study of specific examples. Much of the fascination of dynamical systems theory comes from the profound beauty of these examples, which have been introduced over time by the many researchers in the subject.

The problems of conjugacy, classification, and the general theory can only be appreciated given a firm grounding in the examples. These help to guide the development of the theory, to provide a constant source of new challenges, and to serve as a test and check on one's wilder intuitive guesses. Indeed, they should probably occupy at least half the time of any dynamics course, and of any student's or researcher's thinking time!

In this course we will begin with some of the classical examples; this will give us something to think about as we introduce the general theory. These will often be related to examples in physics, or from ODEs, although we will not emphasize this aspect of the material, providing only a glimpse in the section referred to aabove. For more in that direction, the reader may consult the references cited there.

3. Measure and randomness.

We recall the basic definitions, just to get started; however in general an analysis background at the level of [Roy68] will be assumed. In addition to Royden my own favorites are: [Bar66], [Oxt80], [Rud73] and the more recent, also excellent [Fol99]. We go into more depth on these matters in §33.1.

Given a set X, an **algebra** $\mathcal{A}$ is a collection of subsets of A such that, where $A^c = X \setminus A$ denotes the complement:

- $X \in \mathcal{A}$;
- $A \in \mathcal{A} \implies A^c \in \mathcal{A}$;
- $A, B \in \mathcal{A} \implies A \cup B \in \mathcal{A}$.

  It is a $\sigma-$**algebra** if in addition
- $A_i \in \mathcal{A}$ for $i = 1, 2, \cdots \implies \cup_{i=1}^{\infty} A_i \in \mathcal{A}$.

A *measurable space* is a pair $(X, \mathcal{A})$ where $\mathcal{A}$ is an algebra of subsets of X.

A function $\mu : \mathcal{A} \to [0, +\infty] = [0, +\infty) \cup \{+\infty\}$ is a **finitely additive** measure if $\mu(\emptyset) = 0$ and for A, B disjoint, $\mu(A \cup B) = \mu(A) + \mu(B)$. It is a **measure** if in addition, $\mathcal{A}$ is a $\sigma-$algebra and μ is **countably** additive, i.e. for $\{A_i\}_{i=1}^{\infty}$ disjoint, then $\mu(\cup_{i=1}^{\infty} A_i) = \Sigma_{i=1}^{\infty} \mu(A_i)$. A *measure space* is a triple $(X, \mathcal{A}, \mu)$. In the special case

where $\mu(X) = 1$ then $(X, \mathcal{A}, \mu)$ is called a **probability space** and μ a **probability measure**.

Given two measurable spaces $(X, \mathcal{A}), (Y, \mathcal{B})$ then $f : X \to Y$ is **measurable** iff $f^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$. This is written as $f^{-1}(\mathcal{B}) \subseteq \mathcal{A}$. The map is **invertible** iff it has an inverse in the category of measurable maps; that is, not only do we require that the function f is bijective, but that the inverse map is itself measurable. Thus, for an invertible map, not only the preimage but also the *forward* image of a measurable set is measurable. (Exercise: find an example of a map and a measurable set whose image is nonmeasurable!)

Given a measure μ on X, then $f_*\mu$ is the measure on Y defined by $(f_*\mu)(B) = \mu(f^{-1}(B))$. This is called the **push-forward** of μ. A function "pushes points forward" while the natural operation on sets is to pull back via the inverse image (as that preserves the set operations of intersection, union, and complement) which then leads to the the natural operation on measures, of pushing forward; by contrast, differential forms pull back hence one uses the upper star, f^*, for that action, see §45.2.

In the case where f maps X to itself, then f is called a **transformation** of X. A set A is **invariant** for the transformation iff $f^{-1}(A) = A$; a measure μ on X is invariant (or is **preserved** by the map) iff $\mu(f^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{A}$, iff $f_*\mu = \mu$. If a group G acts on a measure space, then A or μ is invariant iff that is true for the transformation given by each $g \in G$.

If $(X, \mathcal{T})$ is a topological space, then the **Borel** $\sigma-$algebra is the smallest $\sigma-$algebra containing the topology (i.e. the collection of all open sets) $\mathcal{T}$. Exercise: this makes sense!

The first idea of measure theory is to generalize length, area, volume from simple to more wild sets. So for instance Lebesgue measure on $\mathbb{R}^d$ is defined to be volume on balls, and then is extended to the Borel $\sigma-$algebra $\mathcal{B}$ (in a unique way, which takes some proof). This can then be further extended to the **Lebesgue** $\sigma-$algebra $\widehat{\mathcal{B}}$, which is the smallest $\sigma-$algebra containing $\mathcal{B}$ which is **complete**: if $\mu(A) = 0$, (a **null set**), then every subset of A is in $\widehat{\mathcal{B}}$.

The second idea of measure theory is to model the notions of probability and randomness. Following probability convention, let us write our measure space as $(\Omega, \mathcal{A}, P)$; here P is a probability measure. An **event** is a measurable subset of the measure space. Thus $P(A)$ is interpreted as the probability that a point, randomly chosen from the "sample space" Ω, in fact belongs to A; that is, $P(A)$ is the probability that the event A will occur. (Note: we shall subsequently in general write μ, ν, ρ and so on for measures).

Example 1. *Choosing a point randomly from the circle.* What does it mean to choose a point randomly from the circle? *Rigorous mathematical answer:* Choose it with respect to (normalized) Lebesgue measure. That is, we can't exactly "choose" the point randomly, but we *can* tell you what the probability should be that the randomly chosen point lies in a given Borel subset.

Thinking about this, the justification is that Lebesgue measure is the unique probability measure on the circle which is invariant by rotations. And by "random" here we probably mean: chosen in a way that is rotation–independent.

Example 2. *Choosing a point randomly from a compact group.* We do the same- but now the measure is **Haar measure**, the unique translation-invariant probability measure on the group.

Example 3. *Choosing a point randomly from a finite set.* We should have started with this example, but we wanted to motivate our answer, which is the "obvious" one: if our set X has n points, then we should give equal mass to each. What is hidden here is a tacit assumption of invariance: this **uniform distribution** μ is the only probability measure invariant with respect to the group of symmetries of X, the **symmetric group** σ_n of all permutations.

Alternatively, we can give X more structure, identifying it with the compact group $\mathbb{Z}_n$, the additive group of integers modulo n, and μ is the Haar measure and is the only translation-invariant probability measure.

The idea of the symmetric group generalizes to:

Example 4. *Group-invariant measures.* Suppose a group G acts continuously on a topological space $(X, \mathcal{T})$, and that μ is a Borel probability measure on X which is invariant for this action. Then μ is a candidate for "choosing a point randomly on X". If there is only one such μ, then the action is **uniquely ergodic**. For a first example, the group of rotations acting on the circle is uniquely ergodic; more generally, this holds for any compact group, action on itself by left translation (i.e. by multiplication on the left) by the uniqueness of left Haar measure. Whether or not this holds for some given transformation or flow will be an important theme in these notes.

Example 5. *Choosing a point randomly from a compact Riemannian manifold.* Here the natural choice is normalized Riemannian volume. Now there may not be a group action to describe the invariance, but at least there is one locally (formally, a groupoid or pseudogroup).

Example 6. *Choosing a point randomly from a fractal set.* For a nice fractal set of Hausdorff dimension d, embedded in Euclidean space $\mathbb{R}^n$, or more generally in a Riemannian manifold, we should use Hausdorff $d-$ dimensional measure H_d (restricted to the set and normalized). This is defined so as to be translation-invariant (in $\mathbb{R}^n$, or with respect to the pseudo-group in the manifold), but there is a hidden scaling invariance as well: H_d is a **conformal measure**: for any Borel set A and any $a > 0$, then $H_d(aA) = a^d H_d(A)$. Noting that this is exactly the scaling property satisfied by Lebesgue measure in R^n for $n = 1, 2, \ldots$, H_d is a natural generalization to "fractional dimension", i.e. to non-integer dimension $d \geq 0$. A basic example is the Cantor set; another is the snowflake curve, see Fig. 1. For the precise definition of H_d see ... below.

The third idea of measure theory is to generalize the Riemann integral to a much wider (and very useful) class of functions. Recall that, given a real-valued function f on X we define its integral with respect to μ as follows. Recall that the **indicator function** of a set A is:

$$\chi_A(x) = \begin{cases} 1 & \text{for } x \in A \\ 0 & \text{for } x \notin A \end{cases}$$

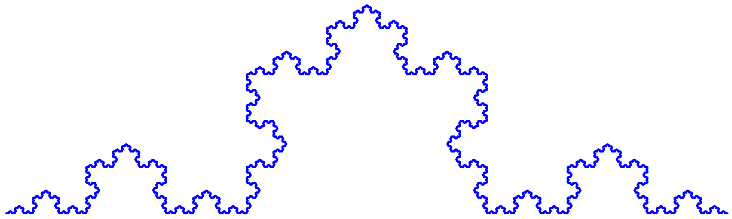


FIGURE 1. The Koch snowflake curve

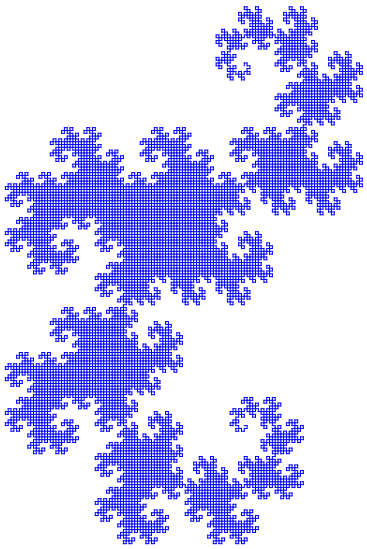


FIGURE 2. The dragon region: its boundary is a fractal curve.

Then $\int_X \chi_A \mathrm{d}\mu = \mu(A)$. We extend this linearly to all finite sums of indicator functions (the so-called **simple functions**) and then extend further to all measurable functions by taking limits (**measurable** means as a function from $(X, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B})$ where $\mathcal{B}$ is the Borel $\sigma-$algebra). One shows this extension is unique [Roy68], [Bar66]. See also Theorem 33.15 below.

This brings us to a fourth idea of measure theory: to define the **mean (or average) value** value of a function f. For the special case when μ is a probability measure, this is written $\overline{f}$, and is its integral: $\overline{f} \equiv \int_X f \mathrm{d}\mu$. Of course any finite measure μ can be turned into a probability measure by **normalization**: dividing by the total mass. Then to compute the mean value of a function with respect to μ we integrate with respect to its normalization.

In probability theory, a (usually real-valued) measurable function has a special name: a **random variable**. Given a probability space $(\Omega, \mathcal{A}, P)$, one often writes $X : \Omega \to \mathbb{R}$ for this function. The idea here is that X is not just a variable (as, say, in elementary algebra) but one to which a notion of randomness has been attached: the randomness is supplied by the **underlying probability space** $(\Omega, \mathcal{A}, P)$, and the probability that X has values in a Borel set $B \subseteq \mathbb{R}$ will be $P(X^{-1}(B))$. Via this

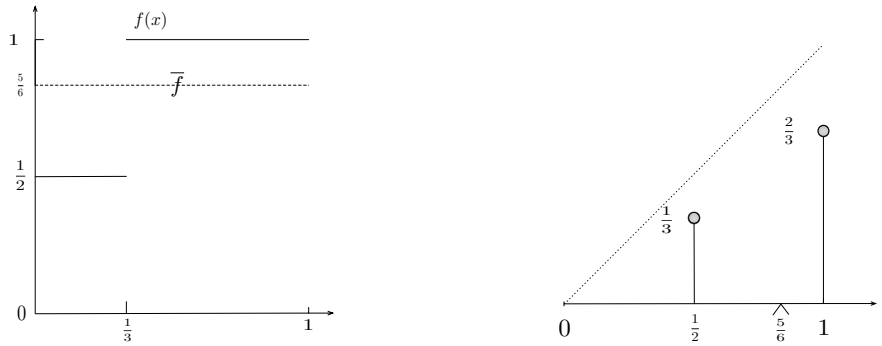


FIGURE 3. The mean value of a random variable (a function) equals the mean value, or center-of-mass, of its distribution (a measure): $\int_X f(x)\mathrm{d}\mu(x) = \int_{\mathbb{R}} y\mathrm{d}\mu_f(y)$.

equation, the measure P has been pushed forward to a (probability) measure P_X on $\mathbb{R}$, that is, $P_X = X_*(P) = P \circ X^{-1}$. In probability notation, the set $X^{-1}(B)$ is commonly written as $[X \in B]$, interpreted as "the event that the random variable X is in B". So the measure P_X, called the **distribution** of X, tells us the likelihood that this random variable assumes a certain value; thus, $P_X(B) = P[X \in B]$, read "the probability that X is is B". Two different random variables X_1, X_2 (whether defined on the same underlying probability space (Ω, P) or on unrelated spaces (Ω_1, P_1) and (Ω_2, P_2) are termed **identically distributed** iff $P_{X_1} = P_{X_2}$.

The **expected value** of a random variable X is probability language for the mean value of that measurable function: $\mathbb{E}(X) \equiv \int_\Omega X\mathrm{d}P$, and indeed this is also called the **mean** of X. The idea is that if you perform the experiment of choosing a point ω from Ω randomly (i.e. with respect to the measure P) then the expected value is the average value of the outcome.

In terms of the distribution P_X, this equals $\int_{\mathbb{R}} x\mathrm{d}P_X$, see Fig. 3. We recall that for a normalized measure μ on $\mathbb{R}$, $\int_{\mathbb{R}} x\mathrm{d}\mu$ is its center of mass. This is the average location of a point with that mass distribution; since the location of the point x is $y(x) = x$, we integrate that function against μ. The equality $\mathbb{E}(X) = \int_{\mathbb{R}} x\mathrm{d}P_X$ says therefore that $\mathbb{E}(X)$ is the center of mass of the distribution P_X of X.

We sketch the proof:

Proposition 3.1. *Let $X : \Omega \to \mathbb{R}$. Then*

$$\int_\Omega X\,dP = \int_{\mathbb{R}} x\,dP_X.$$

Proof. Suppose first that $X = \chi_A$. Then

$$P_X(B) = \begin{cases} P(A) & \text{if } 1 \in B, \\ 0 & \text{if } 1 \notin B. \end{cases}$$

So

$$\int_\Omega X\mathrm{d}P = P(A)$$

but also $\int_{\mathbb{R}} x \mathrm{d}P_X = x(1) \cdot P(A) = P(A)$ so the result holds in this case.

Next consider a linear combination $X = aX_1 + bX_2$. Then $P_{aX_1+bX_2} = aP_{X_1} + bP_{X_2}$, by the properties of inverse images of sets. Now the integral is also linear (in the measures) so the formula hols for X, and hence for any finite linear combination $X = \sum_1^n a_i X_i$; taking $X_i = I_{A_i}$, these are the *simple functions*. Lastly this equation also respects increasing limits, due to the Monotone Convergence Theorem ([Roy68], p. 84): for $f_n \geq 0$ and increasing, then $\int f_n \to \int f$. Since simple functions approximate measurable functions, this holds for all nonnegative measurable functions, and, including negative values, for all integrable functions.

□

Proposition 3.2. *Let $f : \mathbb{R} \to \mathbb{R}$ be measurable. Then*

$$\int_\Omega f(X) dP = \int_{\mathbb{R}} f(x) dP_X$$

where one side exists iff the other does.

In analysis notation, this is a change-of-variables theorem: given a measure space $(X, \mathcal{A}, \mu)$ and measurable spaces $(Y, \mathcal{C})$, $(\mathbb{R}, \mathcal{B})$ with B the Borel sets, $f : X \to Y$ and $g : Y \to \mathbb{R}$ then

$$\int_X g \circ f \, d\mu = \int_Y g d(f_* \mu).$$

Proof. Sketch of proof: just like that above, except we now use step functions for the function g, instead of for f. (The reader should make this precise!) □

See also pp. 5,6 [Lam66].

In probability theory one also can use the term *distribution* more generally to refer to a probability measure (countably additive whenever possible!) on a measurable space with some additional structure, more than just that of a measure space: possibly the reals (as above) but possibly a topological vector space, a metric space, or a manifold.

Calling this space M, with Borel $\sigma-$algebra $\mathcal{B}$, then the more general definition of a random variable is a measurable function X from some probability space Ω to M. Given a distribution μ on M (by which we mean a Borel probability measure) there are many possible choices for this underlying space $(\Omega, \mathcal{A}, P)$; the simplest such choice is $(M, \mathcal{B}, \mu)$ itself, with the random variable X the identity map, and so $P = P_X = \mu$, but it is often both natural and useful to make some other choice.

Remark 3.1. In probability theory, one is often emphasizing distributions (measures) on $\mathbb{R}$, where it can be convenient to use the Riemann-Stieltjes instead of the Lebesgue integral. We remind the reader how this works notationally. Given a measure μ on $\mathbb{R}$, the function $F : \mathbb{R} \to [0, +\infty]$ defined by

$$F(x) = \mu((-\infty, x])$$

is known as the **cumulative distibution function** of μ. Note that this is a nowhere decreasing, right-continuous function (thus a **càdlàg** function: *continu à droite, limites à gauche*: right-continuous, with limits existing from the left). One then can

reasonably write also $\mathrm{d}F$ for $\mathrm{d}\mu$, for the following reason. Supposing for simplicity that μ is given by a distribution $f(x)\mathrm{d}x$; then by the Fundamental Theorem of Calculus, $F'(x) = f(x)$, whence

$$\int_{-\infty}^{a} \mathrm{d}\mu = \int_{-\infty}^{a} f(x)\mathrm{d}x = \int_{-\infty}^{a} \frac{\mathrm{d}F}{\mathrm{d}x}\mathrm{d}x = \int_{-\infty}^{a} \mathrm{d}F.$$

In Riemann-Stieltjes integration theory, this notation is extended to measures μ for which such a function $f(x)$ may not exist: e.g. if μ is a **singular** measure, whose mass lives on a subset of Lebesgue measure zero. The simplest example is point mass at 0, $\mu = \delta_0$, where we have

$$F(x) = \begin{cases} 0 & \text{for } x < 0, \\ 1 & \text{for } x \geq 0 \end{cases}$$

and indeed the "derivative" of this discontinuous function is the measure δ_0.

Precisely what we mean by this is explained by the theory of differentiation of measures (see e.g. [Roy68], [Rud70]). In the measure theory (Lebesgue integral) notation, writing m for Lebesgue measure on $\mathbb{R}^d$, then if μ is absolutely continuous with respect to m ($\mu << m$), this theory gives $f = \mathrm{d}\mu/\mathrm{d}m$ whence $f\mathrm{d}m = (\mathrm{d}\mu/\mathrm{d}m)\mathrm{d}m = \mathrm{d}\mu$. Thus the Riemann-Stieltjes theory allows one to find antiderivatives for singular measures on $\mathbb{R}$.

In what sense this can be further extended is addressed by the beautiful functional analysis theory of Schwartz distributions, see e.g. [Rud73]. The terminology can lead to confusion, since Laurent Schwartz' distributions are linear functionals defined on special function spaces, but are not always measures (whereas in probability theory a distribution is always a probability measure, by definition). For a nice example the derivative of F is the measure δ_0, either by the theory of differentiation of meaures or by Schwartz' theory, but what is the derivative of δ_0? (Answer: a *dipole*, which is a Schwartz distribution but not a measure! And then one can differentiate the dipole, differentiate that....)

Although here we mostly will use the Lebesgue integral notation, the probability theory notion of distribution function can be very useful, for instance when trying to visualize measures on $\mathbb{R}$ which are more complicated than say the point masses pictured in Fig. 3. A good example is provided the Hausdorff measure on the Cantor set (see §4.5 below), where the distribution function $\widehat{\beta}$, pictured in Fig. 6, is an everywhere continuous, almost-everywhere constant function, whose derivative on the senses just discussed is this singular measure!

We return to the question of what we mean by randomness of a distribution, that is, to how one might (usefully and meaningfully) model the intuitive notions of randomness, probability, and chance that we have.

In our discussion so far, our idea of randomness comes from **invariance** of the distribution on a space M with respect to some symmetries, that is, with respect to the action of some group G on M. If it is unique, the resulting symmetry–invariant distribution is called the **uniform** distribution on M. (If it isn't, one might look for a larger symmetry group which will give uniqueness!)

3.1. **Product spaces and independence.** A deeper idea of randomness comes via the notion of *independence*. Given a measure space $(X, \mathcal{A}, \mu)$, and subset A with $0 < \mu(A) < \infty$, we define the **restriction** of μ to A by $\mu|_A(E) = \mu(A \cap E)$; we then normalize this to a probability measure, defining for $E \in \mathcal{A}$,

$$\mu_A(E) = \mu(E \cap A)/\mu(A).$$

This is the **relative measure** on A. In probability terminology, this defines the **conditional probability** $\mu(E|A) = \mu_A(E)$, and describes the probability of the event E given the event A.

Two measurable subsets sets A, B of a probability space are **independent** iff $\mu(A \cap B) = \mu(A)\mu(B)$. Equivalently, $\mu_A(B) = \mu(B)$, so if X is a probability space the probability of the occurence of the event B does indeed not depend on whether (or not) A occured.

Remark 3.2. One might think intuitively that two events A, B are independent if they "have nothing to do with each other", meaning that they are mutually exclusive, i.e. that $P(A \cap B) = 0$. But that would mean they are in fact dependent, since if one occurs the other can't. There is one exception to this:

Exercise 3.1. Show that:
(i)

$$(P(A \cap B) = 0 \text{ and } A \text{ is independent of } B) \implies (P(A) = 0 \text{ or } P(B) = 0).$$

Furthermore:
(ii)

$$(\forall B, B \text{ is independent of } A) \iff (P(A) = 0 \text{ or } 1.)$$

We thank Marina Talet for conversations about these matters.

Two random variables X, Y defined on the same underlying probability space (Ω, P) are defined to be independent iff for every Borel set B, the sets $X^{-1}(B)$ and $Y^{-1}(B)$ are independent, that is, iff the events $[X \in A]$ and $[Y \in B]$ are.

A geometric model for independence comes from product measure. Given two measurable spaces $(X, \mathcal{A})$,$(Y, \mathcal{B})$, with algebras $\mathcal{A}, \mathcal{B}$ then a **rectangle** or **cylinder** is a set of the form $A \times B$, for $A \in \mathcal{A}, B \in \mathcal{B}$. The *product algebra* is the smallest algebra containing all the rectangles. We note that:

Exercise 3.2. The product algebra is the collection of all finite disjoint unions of rectangles.

See [Bar66], Lemma 10.2. The *product* $\sigma-$*- algebra* is the smallest $\sigma-$algebra containing all the rectangles.

Proposition 3.3. *Suppose we are given ... there exists a unique xextension... (???TO DO)*

Given two measure spaces $(X, \mathcal{A}, \mu)$,$(Y, \mathcal{B}, \nu)$, then the product measure space is $(X \times Y, \mathcal{A} \times \mathcal{B}, \mu \times \nu)$ where by definition, $\mathcal{A} \times \mathcal{B}$ is the product $\sigma-$algebra, while $\mu \times \nu$ is defined as follows: on rectangles it is simply the product, and then on the

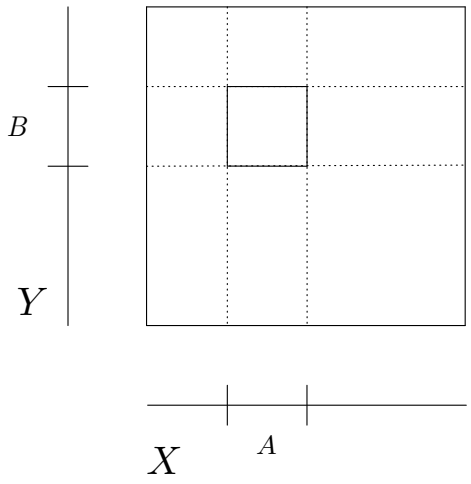


FIGURE 4. A geometric model of independence: product measure.

rest of the product $\sigma-$algebra one takes the unique extension (which one proves to exist!). See ??? below.

Assuming X, Y are both probability spaces, then the sets $A \times Y$ and $X \times B$ are independent; see Fig. 4.

Exercise 3.3. Given a a probability space $(A, \mathcal{A}, \mu)$, verify that A is independent of B iff A^c is. Show A, B are independent iff the random variables χ_A and χ_B are. Prove that if X, Y are independent random variables, then

$$\int_\Omega XY \mathrm{d}P = \int_\Omega X \mathrm{d}P \int_\Omega Y \mathrm{d}P.$$

For m Lebesgue measure on $I = [0, 1]$, show that if two complex- or real-valued square-integrable functions f, g with mean zero are independent then they are orthogonal in L^2, but not necessarily conversely. Defining F, G on the square $I \times I$, with Lebesgue meaure $m \times m$, by $F(x, y) = f(x)$, $G(x, y) = g(y)$, show that F and G are independent random variables.

How should the definition of independence be changed if we have a finite measure space of measure $c > 0$? (See Remark 8.3).

Product measure can be used to model the following situation. We are given a probability space $(X, \mathcal{A}, \mu)$, which defines our idea of randomness for choosing a point from X. Thus, in our previous examples, X may have some symmetry group acting on it for which μ is the unique invariant measure; X may be a finite set with equal mass on every point; μ may be the Gaussian distribution on $X = \mathbb{R}$ if we have some reason to say that is the way our probability is distributed. Then the experiment of choosing not one point from X, but three points in succession, and independently, is modelled by product measure $\mu \times \mu \times \mu$ on $X \times X \times X$. Product measure can also be written $\mu \otimes \mu \otimes \mu$, since it can be viewed as a tensor product (see Example 63) with measures viewed as elements of the dual space of continuous functions.

Product measure is extended to countable infinite products as follows. Given $(X_i, \mathcal{A}_i, \mu_i)$ for $i = 1, 2, \dots$ then $\Pi_{i=1}^{\infty}\mathcal{A}_i$ will be the $\sigma-$algebra generated by all *finite* cylinders, i.e. by all sets of the form $(\dots X_k \times X_{k+1} \times A_{k+2} \times \cdots \times A_n \times X_{n+1} \times \dots)$. Again, there is a unique additive extension, which will be countable additive if each μ_i is.

This leads us to the model for an infinite sequence of independent choices, for instance an infinite sequence of tosses of a fair coin:

Example 7. *Coin-tossing.* Recalling the Dirac delta notation δ_x for point mass on x, i.e. this is the measure defined by $\delta_x(A) = 1$ if $x \in A$, 0 otherwise, then we give the **alphabet** $\mathcal{A} = \{1, -1\}$ the discrete topology, i.e. every subset is open, and the measure $\mu_i = 1/2(\delta_1 + \delta_{-1})$ for each $i \geq 0$; we then define $\Omega = \Pi_0^{\infty}\{1, -1\}$ with infinite product measure $\mu = \otimes_0^{\infty}\mu_i$ and the Borel $\sigma-$algebra $\mathcal{B}$ generated by the product topology. A point in Ω is written $X = (X_0, X_1, \dots)$. We define $X_i : \Omega \to \mathcal{A}$ to be the i^{th} coordinate function, thus $X_i(X) = X_i$. Then $(X_i)_0^{\infty}$ is a sequence of independent random variables.

Note that the following two procedures are equivalent: choose a single point X randomly from Ω (randomly means with respect to the measure μ); making an infinite sequence of independent random choices from $\mathcal{A}$. This last corresponds to flipping a fair coin infinitely many times.

According to Billingsley's colorful account, [Bil68], it is the Greek goddess of chance, Tyche, who makes these random choices! Although mathematically equivalent, the point of view and intuition behind these two operations is completely different, the first being more measure-theoretic and the second probabilistic. Much of the power of modern probability and ergodic theory comes from a constant interchange between these two.

Now consider the map $\Phi : \omega \mapsto \underline{X} = (X_0, X_1, \dots)$; if the random variables take values in a set $\mathcal{R}$, then $\underline{X}$ is an element in sequence space $\Pi = \Pi_0^{\infty}\mathcal{R}$. The map Φ pushes forward the measure P to a measure $P_{\underline{X}}$ on Π. More precisely, if $\mathcal{A}$ is the $\sigma-$algebra on $\mathcal{R}$, then we define $\widetilde{\mathcal{R}}$ to be the $\sigma-$algebra on Π generated by the algebra of **cylinder sets**. These are the subsets of Π which are finite intersections of sets of the form $[X_i \in A_i]$; that is, the finite collections of events. The stochastic process is now represented by a measure on sequence space, and we can do away with the underlying space altogether, and simply consider the measure space $(\Pi, \widetilde{\mathcal{R}}, \mu)$ with $\mu = P_{\underline{X}} = P \circ \Phi^{-1}$. From a different viewpoint, the underlying space has been replaced by this sequence space, and the random variables $\omega \mapsto X_i(\omega)$ by the coordinate functions $\underline{X} \mapsto X_i$. This is the **path space model** of the stochastic process: a path space with a probability measure, and the sequence of coordinate functions, with the index interpreted as times $\dots, 1, 2, 3, \dots$.

The question of "what is randomness" has been converted, in this way, to the question of how to choose an appropriate measure: on the underlying space Ω, on the space of possible values $\mathcal{R}$, or, for a stochastic process, on the sequence space Π; as we saw in the examples, what we really mean by "random" is that this choice should exhibit some sort of uniformity or invariance, frequently expressed formally by the action of a group.

Returning to the example of a finite set, flipping a *fair coin* will produce the outcome heads (= 1) or tails (= 0) with equal probability; this is modelled by the group $\mathbb{Z}_2 = \{0, 1\}$ with Haar measure.

3.2. **From maps to stochastic processes.** We have just encountered, via Kolmogorov's extension, a sequence $(X_i)_{i\in\mathbb{N}}$ of random variables defined on the same underlying probability space (Ω, P). This defines a discrete-time **stochastic** (or **random**) **process**. The choice of $\omega \in \Omega$ determines the sequence of values $X_1(\omega), X_2(\omega), \ldots$. That is, an element of a stochastic process is a random *sequence*. We also encountered $(X_t)_{t\in\mathbb{R}}$; this defines a continuous-time stochastic process: a random element of $\mathbb{R}^{\mathbb{R}}$, that is, a random *function*, $X_t = X(t)$.

How, then, do we get reasonable or interesting examples of stochastic processes? A central example was given in the previous section: that of coin-tossing, based on the fundamental notion of independence. That is, independence defines the joint distributions, and Kolmogorov's extension theorem in this case just yields infinite product measure.

A basic and simple way to define joint distributions beyond the setting of independence is via the *Markov property*; we describe this below.

Moving beyond even this brings us to the fundamental link between dynamics and probability. If we are given a measure-preserving transformation T of a measure space $(X, \mathcal{A}, \mu)$ of total mass 1, and a measurable real-valued function f on X, then the sequence of functions $f_n = f \circ T^n$ defines a discrete-time stochastic process: these are the random variables, with $(X, \mathcal{A}, \mu)$ the underlying probability space. The measure-preserving property of the dynamical system is reflected in the **stationarity** of this process (by definition, that the joint distributions of the random variables are shift-invariant).

Indeed, coin tosses can be modelled geometrically by means of the measure-preserving map $g(x) = 2x (\mathrm{mod}\, 1)$ of the unit interval $I = [0, 1]$; the link to the sequence space of coin-tossing just described comes from expressing each point in its binary expansion. See §4.2 below.

Similarly, to find natural examples of continuous-time processes, we can begin with a measure-preserving flow τ_t on (X, μ), and an integrable function f; then $f_t(x) = f \circ \tau_t(x)$ defines our process. Here we do not need to worry about Kolmogorov's extension theorem, as the underlying space (Ω, P) is just our measure space (X, μ), and again the stationarity of the process is implied by the fact that τ_t preserves the measure.

3.3. **Space averages and time averages.** For both discrete and continuous time, the random sequence determined by the choice of the point x, respectively $f_n(x)$ (repectively $f_t(x)$) is called a **path** of the process. Next consider the following question: given a path, what is its average value? Thinking of the parameter n or t as time, a natural answer is given by the **time average:**

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f_n(x)$$

respectively,

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T f_t(x)\mathrm{d}t.$$

This brings us to the statement of this remarkable theorem of Birkhoff (see Theorem 38.1 for a proof); we give three versions of it.

Theorem. (Birkhoff Ergodic Theorem)
(i) (ergodic case) Let T be a measure-preserving transformation of a probability measure space $(X,\mathcal{A},\mu)$, which is **ergodic**, i.e. it has no nontrivial invariant subsets. Then for any $f\in L^1$ (see §6.2), for almost any $x\in X$, the limit

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^i(x))=\overline{f}\equiv\int_X f(x)\mathrm{d}x.$$

(ii) In the above situation, let $A\in\mathcal{A}$, and write $N_A(x)$ for the number of returns of the point x to the set A up to time n. Then almost any $x\in X$,

$$\lim_{n\to\infty}N_A(x)/n=\mu(A).$$

(iii)(general case) Let T be a measure-preserving transformation of a probability measure space $(X,\mathcal{A},\mu)$. Then for any $f\in L^1$, the limit

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^i(x))=f^*(x)$$

exists almost surely; the function f^* is invariant and integrable, and for any measurable $E\subseteq X$ which is invariant i.e. $T^{-1}(E)=E$, $\int_E f d\mu=\int_E f^* d\mu$.

For more on ergodicity, see Definition 5.1 below.

Note that (iii) implies (i) which implies (ii) (just take $f(x)=\chi_A(x)$). There is also a flow version. The probability theory version (which by Kolmogorov's theorem, see §33.9, is equivalent to (i)) for discrete time is:
(iv) Let $(X_i)_{i\geq 1}$ be an ergodic stationary stochastic process with finite mean $\mathbb{E}(X_i)$. Then for a.e. ω,

$$\frac{1}{N}(X_1+\ldots X_N)\to\mathbb{E}(X_1)\text{ as }N\to\infty.$$

For continuous-time stochastic processes the statement is equivalent to the theorem for flows.

Birkhoff's ergodic theorem can be expressed in words as follows:

For an ergodic transformation or flow, the time average of a function equals its space average.

Part (ii) says that the frequency of time the point x spends in the set is equal to the measure of that set; that is, the frequency of times the event is observed equals the overall probability of that event.

In the next sections we describe how these questions fit into a wider context.

3.4. **A fifth motivation for measures: linear functionals and the dual space as a system of coordinates.** Let us recall, given a finite-dinensional vector space V over a field K (for example, $\mathbb{R}$ or $\mathbb{C}$) the definition of the *dual space* V^* of V.

Definition 3.1. The dual space is the collection of all *linear functionals*, that is, all linear maps $\lambda : V \to K$.

Now V^* itself is a vector space, and one proves in Linear Algebra that it has the same dimension as V.

We remark on how we prove that V^* itself is a vector space: to define the sum operation, we have to specify which function is $\lambda_1 + \lambda_2$, and for this we simply add the values at each point: $(\lambda_1 + \lambda_2)(\mathbf{v}) = \lambda_1(\mathbf{v}) + \lambda_2(\mathbf{v})$. This seems "obvious" but what is *not* so obvious is that it is necessary to make this definition!

Indeed, in the same way, the collection of functions from any set X to any vector space W forms a vector space, verification of the axioms being immediate. A key example is for $\mathcal{C}(\mathbb{R})$, the collection of all continuous real-valued functions on the reals. Other examples are the L^p spaces, see below.

One reason for being careul with the definitions is that things change radically exactly in these cases, where the dimension of V is infinite. Here one should specify a topology (it is then called a *topological vector space*) and then V^* is defined to be the space of *continuous* linear functionals.

Now for V finite dimensional, suppose we are given an inner product defined on V. Then we can *represent* an element $\lambda \in V^*$ by an element of V itself, since given $\mathbf{w} \in V$, then

$$\lambda_{\mathbf{w}}(\mathbf{v}) = \mathbf{w} \cdot \mathbf{v} = \langle \mathbf{w}, \mathbf{v} \rangle$$

defines a linear functional, and conversely, given $\lambda \in V^*$, there exists a unique such $\mathbf{w}$ which represents it (consider what
lambda does on a basis).

Similarly for (signed) measures one often writes:

$$\langle \mu, f \rangle = \mu(f) = \int_J f(x) \mathrm{d}\mu.$$

For an example, as shown by the Riesz Representation Theorem (see... below), for $\mathcal{C}([a,b])$, the topological vector space of continuous real-valued functions on the compact interval $J = [a,b]$, with the topology of uniform convergence (i.e. the sup norm), see below???), then each finite measure μ defines a linear functional λ_μ via

$$\lambda_\mu(f) = \int_J f(x) \mathrm{d}\mu.$$

Moreover if we include *signed measures*, that is, $\mu = \mu_1 - \mu_2$ where μ_1, μ_2 are finite measures, then we have found *all* linear functionals.

We then think of these are providing a system of coordinates for V. These coordinates are indexed by V^* itself; that is, the "μ^{th}-coordinate" of f is

$$\mu(f) \equiv \int_J f(x) \mathrm{d}\mu.$$

For example, if $\mu = \delta_x$ is point mass at x, then $\delta_x(f) = f(x)$. Thus, the value of f at x is the x^{th}-coordinate of f. In fact, this agrees with our usual understanding of coordinates of $\mathbb{R}^n$: for $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ then this vector is a finite sequence, hence a function $\mathbf{x} : \{1, 2, \ldots, n\} \to \mathbb{R}$ and $\mathbf{x}(j) = x_j = \delta_j(\mathbf{x})$.

Indeed, this provides a way of visualising the coordinate axes of the vn- dimensional space $\mathbb{R}^n$ or, just as easily, of the infinite-dimensional space $\mathcal{C}([a, b])$. If $n = 3$, we can locate the axes perpendicular to each other, but already for $n = 4$ this no longer works, but what we can do is to place these four real lines vertically parallel. And the same works for $\mathcal{C}([a, b])$! Every vertical line in the plane is now a coordinate axis.

Every other measure can be approximated by a linear combination of point masses, but rather than trying to find the most efficient system of coordinates, by choosing a topological basis of V^*, it is often more convenient to simply take all of V^* at once.

This is the "idea" of *Functional Analysis*: we use the dual space to coordinatize the infinite-dimensional topological vector sspace (TVS) V, and then we "do analysis" on these coordinates, which are just real numbers!

Thus it becomes important to identify the dual spaces of the TVS of interest. Here are some examples encountered below:

– For X a compact metric space, then for $V = \mathcal{C}(X, \mathbb{R})$ then V^* is the space of finite signed (countably additive) measures;

– For X a noncompact metric space, then for the space of continuous bounded functions with the sup norm, $V = \mathcal{CB}(X, \mathbb{R})$ then V^* is the space of finite signed *finitly* additive measures.

– For $0 < p < \infty$, where $\frac{1}{p} + \frac{1}{q} = 1$, then for the spaces $L^p(X)$ where X is some measure space, then the dual space of L^p is L^q (and vice-versa). The case L^∞ is similar to $\mathcal{C}$. The special case of $p = q = 2$ is Hilbert space, the only one of these Banach spaces where the norm comes from an inner product.

– the spaces of Schwartz distributions also come up as dual spaces of appropriate function spaces. See [Rud73].

In some sense the whole point in Functional Analysis is to choose the function space and topology (and hence dual space) which is suited to the problem being studied. And unlike the case of finite dimensions, where all norms are equivalent, these choices bring up interesting and subtle differences.

As we have just mentioned, for a noncompact space like $\mathbb{R}$, the the linear functionals of $\mathcal{C}$ are represented by the *finitely* additive signed measures. A concrete and very important example of this is given by the time averages of the Birkhoff ergodic theorem, which can be extended to a linear functional, as we next describe.

The important tool here is:

Theorem 3.4. *(Hahn-Banach) Let V be a real Banach space (a topological vector space with a norm). Let W be a subspace of V, and suppose that $\lambda : W \to \mathbb{R}$ is continuous linear; that is, $\lambda \in W^*$. Then there exists a continuous linear extension $\widetilde{\lambda}$ to V, that is, $\widetilde{\lambda} \in V^*$.*

Proof. We follow the proof of Theorem 3.2 in [Rud73].

□

Remark 3.3. The theorem is (with basically the same proof) valid for a locally convex complex TVS, see [Rud73] Theorem 3.6.

Example 8. Consider $\mathcal{C} = \mathcal{CB}(\mathbb{R}, \mathbb{R})$ with the sup norm. Let W denote the subspace such that

$$\lambda(f) = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(x)\mathrm{d}x$$

converges. By Theorem 3.4 there exists an extension $\widetilde{\lambda}$ to all of $\mathcal{C}$. One immediately verifies that $\widetilde{\lambda}$ is translation-invariant, positive and normalized.

We could instead have considered L^∞. Thus $\widetilde{\lambda}$ is an *invariant mean* on $\mathbb{R}$, see below.

This extension of the time average gives a reasonable version of the mean value of f. Note however that $\widetilde{\lambda}$ is *weighted at* $+\infty$; that is, it is unaffetced by any changes to f on $(-\infty, a]$ for any a. To have a more symmetric version we can define $\nu^+ = \widetilde{\lambda}$ and ν^- by $\nu^-(f) = \widetilde{\lambda}(f(-x)$, and then set $\nu = 1/2(\nu^+ + \nu^-$. Or, equivalently we could have begun with the symmetric time average $\lim \frac{1}{2T} \int_{-T}^T f\mathrm{d}m$.

The Hahn-Banach extension is in general highly non-unique. This is already showed by $\widetilde{\lambda} \neq \nu$. Further nonuiniqueness can be seen by replacing the usual (Cesáro) time averages by e.g. the log averages described below.

(To DO ???)

3.5. **Time averages and mean values on infinite groups.** Given a group G acting on a space, say a left action, a *time average* is in its essence an average value calculated along the group. Thus we want to first understand averages on the group, as a group orbit is an image of G or of a factor group of G. If the group is finite, then we just average over this finite set of values, and note that this corresponds to integration over the natural invariant probability measure on the group, and this idea extends to compact groups via the Haar measure, the unique left-action-invariant probability measure on the group.

For infinite groups this notion is replaced by that of an *invariant mean*, equivalently a *finitely* additive (left)-invariant probability measure. However new phenomena now appear, which make the subject both more problematic and more interesting. In particular, invariant means are nonunique; however in the nicest cases (finite measure spaces and ergodic transformations) this is not much of an issue as, by the Birkhoff Theorem, the usual Cesáro average works just fine.

A further complication arises in the case of certain infinite groups which are *non*amenable, by definition those for which no such invariant mean exists.

Nevertheless there is still a way to proceed! Then we can replace the notion of average *value* by that of an average *function*; the value depends on the point in the orbit we are looking out from. The result defines an operator from bounded functions, taking values in the class of harmonic functions. This *harnonic projection* then replaces the notion of time average.

Even in the case of space average, there can be ambiguities when the measure is infinite. In particular, the interaction between both nonamenable groups and infinite ergodic theory leads quickly to unexplored questions of active research.

To begin our discussion, we focus on the already interesting case of the simplest infinite group, the integers.

There the idea of time average brings us to an earlier theme in a different setting:

Example 9. *Choosing a point randomly from the integers and from the real line.*

Beginning with $\mathbb{R}$, by analogy with the above examples, we would like to find a translation-invariant probability measure on the real line. However this is impossible. Indeed, suppose that μ is a translation-invariant probability measure; since $\mathbb{R}$ is a countable union of intervals $[k, k+1)$, all translates of each other, each must have zero measure, giving a contradiction. What is hidden here is the use of countable additivity; indeed, if we are willing to throw out the word "countable" from our definition of measure, there may still be chance of finding something reasonable.

For this purpose, let us begin with the usual notion of time average just encountered.

A set $A \subseteq \mathbb{R}^+$ has **(Cesàro) density** $c \in [0,1]$ iff the Cesàro average of χ_A exists and equals c, i.e.

$$\lim_{T\to\infty} \frac{1}{T} m(A \cap [0,T]) \to c \text{ as } T \to \infty.$$

Let us denote by $\mathcal{A}$ the algebra of sets where the limit exists, defining a function $\mu : \mathcal{A} \to [0,1]$ by $\mu(A) = c$, the Cesàro density of A. This satisfies, for $A, B \in \mathcal{A}$ disjoint,

$$\mu(A \cup B) = \mu(A) + \mu(B),$$

$$\mu(A + t) = \mu(A),$$

and

$$\mu(\mathbb{R}) = 1.$$

That is, μ defines a *finitely* additive, translation-invariant probability measure on the algebra $\mathcal{A}$. Now in fact integration makes perfect sense also for finitely additive measures: first we define

$$\int_{\mathbb{R}} \chi_A(x) \mathrm{d}\mu(x) \equiv \mu(A) = \lim_{T\to\infty} \frac{1}{T} m(A \cap [0,T])$$

and then we extend this to all bounded $\mathcal{A}-$ measurable functions by linearity. We then denote

$$\lambda^+(f) \equiv \int_{\mathbb{R}} f(x) \mathrm{d}\mu(x) = \lim_{T\to\infty} \frac{1}{T} \int_0^T f \mathrm{d}m$$

when the limit exists. To make this symmetric in time, we define for $f : \mathbb{R} \to \mathbb{R}$ $\lambda^- f(t) = \lambda^+(f(-t)$ and then set

$$\lambda(f) \equiv (\lambda^+(f) + \lambda^-(f))/2 = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^T f \mathrm{d}m$$

when this exists. This is a *two-sided* Cesàro average.

We mention that this Cesàro average is just one of an enormous collection of possible *averaging methods* which have been studied in the literature. Other possibilities would be for instance to average over any sequence of intervals whose length goes to infinity.

To test some given choice of averaging method, we can consider various naturally occuring examples and see, first, whether or not the limit exists, and second, whether the answer agrees with our intuition as to what the average value should be, and third, whether the method exhibits various desirable symmetry or invariance properties.

A rich source of examples of such test functions is provided by dynamics; given an integrable function φ on a measure space acted on by a flow, we choose a point x in the space and then sample our function along the orbit of a flow, giving the test function $f(t) = \varphi(\tau_t(x))$.

The Birkhoff ergodic theorem then states exactly that our choice of averaging method works: for an ergodic flow on a probability space, for any such observable φ, for almost any choice of x, then the time average $\lambda^+(f)$ exists and equals the space average $\int \varphi$ of our observable, which is the value our intuition would have chosen.

So to test this method with more difficult functions, we have to move beyond dynamics, or at least beyond the dynamics of finite measure spaces.

We note first that there are other simple examples where the Cesáro average exists, occuring in harmonic analysis. The Cesáro average of any bounded periodic function on $\mathbb{R}$ exists, since the "tail effects" due to averaging over different long intervals vanish in the limit. And, moreover, the same logic works for any *almost* periodic function, such as $\sin(x) + \sin(\sqrt{2}x)$. See ??? below.

Intriguingly, any almost periodic function can be realized in a natural way from dynamics; see **??**?? , so this gives nothing new!

So perhaps we should look for examples where this limit does *not* exist, and then ask if there is some reasonable way to extend the definition of our measure to a wider class of sets, or functions? And then consider whether, *despite* the Birkhoff theorem, it might not be possible to find such functions arising in a natural way in a dynamical context?

First, there is an abstract general approach:

Definition 3.2. An **invariant mean** on $\mathbb{R}$ is a continuous positive normalized translation invariant linear functional on $L^\infty(\mathbb{R})$.

One way to define an invariant mean is to begin with a reasonable averaging method, such as the Cesáro average, defining that to be the mean value when it exists, and then extending to all bounded functions, via the Hahn-Banach theorem. This does give some notion of average value, as it is translation-invariant, yet this abstract approach is not completely satisfying as the value depends on the choice of intervals, and furthermore because the extension, and hence the mean value, is not uniquely determined.

Let us consider a concrete example of a function for which the Cesáro average does *not* exist: $f(x) = \sin(\log x)$. Here perhaps one should look for a different averaging method which will work.

Among the other possible averaging methods, also with nice symmetry properties, are the hierarchy of **log averages** : The Hardy-Riesz **log average** of f is

$$\lim_{T\to\infty} \frac{1}{\log T} \int_1^T f(x) \frac{1}{x} \mathrm{d}m;$$

the **log log** average is

$$\lim_{T\to\infty} \frac{1}{\log\log T}\int_e^T f(x)\frac{1}{x\log x}\mathrm{d}m,$$

and so on.

And indeed, these are all consistent in that if one exists, the limit for the next stronger method exists as well.

One can, moreover, show that there exists an invariant mean λ on $\mathbb{R}$ which is Cesáro- invariant and also is invariant with respect to an exponential change of scale; this means it is compatible with all the log averages! For an explanation, see Theorem 48.1 below.

Now the time average can be thought of as integration with respect to a *finitely* additive measure on $\mathbb{R}$ and the meaning of measure-linearity is that Fubini's theorem still holds, for the product of a countably additive measure with this special type of finitely additive measure (while in general that fails). In our setting, this means that the time-average and space- average can always be interchanged, which sounds just like Birkhoff's theorem!

Indeed, from this point of view, the content of Birkhoff's theorem is the following: that for functions arising from an integrable observable on an ergodic flow, one doesn't need the full power of such an invariant mean, as the Cesáro time average is sufficient.

We mention that the relationship between finite and countable additivity is explored in §33.1: on a compact space, a finitely additive measure is in fact countably additive. The difficulty and beauty of the Birkhoff theorem comes from the time average being taken over the noncompact group $\mathbb{R}$ (or $\mathbb{Z}$); see §33.1.

We mention that although the Cesáro average is sufficient for the classical situation of ergodic theory, as soon as infinite measures come into the picture, the log averages can indeed play a role; see §49.

4. Basic examples of dynamics

We have encountered in the last section the most basic independent system (the Bernoulli coin toss) and next we contrast this with some of the most basic non-independent systems. In fact a theme we find in this section (and throughout the notes) is that often these two types of behavior go hand-in -hand.

4.1. **Rotations of the circle and torus.** The $d-$**dimensional torus** is the additive topological group $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$. The one-dimensional torus $\mathbb{T}^1$ has another model, in complex notation: defining $S^1 = \{z \in \mathbb{C} : |z| = 1\}$, then S^1 is a subgroup of the multiplicative group $\mathbb{C}\setminus\{0\}$, and the map $\alpha : \mathbb{T}^1 \to S^1$ defined by $x \mapsto e^{2\pi i x}$ is a group isomorphism; similarly the torus $\mathbb{T}^d$ is isomorphic to $S^1 \times \cdots \times S^1 (d$ times).

If $x, y \in \mathbb{R}/\mathbb{Z}$ we write the group operation additively, as $x + y$, since the group is abelian. Another way of writing this $x + y (\mathrm{mod}\, 1)$ where x, y are real numbers; the word "mod" means "modulo the equivalence relation"; that is, addition is defined up to x being considered "equal" to y iff they are equivalent, $x \sim y$.

Now consider an irrational number θ. On $\mathbb{R}$ the translation map $R_\theta : x \mapsto x + \theta$ has a quite boring dynamics, but on the factor space $\mathbb{R}/\mathbb{Z}$ the induced map is much

more interesting. This is the map $x \mapsto x + \theta(\text{mod}1)$, an **irrational rotation on the circle**.

When we write the circle as S^1 this is a multiplicative group, and the corresponding irrational rotation map is $z \mapsto wz$ for $w = e^{2\pi i\theta}$.

Definition 4.1. A topological transformation (or flow) is **minimal** iff every orbit is dense.

It is **uniquely ergodic** if there is a unique invariant probability measure.

Exercise 4.1. Show that R_θ is minimal if and only if it is an **irrational rotation**, i.e. $\theta \neq \mathbb{Q}$.

Show that with respect to normalized Lebesgue measure μ, R_θ is ergodic iff θ is irrational.

Show that in fact it is in this case uniquely ergodic.

Let $\mathbf{v} = (a, b) \in \mathbb{R}^2$, and define $\tau_{\mathbf{v},t} : \mathbb{T}^2 \to \mathbb{T}^2$ by $\tau_{\mathbf{v},t} : (x, y) \to (x, y) + t(a, b)$. This is the **rotation flow** on the torus of velocity $\mathbf{v}$. The time-one map $R_\mathbf{v} = \tau_{\mathbf{v},1}$ is a **rotation of the torus**. We say this is an **irrational** rotation if one (hence every) orbit is dense.

Exercise 4.1. *Show that $\tau_{\mathbf{v},t}$ is a minimal flow iff it is an* **irrational** *rotation flow, i.e. $b/a \notin \mathbb{Q}$. Show this is false in general for $R_\mathbf{v}$, and find a condition on a, b which is equivalent to minimality of this transformation. Extend these results to the $d-$torus.*

Now we recall:

Definition 4.2. A **relation** from a set X to a set Y is a subset of the product space, $R \subseteq X \times Y$; one writes xRy iff $(x, y) \in R$, read "x is related to y". An **equivalence relation** is a relation on X (i.e. $R \subseteq X \times X$) which satisfies:
(i) (symmetry) xRx
(ii) (reflexivity) $xRy \implies yRx$
(ii) (transitivity) xRy and $yRz \implies xRz$.

A **partition** $\mathcal{P}$ of a set X is an indexed collection $\mathcal{P} = \{P_i\}_{i\in I}$ of subsets of X such that:
$(1) P_i \cap P_j = \emptyset$ iff $i \neq j$
$(2) \cup_{i\in I} P_i = X$.

A **fundamental domain** for an equivalence relation is a subset A of X which contains exactly one point from each equivalence class. We shall say a fundamental domain is *nice* if it is, in the topological category, well-behaved topologically in that it is locally homeomorphic to X; if X is connected, one may require A to share this property as well. In the measure-theoretic category, a nice fundamental domain should be a measurable subset. An example is the torus, where the group $\mathbb{R}^d/\mathbb{Z}^d$ is given the quotient topology; a fundamental domain is the product of d half-open intervals, $[0, 1) \times \cdots \times [0, 1)$. This verifies what was stated above, that $\mathbb{T}^d$ is the product of d circles: $\mathbb{T} \times \cdots \times \mathbb{T}$.

An important example is given by the orbit equivalence relation of a group action, defined by its name: $x \sim y$ iff $x \in \mathcal{O}(y)$.

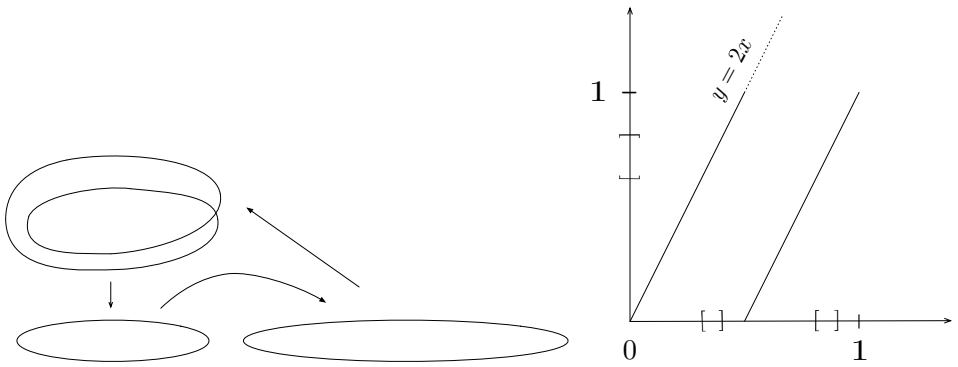


FIGURE 5. Doubling map on circle and interval; measure is preserved: the inverse image of an interval has two pieces, each half its size.

Exercise 4.2. Verify that the notions of equivalence relation and partition are equivalent (what are the equivalence classes?) Verify that group orbits do partition the space.

Exercise 4.2. *On the topological space $\mathbb{T}^d$, $d \geq 1$, consider a rotation transformation, and for $d \geq 2$ a flow. Show the quotient topological space $\mathbb{T}^d/\sim$ is a non-Hausdorff space iff this dynamical system is minimal.*

4.2. **Doubling map on the circle and interval.** Define a map $T : \mathbb{T}^1 \to \mathbb{T}^1$ by $T : x \mapsto 2x (\mathrm{mod}\, 1)$. Define $f : S^1 \to S^1$ by $f : z \mapsto z^2$.

Exercise 4.3. *For α as defined above, show the following diagram commutes:*

$$\begin{array}{ccc} \mathbb{T} & \xrightarrow{T} & \mathbb{T} \\ \downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle \alpha} \\ S^1 & \xrightarrow{f} & S^1 \end{array}$$

Define $g : I \to I$ for $I = [0, 1]$ by $g(x) = 2x - [2x]$ where $[a]$ denotes the greatest integer $\leq a$ (the **integer part of** a), so

$$g(x) = \begin{cases} 2x & \text{for } x \in [0, 1/2) \\ 2x - 1 & \text{for } x \in [1/2, 1]. \end{cases}$$

(One often sees this map written as $g(x) = 2x (\mathrm{mod}\, 1)$, though strictly speaking that isn't correct, as $0 = 1 (\mathrm{mod}\, 1)$.)

The three maps T, f, g are called **doubling maps** (of the circle, and the interval). On the circle one can imagine streching out a rubber band to twice its length, doubling it over and then projecting, see Fig. 5. Note that while Lebesgue measure is not preserved locally in the forward direction (it is doubled!) it is preserved by the inverse map, and so fits the definition of invariant measure.

Exercise 4.3. Prove this last statement.

4.3. **Shift maps.** Given a finite set $\mathcal{A}$, called the **alphabet**, define $\Sigma = \Pi_{-\infty}^{\infty}\mathcal{A}$, $\Sigma^+ = \Pi_0^{\infty}\mathcal{A}$ and $\Sigma^- = \Pi_{-\infty}^{-1}\mathcal{A}$, so $\Sigma = \Sigma^+ \times \Sigma^-$.

A point in Σ will be written as $(\dots ab.cde \dots)$ where the "decimal point" serves to locate the 0^{th} coordinate, c in this case.

A map $\sigma : \Sigma \to \Sigma$ is defined, for $\underline{x} = (\dots x_{-1}.x_0x_1 \dots)$, by $\sigma(\underline{x}) = (\dots x_{-1}x_0.x_1 \dots)$.

The transformation (Σ, σ) is called the **(left) shift on k symbols**, where $k = \#\mathcal{A}$. This is also known as the **bilateral** or **two-sided** shift. We define a map, also denoted σ, on Σ^+ by $\sigma(.x_0x_1 \dots) = (.x_1x_2 \dots)$. This is the **one-sided shift**. We call x_0 the **present** coordinate of $\underline{x}$; the coordinates $x_0, x_1 \dots$ are the **future** and $\dots x_{-2}, x_{-1}$ the **past**. Given a point $\underline{x} \in \Sigma$, we write $\underline{x}^+ = (.x_0x_1 \dots) \in \Sigma^+$ and $\underline{x}^- = (\dots x_2x_1.) \in \Sigma^-$.

We give $\mathcal{A}$ the discrete topology and Σ the corresponding product topology. Note that Σ is compact, by Tychonoff's product theorem; an example of a compatible metric is $d(\underline{x}, \underline{y}) = 1$ if $x_0 \neq y_0$, otherwise $d(\underline{x}, \underline{y}) = 2^{-i}$ where i is $\inf\{|k| : x_k \neq y_k\}$.

This space itself is certainly not discrete, as:

Exercise 4.4. *Show that the product $\Pi_{i\in I}X_i$ of discrete topological spaces $(X_i, \mathcal{T}_i)$ is discrete if and only if the index set I is finite.*

A **word** is a finite sequence of symbols from the alphabet, e.g. $x_0x_1 \dots x_n$. A **string** is a one- or two-sided infinite such sequence, $.x_0x_1 \dots$ or $\dots x_{-1}.x_0x_1 \dots$ (so strings are the points in Σ^+, Σ.) Following Billingsley [Bil65], we define a **thin cylinder set** for $l \leq n$ to be a subset of Σ of the form $[a_l \dots a_n] = \{\underline{x} \in \Sigma^+ : x_l = a_l, \dots, x_n = a_n\}$, thus all the strings which begin with a given word. A **general cylinder set** is a finite intersection of thin cylinders. We use the symbol "$*$" to denote "any symbol", so the cylinder set $[a_0a_1] \cap [a_3a_4]$ can be written as $[a_0a_1 * a_3a_4]$. We also make use of the decimal point here, writing e.g. $[ba. * *cb]$. The reason for the name **cylinder** is that these are cylinders in the infinite product space $\Pi\mathcal{A}$; indicating time by a subscript; thus $[.ab] = \dots \mathcal{A}_{-2} \times \mathcal{A}_{-1} \times \{a\}_0 \times \{b\}_1 \times \mathcal{A}_2 \times \dots$.

Exercise 4.5. *Show that cylinder sets form a subbase for the topology on Σ, and that these are clopen (both closed and open) sets.*

Show that the map $\pi : \underline{x} \mapsto \underline{x}^+$ is continuous, and identify the fiber over $\underline{x}^+$ (its inverse image). Check that the following diagram commutes, semiconjugating this invertible map with the everywhere $k-$to-one map on the one-sided space. More generally, this works for subshifts: let $\Omega \subseteq \Sigma$ be a closed invariant subset; then the bilateral shift (Ω, σ) factors onto the unilateral shift (Ω^+, σ):

$$\begin{array}{ccc} \Sigma & \xrightarrow{\sigma} & \Sigma \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\ \Sigma^+ & \xrightarrow{\sigma} & \Sigma^+ \end{array} \qquad \begin{array}{ccc} \Omega & \xrightarrow{\sigma} & \Omega \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\ \Omega^+ & \xrightarrow{\sigma} & \Omega^+ \end{array}.$$

The most basic case is the **two-shift**, where $\mathcal{A}$ has two letters, usually labelled 0 and 1, so the one-sided shift is $\Sigma^+ = \Pi_0^{\infty}\{0, 1\}$. We claim that we can visualize this geometrically as the doubling map on the interval.

Exercise 4.6. *Defining a map* $\pi : \Sigma^+ \to I$ *by*

$$\pi(\underline{x}) = \sum_{i=0}^{+\infty} x_i 2^{-(i+1)}, \tag{1}$$

show the following diagram commutes.

$$\begin{array}{ccc} \Sigma^+ & \xrightarrow{\sigma} & \Sigma^+ \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\ I & \xrightarrow{g} & I \end{array} \tag{2}$$

Show that π *is one-to-one except at a countable set of points, where the map is two-to one.*

Show that the $(1/2, 1/2)$ *infinite product measure* μ *on* Σ^+ *is preserved by that shift map, and that it pushes forward via* π *to Lebesgue measure on* I*.*

So the inverse of the map assigns the binary expansion $\underline{x}$ to a point $x \in I$; this is uniquely defined except at a countable set where there are two possible expansions.

Thus the doubling maps on T and on I provide geometric models for the shift map. Including the measures, since infinite product measure models the infinite toss of a fair coin, the doubling maps give geometric, "deterministic" models for the thoroughly random process of coin-tossing.

If we replace the above coin-tossing measure on Σ^+ by infinite product measure ν where the symbols $0, 1$ are given probabilities p, q for $p, q > 0$ with $p+q = 1$, then the measure is again shift-invariant, and now models tosses of an *unfair* coin. To study tosses of a fair die we would use an alphabet with six letters, $1, 2, \ldots 6$, each with probability $1/6$.

4.4. **Symbolic dynamics.** We have just seen how independent coin-tosses provide a shift map with an invariant measure which models the doubling map. The homomorphism $\pi : \Sigma^+ \to I$ was defined by the arithmetic expression of writing a point in the binary expansion of (1).

Symbolic dynamics can be viewed as the converse procedure, that of starting with a dynamical system defined in some other way, e.g. geometrically or algebraically and associating with it the stationary process of a shift map; sometimes a converse arithmetic expansion for a point is possible, with the symbols serving as the digits, but the general idea goes far beyond that most ideal situation. What is achieved thereby is that the geometric and dynamical point of view ergodic theory is brought into probability theory, and conversely.

Given a map $T : X \to X$ and a partition $\mathcal{P}$ with a countable (i.e. finite or countably infinite) index set $\mathcal{A}$ (called as above the **alphabet**), then we consider for T invertible the shift space $\Pi \equiv \Pi_{i\in\mathbb{Z}}\mathcal{A}$, for the noninvertible case $\Pi^+ \equiv \Pi_{i\in\mathbb{N}}\mathcal{A}$, with the left shift map σ. Then there is a map from X to Π defined by $x \mapsto \underline{x} = (\ldots x_0 x_1 \ldots)$ where $x_0 = a$ iff $x \in P_a$. By definition, the diagram commutes:

$$\begin{array}{ccc} X & \xrightarrow{T} & X \\ \downarrow{\scriptstyle \varphi} & & \downarrow{\scriptstyle \varphi} \\ \Pi & \xrightarrow{\sigma} & \Pi \end{array}$$

and similarly for the noninvertible case.

One says that the partition **generates** iff it **separates points** in the sense that for $x \neq y$, there exists k such that $T^k x, T^k y$ are in different partition elements. It is clear that this happens exactly when φ is injective. If the partition doesn't generate, then we have a factor map to the image $\alpha(X) \subseteq \Pi$.

Any partition which generates gives in this simple way a combinatorial representation of the dynamical system. This has some clear advantages:
- the shift dynamics is very simple, in particular we know exactly where $\underline{x}$ will be at time k;
-since the index set $\mathcal{A}$ is discrete, its points (called **symbols** or **letters**, or **digits** if they are natural numbers) can be treated analogously to letters of words in a language, and one can bring in ideas from linguistics, coding theory, and information theory; in particular, a **code** can be considered to be a map between two symbolic dynamical systems, with the infinite **string** of letters $\underline{x}$ representing an infinitely long message.

Moreover, given an invariant measure for $(X, \mathcal{A}, \mu, T)$ and a partition $\mathcal{P}$ consisting of measurable sets, the measure μ pushes forward to an invariant measure $\widetilde{\mu}$ on the shift space.

(A side remark is that there must indeed be *many* invariant measures on Π, as *any* dynamical system can be modelled in this way. Another way of looking at this is: *any* system can be modelled by coin-tosses- where the tosses are stationary but in general very far from independent).

However, there are some problems with this approach:
-*it is too general*: unless the partition is chosen with care, to reflect geometrical, arithmetic, algebraic or dynamical properties of the system, it may be next to useless; indeed, by a theorem of Krieger, any finite entropy ergodic measure-preserving transformation has a finite generating partition, and any infinite entropy map a countable generating partition. When the invariant measure is transported from the original map to the shift space, that means we have a measure-theoretically isomorphic model for *any* measure-preserving ergodic map.
-*in choosing a symbolic model we may lose a great deal of information*: for instance, topological information: if T is a continuous map of a topological space, then the best we can do may be to have a partition into clopen sets which form not a partition but a **partition mod zero**, that is, after throwing away a null set; the partition boundaries are thus ripping apart the space, (an example is given by the baker's transformation, which we encounter shortly) and points there have an ambigous **name** (the string of symbols).

In this case, if $\mathcal{P}$ generates, then it is more natural to draw the diagram in the opposite way, with the shift space on top, as we may then have a topological factor map from a set which is topologically disconnected (the shift space) to the space X,

with partition elements being glued along their boundaries to form X; if the identifications of symbolc sequences can be nicely specified in a given case, the symbolic model is more useful. This is exactly what happens in Figure 2 above.

How to choose such a partition appropriately, and to understand which properties will or will not be preserved, is, then, the real work of symbolic dynamics, and will be a recurrent theme in these notes.

If we choose such a nice partition, one speaks of the resulting measure-preserving homomorphism to the shift space as giving a nice way of "coding" the original dynamical system.

???

We saw above how symbolic dynamics gives a semicongugacy from the shift map to the the doubling map on the interval. Next we examine a dynamical model which is conjugate to the shift.

4.5. **The Cantor set.** Recall that the (middle-third) Cantor set C is constructed by removing succesive open middle third intervals from I. Thus, writing $C_0 = I$, $C_1 = I_0 \cup I_1$ where $I_0 = [0, 1/3]$ and $I_1 = [2/3, 1]$ and $C_2 = I_{00} \cup I_{01} \cup I_{10} \cup I_{11}$ where I_{00}, I_{01} are the left and right closed thirds of I_0 and similarly for I_1. Continuing in this manner, the Cantor set is, by definition,

$$C = \cap_{n=0}^{+\infty} C_n.$$

This is a compact set which is nowhere dense (so it contains no open intervals) and is dense in itself (there are no isolated points, i.e. every point is a limit point of other points in C). It has Lebesgue measure zero, since $m(C_{n+1}) = 2/3m(C_n)$ for all n. Furthermore it has the cardinality of the continuum: each point in C is the endpoint of an infinite binary tree, branching to the left or right at level n depending on which subinterval the point belongs to; the set of infinite branches corresponds to $\Pi_0^\infty\{0, 1\}$, which has cardinality of the continuum.

The **tripling map** on the interval is defined by $h(x) = 3x (\mathrm{mod}\, 1)$. Exactly as for the doubling map, there is a projection from the shift space $\Sigma_3^+ = \Pi_0^\infty\{0, 1, 2\}$ to I given by the **ternary** (or base three) expansion, $\pi(\underline{a}) = a = \sum_{i=0}^{+\infty} a_i 3^{-(i+1)}$ for $a_1 = 0, 1$ or 2.

This defines a dynamics on the Cantor set: writing each point $x \in I$ in ternary expansion $x = \sum_{i=0}^{+\infty} a_i 3^{-(i+1)}$ for $a_i \in \{0, 1, 2\}$, then each point in $I_0 \cup I_1$ has expansion $\underline{a} = (.a_0 a_1 \dots)$ with the restriction $a_0 \in \{0, 2\}$; for C_2 we have a_0 and $a_1 \in \{0, 2\}$ and so on. Therefore, C is the collection of points $a = \sum_{i=0}^{+\infty} a_i 3^{-(i+1)}$ for $a_1 = 0$ or 2. Let us write Σ_2^+ for $\Pi_0^\infty\{0, 1\}$ and Σ_3^+ for $\Pi_0^\infty\{0, 1, 2\}$.

Next we define a function from C to I which sends a point in ternary expansion to binary expansion with the same $0 - 1$ digits.

Thus, defining $\alpha : \Sigma_2^+ \to I$ by $\alpha(\underline{x}) = \sum_{i=0}^{+\infty} 2x_i 3^{-(i+1)}$, and $\beta : C \to I$ by $\beta : \sum_{i=0}^{+\infty} 2x_i 3^{-(i+1)} \mapsto \sum_{i=0}^{+\infty} x_i 2^{-(i+1)}$, then β gives a map from C to I which is bijective except at countably many points. These points are exactly the interior endpoints of the subintervals $I_0, I_1; I_{00}, I_{01} \dots$, which get glued together by the map β to form the continuum I. We have the following commutative diagram, where g is the doubling map on I.

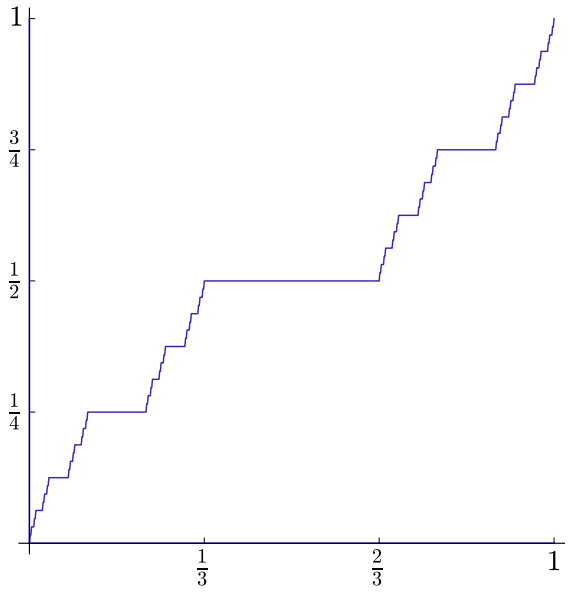


FIGURE 6. The conjugacy β extends to the Cantor function $\widehat{\beta}$: the distribution function of the Cantor measure.

$$\begin{array}{ccc} \Sigma^+ & \xrightarrow{\sigma} & \Sigma^+ \\ \downarrow \alpha & & \downarrow \alpha \\ C & \xrightarrow{h} & C \\ \downarrow \beta & & \downarrow \beta \\ I & \xrightarrow{g} & I \end{array} .$$

Exercise 4.7. *Show that α is a homeomorphism and that β is continuous.*

The function β has a unique nondecreasing extension to I, giving a continuous function $\widehat{\beta}$ from I onto I which is flat on the gaps of the Cantor set. This is the **Cantor function**, see Fig. 6.

We have given two definitions of C, by removing middle thirds or by expressing the points in ternary expansion with no 1's. Here is a third definition, which, since it is dynamical rather than combinatorial or geometric, leads to interesting generalizations.

First we need:

Definition 4.3. Let $f : X_0 \to X$ for $X_0 \subseteq X$. The **eventual domain** of f is $X_\infty^- = \cap_{k=1}^{+\infty} f^{-k}(X)$.

Proposition 4.1. *The eventual domain is the largest subset of X such that $f(x)$ is defined for all $k \geq 0$.*

Proof. Since $f^n(x) \in X \iff x \in f^{-n}(X)$, the statement follows.

□

We now consider the map $f : C_1 = I_0 \cup I_1 \to I$ defined to be the restriction of the tripling map to this set, except at $1/3$ where we define $f(1/3) = 1$ (in order to make f continuous).

Then the Cantor set C is the eventual domain of this map, since $f^{-1}(I) = I_0 \cup I_1 = C_1$, $f^{-2}(I) = f^{-1}(I_0 \cup I_1) = I_{00} \cup I_{01} \cup I_{10} \cup I_{11} = C_2$, and so on.

Now we use the same idea to define a much wider class of Cantor sets. Let I_0, I_1 be two disjoint closed subintervals of I, and let f_i be a $C^{k+\alpha}$ diffeomorphism from I_i onto I. Write C_f for the eventual domain of this map. Again there is a homeomorphism α from Σ_2^+ onto C_f; if f is a nonlinear map then C_f is called a **hyperbolic** $C^{k+\alpha}$ **Cantor set** (or **cookie-cutter set**), and the sequence of 0's and 1's which is the itinerary of the point $x \in C_f$ with respect to the sets I_0, I_1 gives a nonlinear analogue of the ternary expansion. See Fig. ?? and e.g. [Sul87], [BF97].

Returning to the middle-thirds Cantor set C, we summarize some of its basic properties:

–C is a compact set which is nowhere dense and dense in itself (i.e , every point $x \in C$ is a limit point of $C \setminus x$). In particular it is a **perfect set** (closed and dense in itself).

–C is an exactly self-similar set: at each scale, it is a union of small pieces which are exact replicas of C; precisely, at scale 3^{-n} it consists of 2^n pieces $C \cap I_{x_0 \ldots x_{n-1}}$ such that $3^n \cdot = C$. It has Hausdorff dimension $d = \log 2 / \log 3$.

–measure-theoretically, the Bernoulli $(1/2), (1/2)$ coin-tossing measure on Σ^+ pushes forward to a probability measure μ on C which turns out to be $H_d|_C$, the Hausdorff measure H_d of dimensionrestricted to C.

–The Cantor function can therefore be understood as the cumulative distibution function of the measure μ, since $\widehat{\beta}(x) = \mu([0, x]$.

Since we have just talked about the eventual domain of a map, this is a good place to introduce a related idea which we need later on:

Definition 4.4. Let $f : X \to X$. The **eventual range** of f is $X_\infty^+ = \cap_{k=0}^{+\infty} f^k(X)$.

Proposition 4.2. *Let X be a metric space and f a continuous map such that the closure of $f(X)$ is compact. Then the eventual range is the largest subset A such that f maps A onto itself.*

For the proof we first recall that a topological space has the **Bolzano-Weirstrass property** iff every sequence has an accumulation point. It is **sequentially compact** iff every sequence has a convergent subsequence. The Bolzano-Weirstrass property implies compactness; for metric spaces sequential compactness is equivalent to compactness and also to separability [Roy68].

Lemma 4.3. *Let $f : X \to X$ be a continuous map on a topological space X, and let K be a compact subset of X which has the Bolzano-Weirstrass property. If $f(K) \subseteq K$, then for $K_n = f^n(K)$ and $K_\infty = \cap_{i=0}^{\infty} K_i$, we have that $K = K_0 \supseteq K_1 \supseteq K_2 \ldots$ and that $f(K_\infty) = K_\infty$.*

Proof. Since $A \subseteq B \implies f(A) \subseteq f(B)$, the nesting of the K_i follows by applying induction to the containment $f(K) \subseteq K$. This also implies that $f(A \cap B) \subseteq f(A) \cap f(B)$, and similarly for infinite intersections, whence $f(K_\infty) \subseteq K_\infty$. To show this

is onto, we give the proof given sequential compactness; this is easily modified to work assuming the Bolzano-Weirstrass property. Let $x \in K_\infty$; we claim there exists $w \in K_\infty$ with $f(w) = x$. Since $x \in K_{i+1}$ for all i, there exists $w_i \in K_i$ with $f(w_i) = x$. By compactness there exists a subsequence w_{i_k} and point w with $w_{i_k} \to w$; $w \in K_{i_k}$ for each k, so $w \in K_\infty$, and by continuity, $f(w) = x$. □

Proof. (of Proposition) From the lemma it follows that $f(X_\infty^+) = X_\infty^+$, and this is clearly the largest such subset. □

This idea is closely related to the following:

Definition 4.5. If W is a sequentially compact topological space with $T : W \to W$ continuous, then given $\mathcal{U}$ open such that the closure of $f(\mathcal{U})$ is contained in $\mathcal{U}$, one calls $C \equiv \cap_{k=0}^{+\infty} f^k(U)$ an **attractor** of the map. The **basin** of the attactor is $\cup_{n\geq 0} f^{-n}(\mathcal{U})$.

Proposition 4.4. *For an attractor C of $T : W \to W$ as above, then $f(C) = C$ and for any x in the basin of C, $f^n(x)$ converges to C in the sense that for any open set $\mathcal{V} \supseteq C$, this orbit is eventually inside of $\mathcal{V}$. The basin is the collection of all points in W which converge to C, hence does not not depend on the choice of open set $\mathcal{U}$ with attractor C.*

Proof. Setting $X = \mathcal{U}$, with the relative topology, we are in the situation of Proposition 4.2, with C the eventual range. □

Remark 4.1. For the definition of attractor we are following [BS02]. We note that by the same proof, in Proposition 4.2 we can conclude that every $x \in X$ is attracted to the eventual range. Proposition 4.4 is stated, for compact topological spaces, in the discussion at the beginning of §1.13 of [BS02]. However we warn that the argument given there that $f(C) = C$ seems to us to be incomplete: reasoning like that above, using the compactness, (and perhaps as we did, the Bolzano-Weirstrass property) is needed.

Exercise 4.4. Investigate whether a counterexample can be found (for a space which is compact but not sequentially compact).

4.6. **The baker's transformation.** This map, written with a small "b" as the baker is someone who is kneading bread, is defined on the half-open square, $X = [0, 1) \times [0, 1)$, Despite the fact that it is not everywhere continuous, it gives a good model for what is happening geometrically with the full (two-sided) shift map, and with hyperbolic (or "chaotic") dynamics in general. This map is also sometimes known as *Arnold's cat map* (evidently because of the illustration on p. 9 of [AA68]) although it appears on p. 9 of Halmos' book of 1956 [Hal60] (but it must predate not only Arnold but Halmos).

We define $F_1 : \mathbb{R}^2 \to \mathbb{R}^2$ by $F_1(x, y) = (2x, 1/2y)$; writing $X_1 = F(X)$, then we define $F_2 : X_1 \to X$ by

$$F_2(x, y) = \begin{cases} (x, y) & \text{for } x \in [0, 1/2) \\ (x, y) + (-1, 1/2) & \text{for } x \in [1/2, 1) \end{cases}$$

and then $F : X \to X$ by $F = F_2 \circ F_1$. See Fig. 7. The map F is a bijection of X which is continuous off of the line segment $x = 1/2$, while F^{-1} is continuous off of

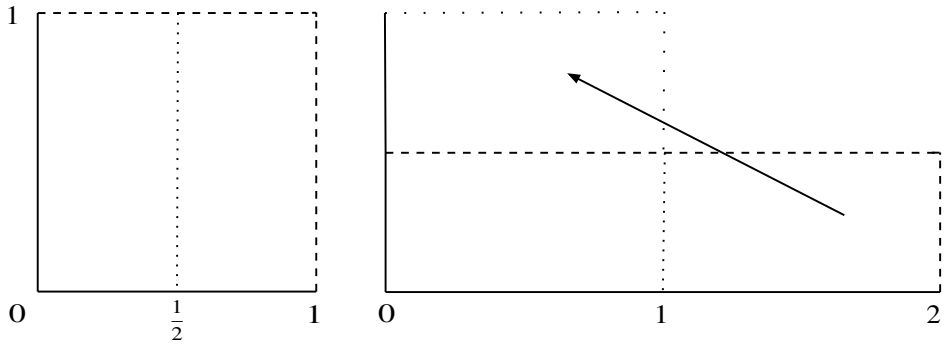


FIGURE 7. The baker's transformation, showing $\mathcal{P}$ and $T\mathcal{P}$.

$y = 1/2$. So it is a homeomorphism off of the union of those two segments. Note that Lebesgue measure on the square is preserved by F.

Exercise 4.8. *Defining the map* $\pi_+ : \Sigma^+ \to [0,1]$ *by* $\pi_+(\underline{x}^+) = \sum_{i=0}^{+\infty} x_i 2^{-(i+1)}$ *and* $\pi_- : \Sigma^- \to [0,1]$ *by* $\pi_-(\underline{x}^-) = \sum_{i=-1}^{-\infty} x_i 2^i$*, and finally* $\pi : \Sigma \to [0,1] \times [0,1]$ *by*

$$\pi(\underline{x}) = (\pi_+(\underline{x}^+), \pi_-(\underline{x}^-)),$$

and setting $E = \{\underline{x} : \underline{x}^+ = .11111\ldots$ *or* $\underline{x}^- = \ldots 11111.\}$*, show that the following diagram commutes, with* $(\frac{1}{2}, \frac{1}{2})$*- infinite product measure on the shift space finite taken by* π *to Lebesgue measure on the square. Find the smallest set of points* $\mathcal{N} \subseteq \Sigma$ *such that the restriction of* π *gives a topological conjugacy of the two dynamical systems. (Remember: by our definition a dynamical system is required to be an* onto *map).*

$$\begin{array}{ccc} \Sigma & \xrightarrow{\sigma} & \Sigma \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\ X & \xrightarrow{T} & X \end{array}$$

Defining a partition $\mathcal{P} = \{P_0, P_1\}$ *of* X *where* $P_0 = \{(x,y) : x \in [0, 1/2)\}$ *and* $P_1 = \{(x,y) : x \in [1/2, 1)\}$*, then show that the digits of* $\underline{x} \in \Sigma \setminus E$ *satisfy* $x_i = j$ *if and only if* $T^i(x) \in P_j$*. Draw* $T^{-1}(\mathcal{P})$ *and* $\bigvee_{i=-2}^{1} T^{-i}(\mathcal{P})$*.*

Exercise 4.9. *Show that for* I *the unit interval with Lebesgue measure* m*,* (I, m) *and* $(I \times I, m \times m)$ *are measure-isomorphic. (*Hint*: coding.)*

4.7. **The odometer transformation.** We define a second map on the one-sided shift space $\Sigma^+ = \Pi_0^\infty \mathcal{A}$ which will in a certain sense be *transverse* to the dynamics of the shift map σ. For $x = (.x_0 x_1 \ldots) \in \Sigma$, we map x as follows, illustrated by example: $(.000\ldots) \mapsto (.100\ldots) \mapsto (.010\ldots) \mapsto (.110\ldots) \mapsto (.0010\ldots) \mapsto (.1010\ldots) \mapsto (.0110\ldots) \mapsto (.1110\ldots) \mapsto \ldots$

This is like watching the odometer of a car, the device which measures miles (or kilometers!) travelled, except written in binary and in reverse.

Note that if the 0 coordinate turns over every second, the 1 coordinate turns over every 2 seconds, the 2 coordinate every 4 seconds, then every 8, 16 seconds and so on.

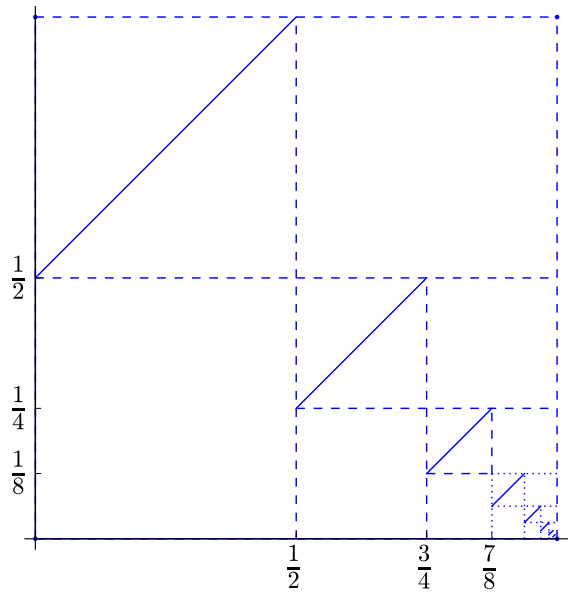


FIGURE 8. Graph of the odometer transformation on the unit interval: an exchange of infinitely many intervals.

From a different point of view we proceed in anti-lexicographic order, where the letters are from the alphabet $\mathcal{A} = \{0, 1\}$ with order $0 < 1$. Given two infinite strings x, y then supposing they are in the same stable set for the shift map σ, thus there exists n such that $x_k = y_k$ for every $k \geq n$, then $x < y$ iff $x_{n-1} < y_{n-1}$.

Then T sends x to its successor in this order. There is one point where the map is not defined: $x = (.111\ldots)$. In this case, it is natural to define the image to be (.000...) the unique point with no *pre*image. As one checks, this is the unique way to extend the map continuously to all of Σ^+. That's like the odometer in your car turning over to 0 after it reaches $99,999$!

This defines the Kakutani-von Neumann **dyadic odometer**. Another name for this is the *adding machine* transformation, since we succesively adding 1 in binary (written in reverse).

Algebraically, Σ^+ is an abelian group with respect to addition, and the map T is a rotation in this group. We have already encountered a group rotation: irrational rotation R_θ on the circle. We mention that just as for this map, the odometer is also minimal and uniquely ergodic (exercise: verify this!)

Representing Σ^+ as the unit inteval via binary expansion, T has the graph given in Fig. 8, which shows it to preserve Lebesgue measure and to be an exchange of countably many intervals (Here we have to remove a set of measure zero, the countably many points in the interval with two binary expansions). Note that restricting to cylinder sets of finite length, T simply permutes them, e.g. $[.000] \mapsto [.100] \mapsto [.010] \mapsto [.110] \mapsto [.001] \mapsto [.101] \mapsto [.011] \mapsto [.111] \mapsto [.000]$. Thus T is a limit of finite permutations of intervals, see Fig.9.

The maps σ and T of Σ^+ are linked by an interesting *commutation relation*:

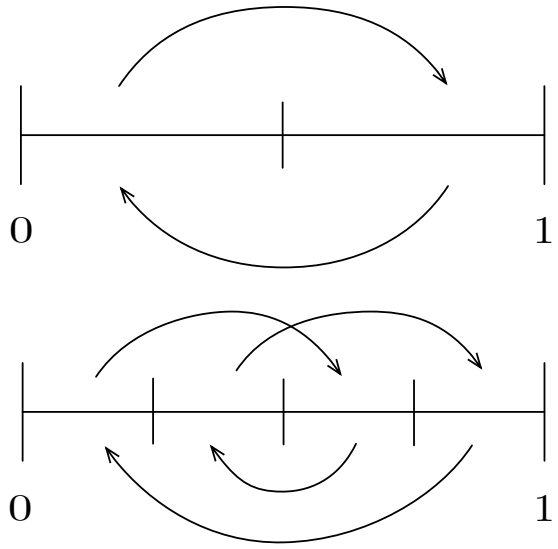


FIGURE 9. Odometer transformation as a limit of finite interval exchanges

Proposition 4.5. *T, σ satisfy*

$$\sigma \circ T^2 = T \circ \sigma.$$

Thus, the following diagram is commutative:

$$\begin{array}{ccc} \Sigma^+ & \xrightarrow{T^2} & \Sigma^+ \\ \downarrow{\scriptstyle \sigma} & & \downarrow{\scriptstyle \sigma} \\ \Sigma^+ & \xrightarrow{T} & \Sigma^+ \end{array}$$

The odometer map on the interval has a different construction, as a cutting-and-stacking construction.

We explain this by picture: beginning with the unit interval I with Lebesgue measure, we dvide it into two halves, I_0, I_1 corresponding to the cylinder sets $[.0], [.1]$. Stacking the second on top of the first, we define the map T to go upward. This is the first stage of the definition; not that it is equivalent to the exchange of two intervals.

In the second stage, we cut the tower into two halves, and stack the right on top of the left.

There are several simple yet important observations:
–at stage n, the map is defined everywhere except at the top of the tower, and its inverse is defined everywhere except at the bottom;
–once defined, the definition never changes;
–Lebesgue measure is preserved, where the map is defined, both forwards and backwards;
–the measure of the set of points where T or T^{-1} are defined goes to one.
–at each stage n, the tower definition is equivalent to the exchange of 2^n intervals, hence in the limit does give the odometer map.

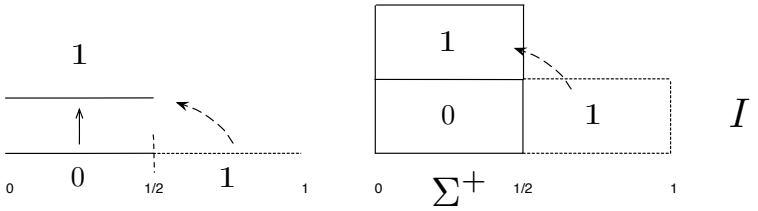


FIGURE 10. Cutting-and-stacking constructions of the odometer map and of the solenoid and baker's transformation.

The importance of this construction is that it generalizes considerably, permitting the definition of a wide variety of interesting maps. We examine other cases below.

4.8. **The solenoid.** The cutting-and-stacking constuction for the odometer should be reminiscent of the definition above of the baker's transformation. This is not accidental; the link between the two can be explained via a further example, the *(dyadic) solenoid*. This is a topological space which is locally a product of a Cantor set and an interval; just as for the odometer, it posesses two types of dynamics, one transverse to the other and satisfying a similar commutation relation. And in fact, the space can be built in two ways: based on extending the doubling map of the interval, or based on extending the odometer map. The first of these will give a hyperbolic map corresponding to the shift, while the odometer map will now be replaced by a *flow*: the *rotation flow* on the solenoid. We explain, first considering the doubling map on the circle $S^1 = \{z \in \mathbb{C} : |z| = 1$, the map $f : z \mapsto z^2$.

This already has two types of dynamics: in addition to f there is the *rotation flow* h_t, defined additively (i.e. on $\mathbb{T} = \mathbb{R}/\mathbb{Z}$) by $R_t : x \mapsto x + t (\mathrm{mod}\, 1)$ or multiplicatively (on S^1) by $M_t : z \mapsto e^{2\pi i t} \cdot z$.

The map f is of course not invertible (it is $2-1$) but we can remedy this by creating the *natural extension* $\widehat{f} : \widehat{S} \to \widehat{S}$; the construction is treated in generality below, but here we give the basic idea. Given $\mathbf{z}_0 \in S^1$, we choose an infinite string of preimages, $\dots z_n \mapsto z_{n+1} \mapsto \dots z_{-1} \mapsto z_0$. This can be continued uniquely to the future as $z_0 \mapsto z_1 \mapsto \dots$, giving a biinfinite string $\underline{z} = (\dots z_{-1}.z_0 z_1 \dots) \in \Pi_{-\infty}^{\infty} S^1$. The reader may recognize this as an *inverse limit* space; with the natural (inverse limit) topology, this defines the topological space $\widehat{S}$ called the solenoid.

One can picture this as follows. ...

(quote two-sided preprint!!!) rotation flow/ doubling map

Natural extension of $z \mapsto z^2$; of shift map, as limits of covers.

4.9. **A Cantor set baker's transformation, and Smale's horseshoe.**

4.10. **Subshifts.** As we have seen, the general measure-theoretic framework of invariant measures on shift spaces is in some sense *too* general, as it models any ergodic transformation. More interesting and more useful symbolic representations will come about by restricting the topological space on which the shift acts.

Thus, given an alphabet $\mathcal{A}$ with $d = \#\mathcal{A}$ and the corresponding shift space Σ, a **subshift** is a closed (hence compact) shift-invariant subset of Σ; by contrast, Σ is then known as the **full shift** on d symbols. On the one hand, given such a restriction,

one will then study the invariant measures which occur, perhaps with additional properties; on the other, given a dynamical system, the aim will be to search for a subshift which models the dynamics as closely as possible, e.g. from a topological perspective. What this means will best be seen through examples.

We shall need a definition:

Definition 4.6. Given a topological dynamical system (X, T), the $\omega-$**limit set** $\omega(x)$ of a point $x \in X$ is

$$\cap_{i=0}^{\infty} \bar{\mathcal{O}}^+ T^i(x) = \cap_{i=0}^{\infty} \text{closure}(\{T^n(x) :\geq i\}).$$

One way to define a subshift is to choose a sequence in Σ^+ or Σ, and define the subshift Ω to be the $\omega-$ limit set $\omega(x)$ Furstenberg in [Fur81] calls this a **Bebutov system**. It is possible that the point itself will not be in Ω, but if blocks in the sequence recur infinitely often it will be. Indeed any point in the subshift will exhibit the recurrence behavior of x. Thus we can choose a pattern we wish to model dynamically, and this construction will build such a model.

Exercise 4.10. *Verify these statements, taking care to make the last one precise. Find an example of a point in the shift space Σ^+ such that its $\omega-$ limit set is strictly contained in its orbit closure $\mathcal{O}^+(\underline{x})$.*

4.11. **Substitution dynamical systems.** An example of a sequence we might wish to model is one which exhibits some *self-similar* behavior: each letter is replaced by a block of k letters, then each of these is replaced and so on. A basic example is the **Thue-Morse sequence**, defined inductively by starting with the symbol 1, replacing this by 10, and thereafter replacing each 1 by 10 and each 0 by 01. Note that the limiting sequence $1001011001101001\ldots$ can be grouped into blocks of length 2^n, each labelled 1 or 0, and which exhibits exactly the same structure.

Another example imitates the self-similar structure of the Cantor set. Beginning with an infinite string of 1's, we replace these by 0 where there is a middle-third interval to be removed, giving the sequence $.10100010100000000101000101\ldots$

Here are our corresponding dynamical systems:

Example 10. The **Thue-Morse** subshift is defined as follows. Consider the sequence $\underline{x} = (\ldots 00000.10010110\ldots) \in \Sigma$; thus $x_i = 0$ for $i < 0$, and $\underline{x}^+$ is the Thue-Morse sequence. Now define Ω to be the $\omega-$limit set of x in Σ.

Example 11. The **Chacon** subshift: given the substitution $\rho(0) = 0, \rho(1) = 1101$ we consider the fixed point asociated to ρ, $\lim \rho^n(.1) = (.1101\,1101\,0\,1101...) \equiv (.a_0a_1\ldots)$. Extending this in an arbitrary fashion to a biinfinite string $a = (\ldots a_{-1}.a_0a_1\ldots)$, we then define $\Omega_{\rho,a} \equiv \cap_{n\geq 0}(\text{cl}\{\sigma^k(a)\}_{k\geq n})$ where cl denotes the closure. That is, $\Omega_{\rho,a}$ is the $\omega-$limit set of a within the compact space $\Pi_{-\infty}^{+\infty}$ acted on by the left shift σ.

Example 12. The **integer Cantor set** example is defined similarly, now taking the $\omega-$limit set of the sequence $\ldots 00000000.101000101\ldots$. (Note that the past of the sequence, chosen to be all zeroes, is completely irrelevant to the definition of the $\omega-$limit set). See [Fis92] and below.....

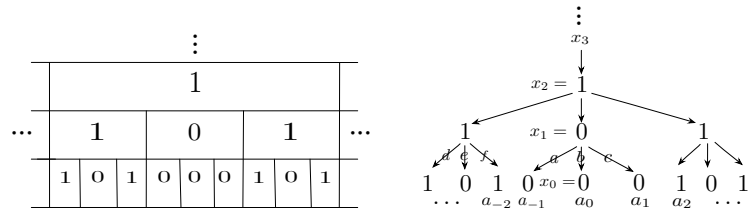


FIGURE 11. A stable equivalence class of the Integer Cantor Set transformation, depicted in the curtain and stable tree models, showing simultaneously the edge paths for the Bratteli diagram and a point in the substitution dynamical system.

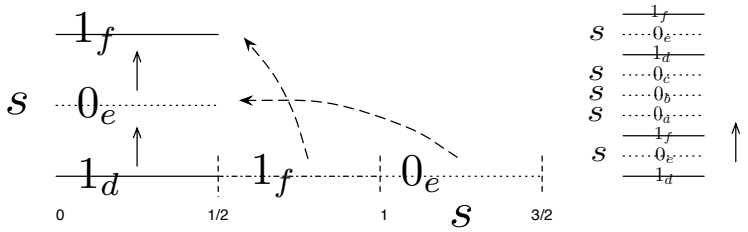


FIGURE 12. Spacer cutting-and-stacking construction of Integer Cantor Set map; induced map on I is the odometer.

These give examples of a **substitution dynamical system**; here is the general definition. We write $\mathcal{A}^*$ for the collection of all finite words from the alphabet; define a product operation on $\mathcal{A}^*$ by concatenation, which forms a semigroup (the empty word $\emptyset$ is the identity element). A **substitution** is a function $\rho : \mathcal{A} \to \mathcal{A}^*$. (In the Morse example, $\rho(0) = 01$ and $\rho(1) = 10$; for the Cantor example, $\rho(0) = 000$ and $\rho(1) = 101$.) In all cases, the function ρ extends to all of $\mathcal{A}^*$ by concatenation, and the map $\rho : \mathcal{A}^* \to \mathcal{A}^*$ is a homomorphism of this semigroup.

The map ρ also extends to Σ^+ and to Σ^- naturally (the images of a sequence beginning with a given word are pushed out towards the right and left respectively) although not to Σ; a central portion of the concatenation of limiting strings x^+ and $\underline{x}^-$ might not be in the image of ρ. Instead we consider first the map $\rho : \Sigma^+ \to \Sigma^+$, and write $\Omega^+ = \cap_n^\infty \rho^n(\Sigma^+)$; then we imbed Ω^+ in Σ and take the limit of the left shift of this set.

We have:

Proposition 4.6. *Defining $\gamma : \Omega^+ \to \Omega$ by $\underline{x}^+ \mapsto \underline{x} = \dots 00000.x^+$, we let $\Omega = \cap_n^\infty \sigma^n(\gamma(\Omega^+)$. Ω^+ is a compact invariant subset of Σ^+, and similarly for $\Omega \subseteq \Sigma$.*

In the special case where $\rho(a) = a * \cdots *$ for some $\mathbf{a} \in \mathcal{A}$, the limit $\rho^n(a * \dots) = \underline{w}$ exists, is a fixed point for ρ, and $\overline{\mathcal{O}}(\underline{w}) = \Omega^+$ in Σ^+, and in Σ, $\overline{\mathcal{O}}(\gamma(\underline{w})) = \Omega$.

In both our examples, for the Morse-Thue system for the integer Cantor set example, ρ is a **constant length substitution**, of length $k = 2$ and 3, and so these substitution dynamical systems exhibit a more direct form of self-similarity than in the general case.

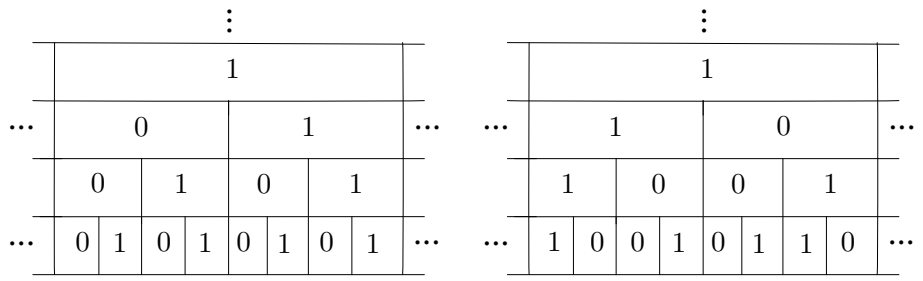


FIGURE 13. Curtain models for odometer and Morse substitutions

Regarding invariant measures, we mention (proved later) that both the Morse and Chacon substitution dynamical systems are uniquely ergodic (have a single invariant probability measure). For the Chacon system this was shown by Ferenczi [Fer95], [Fer02]. The Integer Cantor Set map is uniquely ergodic in an infinite measure sense: there exists a unique invariant measure, up to multiplication by a constant, which is positive finite on some open subset. See [Fis92].

4.12. **Subshifts of finite type.** Now let A be a $(k \times k)$ matrix with entries 0 and 1, where $k = \#\mathcal{A}$. We assume $\mathcal{A}$ is ordered, and so we write $\mathcal{A} = \{0, 1, 2, \ldots, k-1\}$. We count rows and columns of the matrix A starting with 0, and we say $\underline{x} \in \Sigma$ is an **allowed string** iff $A_{x_i x_{i+1}} = 1$, and write Σ_A for the collection of all allowed biinfinite strings $\underline{x} = (\ldots x_{-1}.x_0 x_1 \ldots)$.

We associate to A a finite graph, with one vertex for each symbol and a directed edge (an arrow) from symbol i to symbol j exactly when the transition from i to j is allowed, i.e. when $A_{ij} = 1$. Note that there is at most one edge between any two symbols. This gives the **vertex-shift** model of a **(two-sided) subshift of finite type (sft)**. The corresponding one-sided vertex **sft** is Σ_A^+, the collection of allowed strings $\underline{x}^+ = (.x_0 x_1 \ldots)$.

Note that the projection $\underline{x} \mapsto \underline{x}^+$ defines a semiconjugacy from the two-sided to the one-sided *sft*.

The simplest example is the **full two-shift** $\Sigma = \Pi_{k \in \mathbb{Z}} \mathcal{A}$ for $\mathcal{A} = \{0, 1\}$, along with its one-sided version $\Sigma^+ = \Pi_{k \geq 0} \mathcal{A}$, which we used above to model coin-tossing. Note that $\Sigma = \Sigma_A$ for $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

To define the **edge-shift model** for a subshift of finite type, we begin with a finite graph G with k vertices, but allow any finite number M_{ij} of directed edges from vertex i to vertex j. We associate to this a $(k \times k)$ matrix M, but now with nonnegative integer entries.

Now we build a new graph $\widehat{G}$, whose vertices are the m edges of G. This will have the single-edge property. We define Σ_M to be the vertex shift on this new graph, so the new alphabet is the edges, labelled $0, 1, \dots m-1$.

Coin-tossing also has an edge-shift model: Σ_M with $M = \begin{bmatrix} 2 \end{bmatrix}$. There is one vertex with two edges, labelled 0 and 1.

For more interesting examples, see Fig. 14, the **golden vertex shift** with alphabet $\mathcal{A} = \{A, B\}$ and matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$. Its graph is indicated on the left side of Fig. 14.

The name "golden" refers to the **golden ratio**

$$\phi = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cdots}}$$

We see from this continued fraction expansion that $1/\phi = \phi - 1$ whence $\phi^2 - \phi - 1 = 0$, which has roots $(1 \pm \sqrt{5})/2$, and so

$$\phi = \frac{1 + \sqrt{5}}{2} = 1.618 \cdots$$

Calculating the eigenvalues of A (see Definition 16.1) these are the roots of the characteristic polynomial of A, $p(\lambda) = \det(A - \lambda I) = \begin{vmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 - \lambda - 1$; this factors as $p(\lambda) = (\lambda - \lambda^+)(\lambda - \lambda^-)$ for $\lambda^{\pm} = \frac{1 \pm \sqrt{5}}{2}$. So $\phi = \lambda^+$ is the largest eigenvalue of A (the Perron-Frobenius eigenvalue, see §16). We shall see the usefulness in calculating the topological entropy in a moment, see Example **??**.

Taking the square of the matrix gives $M = A^2 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ we have a nonnegative integer matrix with graph on the right side of Fig. 14, giving the **golden edge shift** Σ_M. The alphabet $\mathcal{A}$ gives the vertex set in both cases; note that for the edge shift, multiple edges are allowed. We can also represent the edge shift Σ_A as a vertex shift. For this we define a new alphabet: the collection of edges; this gives three symbols $\{e, f, g\}$ with a (3×3) $0-1$ matrix

$$N = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

The shift spaces Σ_A and Σ_N are naturally isomorphic via a **two-block code**: given $\underline{x} = (\dots x_1.x_0x_1 \dots)$ we send this to the sequence $(\dots (x_{-1}, x_0).(x_0, x_1)(x_1, x_2) \dots)$ of *two-blocks* and then on to the sequence of edges $(\dots e_{-1}.e_0e_1 \dots) \in \Sigma_N$ such that e_i goes from vertex x_i to vertex x_{i+1}. For an excellent introduction to this coding and information theory perspective on ergodic theory, see [LM95].

We can do this also for the golden edge shift. Our new alphabet is the edge set $\mathcal{E}$ with the 5 symbols $\{a, b, c, d, e\}$; then the edge set $\mathcal{E}$ has become the vertex set for the corresponding vertex shift, with now a (5×5) $0-1$ matrix. This illustrates the advantage of dealing with edge shifts rather than vertex shifts: the (2×2) matrix is easier to handle than the (5×5) version.

FIGURE 14. Graphs for the golden vertex shift and edge shift.

For example, the (2×2) matrix leads to these observations. First, M has a second nice factorization, as

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

Secondly, for $n \geq 0$,

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix}, \begin{bmatrix} 8 & 5 \\ 5 & 3 \end{bmatrix} \cdots$$

with entries in the upper left corner given by the Fibonacci sequence $1, 1, 2, 3, 5, 8, 13, \ldots$ We note also that the sum of the matrix entries gives this sequence beginning at 2.

Exercise 4.11. *Show that for a nonnegative matrix A, the number of allowed words of length n in Σ_A is the sum of the entries in the matrix A^n.*

We can give here a preliminary definition of a much more general concept:

Definition 4.7. Given a subshift $\Omega \subseteq \Sigma$, the **topological entropy** of (Ω, σ) is

$$\lim \frac{1}{n} \log \#(\text{ allowed words of length } n)$$

Proposition 4.7. *In the case of the golden shift, the topological entropy is* $\log \phi$.

Proof. We recall (see Definition 35.10 below) that the operator norm of a matrix is $||A||_{op} = \sup_{||\mathbf{v}||=1} ||A\mathbf{v}||$. To apply this we take the L^1 norm on R^d, that is $||\mathbf{v}||_1 = \sum_{i=1}^n |v_i|$. (See §6.2.)

As shown in Lemma 35.41, in a finite dimensional vector space all norms are equivalent.

Now we refer to Proposition 16.9: the number of words of length n is equal to the he L^1-norm of A^n, where $||A||_1 = \sum_{i,j} |A_{ij}|$. Then, as in the third proof of Corollary 16.10, this is within constant multiples of $(\lambda^+)^n$. It follows that the topological entropy is $\log \lambda^+$.

□

Remark 4.2. In fact the limit used in the above definition of topological entropy always exists; this follows from a subadditivity lemma applied to the operator norm, see below ???.

Invariant measures for subshifts of finite type are a much-investigated and fascinating subject; we return to this below, in §16.6.

4.13. **Toral endo- and automorphisms.** We have noted above that the $d-$ torus is the factor space (quotient topological space, and factor group) $\mathbb{R}^d/\mathbb{Z}^d$. Now let A be a $(d \times d)$ matrix with integer entries. Acting on, say, column vectors, this maps $\mathbb{Z}^d$ into itself which implies that A gives a well-defined map on the torus: $\mathbb{R}^d/\mathbb{Z}^d$ is the collection of cosets $\mathbf{v} + \mathbb{Z}^d$ and $A(\mathbf{v} + \mathbb{Z}^d) = A\mathbf{v} + A\mathbb{Z}^d \sim A\mathbf{v} + \mathbb{Z}^d$. This is called a *toral endomorphism*.

We claim that this map of the torus is invertible iff determinant A is ± 1.

Consider first the (2×2) case.

Now more generally, if $\det A = \pm 1$, then $A \in GL(d, \mathbb{Z})$ so its inverse also has integer entries. ??

For one of the the simplest examples, consider $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$, which we encountered already in a different setting (defining a subshift of finite type).

Example 13. The **golden toral automorphism** is the map of the torus given by the action of $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ on column vectors.

Our aim here is to show how the geometry of this map gives a continuous analogue of the (very discontinuous!) baker's transformation.

For this purpose, we diagonalize the matrix by finding its eigenvalues

$$\lambda^{\pm} = \frac{1 \pm \sqrt{5}}{2}$$

An eigenvector corresponding the eigenvalue λ is $(x, 1)$ satisfying $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix} = \begin{bmatrix} x+1 \\ x \end{bmatrix} = \begin{bmatrix} \lambda x \\ \lambda \end{bmatrix}$ so $x = \lambda$.

Noting that $\lambda^+\lambda^- = -1$, the eigenvectors $(\lambda^{\pm}, 1)$ are orthogonal and can be normalized to have length one, as $\mathbf{v}^{\pm}$. Letting Q be the matrix with columns $\mathbf{v}^-, \mathbf{v}^+$, then $Q^{-1}AQ = D$ is diagonal with $D = \begin{bmatrix} \lambda^- & 0 \\ 0 & \lambda^+ \end{bmatrix}$.

(That one can find an orthogonal change-of-basis matrix Q is a consequence of a general fact for $(d \times d)$ real symmetric matrices, the *Spectral Theorem*. See Theorem 35.56.)

We find the continued fraction expansion of the eigenvalues. Setting $\lambda = \varphi^{-1}$, let us first find the expansion for $\varphi^+ \equiv (\lambda^+)^{-1}$. Since λ satisfies $\lambda^2 - \lambda - 1 = 0$, we have $\varphi^{-2} - \varphi^{-1} - 1 = 0$ whence $\varphi^2 + \varphi^{-1} - 1 = 0$, so from the quadratic formula, $\varphi^{\pm} = \frac{-1 \pm \sqrt{5}}{2}$ which indeed is the inverse of $\lambda^{\pm}$. Consider $x = [111\ldots]$; then $1/x - x = 1$, whence $x^2 + x - 1 = 0$ and so indeed $x = \varphi^+$. Thus, $\lambda^+ = 1/x = 1 + [111\ldots]$. This is in the literature usually written as $[1; 111\ldots]$, denoting the continued fraction expansion of an irrational in $(0, +\infty)$.

Example 14. In discussing subshifts of finte type, we have already encountered $A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^2$, one of the most famous maps in ergodic theory. Now $\det A = 1$, so

it is orientation-preserving. Of course the eigenvectors are the same as for $\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}$, and its eigenvalues are the squares. As mentioned above, it has a second factorization, as $A = \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}$, which will have dynamical as well as geometric significance, and will greatly aid in a study of related maps, see below.

Example 15. We return to further examples throughout these notes. But for now we mention another example: an *endo*morphism of the torus which perhaps deserves to be called the *doubling map on the torus*. This is given by $A = \begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix}$. Note that in complex notation this is the map $T : \mathbb{C} \to \mathbb{C}$ given by

$$T : z \mapsto (1+i)z.$$

Since $\det A = 2$, this doubles area, and is two-to-one, and for these reasons is a two-dimensional analogue of the doubling map of the circle.

Furthermore,as seen below in §22.1, one can code the map in the same way as for the doubling map of the circle: by a one-sided Bernoulli two-shift, modelled probabalistically by the tosses of a fair coin. This coding is given by by the construction of special Markov partitions, of both geometric and arithmetic origin, which have fractal boundaries, and this will lead us off into fascinating directions.

4.14. **A Markov partition for a hyperbolic toral automorphism.**

Here we shall link the previous two sections by showing how a hyperbolic toral automorphism can be represented symbolically by a subshift of finite type.

Definition 4.8. A finite partition $\mathcal{P}$ for a continuous map T of a topological space $(X.\mathcal{T})$ is a **(topological) Markov partition** if there exists a $(d \times d)$ nonnegative integer matrix M which codes the map as an (edge or vertex) subshift of finite type. More precisely, for the case of T invertible, there exists a semiconjugacy

$$\begin{array}{ccc} \Sigma_M & \xrightarrow{\sigma} & \Sigma_M \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\ X & \xrightarrow{T} & X \end{array}$$

such that π is a continuous surjection which is a homeomorphism off of an invariant set of measure zero; for T noninvertible we replace Σ_M by the one-sided shift space Σ_M^+.

The term "Markov partition" indeed comes from a connection with the Markov processes of probability theory; see §15.3 regarding this part of the theory.

The "bad" set of measure zero can be thought of as the forward and backward iterates of the partition boundaries. In the two-dimensional case, that is, for hyperbolic automorphisms of $\mathbb{T}^2$, the partitions are rectangles and the boundaries are just line segments: subsets of the stable and unstable eigendirections at the point $(0, 0)$. For higher dimensions, partitions become much more complicated, typically with fractal boundaries; see §22.1.

We have already encountered the simplest examples: the doubling map $g : I \to I$, and the baker's transformation $\widehat{g}$ of the square $I \times I$. For the doubling map the Markov partition of I is $\mathcal{P} = \{P_0 = [0, 1/2], P_1 = [1/2, 1]\}$, leading to the binary expansion of a point $x \in I$ and so to a coding of g by the one-sided Bernoulli shift (Σ^+, σ). For the baker's transformation the pair of rectangles codes the map by the bilateral shift.

The first step beyond this was carried out by Adler and Weiss for hyperbolic automorphisms of the two-torus $\mathbb{T}^2$ in [AW70]; their construction of Markov partitions represents these maps symbolically by subshifts of finite type (Σ_M, σ). The next step is to find an appropriate measure on this shift space; this will represent the toral automorphism, with Lebesgue measure on the (square or parallelogram, see below) torus, as a Markov shift of probability theory, see §15.3 regarding the connection between the topological and probability Markov properties in general.

Now for a toral automorphism the natural invariant measure is Lebesgue measure on the square $[0, 1) \times [0, 1)$, since this is a fundamental domain for $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$. What Adler and Weiss discovered was that when transferred to the symbolic space, this has a wonderfully simple and elegant combinatorial expression as a very special Markov measure, called the **Parry measure** of the *sft*. This can be characterized as the unique measure of maximal entropy (for both maps: the toral automorphism and the topological shift space defined by the *sft*). We explain this below in §16.6. The situation is exactly analogous to the the doubling map and baker's transformation on the interval or square, where Lebesgue measure corresponds to the infinite product measure of coin-tossing on the symbol space (the full shift). The maximum entropy property is related to ideas from both information theory and from the physics of lattice models.

The key new idea of Adler and Weiss is essentially this. They interpret the Markov property of probability theory geometrically, in terms of how the preimage of the partition meets the partition. This property is intrinsically related to the hyperbolicity of the map, an insight so striking and important that that it was soon pushed far beyond the original setting, to Anosov and Axiom A diffeomorphisms and to the "thermodynamic formalism" of Sinai, Bowen and Ruelle [Bow75], [Bow77]; see e.g. [KH95].

We mention that Ken Berg had a similar idea to that of Adler and Weiss, at about the same time, in his (unpublished) thesis at the University of Maryland.

In this section we explain the symbolic part of this (construction of the Adler-Weiss Markov partition) for the simplest example, the *golden* toral automorphism $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$, see Example 13. Our partition will have the remarkable property that $M = A$, i.e. the transition matrix is that same as that for the map itself.

It will be convenient to choose a different change-of-basis matrix from the orthogonal matrix Q used above in §4.12. For this, beginning with the matrix whose columns are the eigenvectors we found, $\begin{bmatrix} \lambda^- & \lambda^+ \\ 1 & 1 \end{bmatrix}$, we leave the first column unchanged and normalize the second so the matrix has determinant 1. The result is $B = \begin{bmatrix} a & c \\ -b & d \end{bmatrix}$,

where $a = \lambda^-$, $b = 1$, $c = \lambda^+/\sqrt{5}$ and $d = 1/\sqrt{5}$. We define $\mathbf{v}^s, \mathbf{v}^u$ to be these columns, also eigenvectors, and again we have the diagonalization $B^{-1}AB = D$ where $D = \begin{bmatrix} \lambda^- & 0 \\ 0 & \lambda^+ \end{bmatrix}$. That is, the following diagram commutes, where matrices are acting on column vectors:

$$\begin{array}{ccc} \mathbb{R}^2 & \xrightarrow{A} & \mathbb{R}^2 \\ \uparrow B & & \uparrow B \\ \mathbb{R}^2 & \xrightarrow{D} & \mathbb{R}^2 \end{array}$$

Now let Λ be the lattice subgroup of $\mathbb{R}^2$ generated by the column vectors of $B^{-1} = \begin{bmatrix} d & -c \\ b & a \end{bmatrix}$. That is, Λ is the image by B^{-1} of the integer lattice $\mathbb{Z}^2$. This implies that the commutative diagram passes on to a conjugacy of maps on the quotient spaces, the **square torus** defined by $\mathbb{R}^2/\mathbb{Z}^2$ and the **parallelogram torus** $\mathbb{R}^2/\Lambda$:

$$\begin{array}{ccc} \mathbb{R}^2/\mathbb{Z}^2 & \xrightarrow{A} & \mathbb{R}^2/\mathbb{Z}^2 \\ \uparrow B & & \uparrow B \\ \mathbb{R}^2/\Lambda & \xrightarrow{D} & \mathbb{R}^2/\Lambda \end{array}$$

Proposition 4.8. *$\mathcal{P} = \{P_0, P_1\}$ defined above is a Markov partition for the toral automorphism $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ of $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$. That is, the map π defined by the symbolic dynamics is a semiconjugacy from (Σ_A, σ) to $(\mathbb{T}^2, A)$ which is 1-1 off of an invariant set of Lebesgue measure zero, consisting of the iterates of the partition boundaries, and the inverse image inside of Σ_A of that set.*

Proof. For $\mathcal{A} = \{0, 1\}$, we define a relation R on $\Pi_{-\infty}^{\infty}\mathcal{A} \times \mathbb{T}^2$ determined by the symbolic dynamics of the map A and partition $\mathcal{P}$. That is, $R \subseteq \Pi_{-\infty}^{\infty}\mathcal{A} \times \mathbb{T}^2)$ is defined by: for $\underline{x} = (\dots x_{-1}.x_0x_1\dots)$, $\underline{x} \sim_R x$ iff $A^i(x) \in P_{x_i}$ for all $i \in \mathbb{Z}$.

We claim that R is a function π from Σ_A to $\mathbb{T}^2$, that is, $(\underline{x}, x) \in R$ iff $x = \pi(\underline{x})$, which satisfies the claimed properties.

□

Fig. 16 illustrates how, using the Markov partition, the toral automorphism A can be thought of as a continuous version of the baker's transformation. The illustration of Fig. 15 has been rotated by angle $\pi/2$ so as to make this analogy clearer. The expansion is now along the $x-$axis, and the automorphism acts via the diagonal map $\begin{bmatrix} \lambda^+ & 0 \\ 0 & \lambda^- \end{bmatrix}$. Since $-1 < \lambda^- < 0$, the two boxes are reflected and contracted in the $y-$axis. A cutting and stacking then takes us back to the original configuration. To this point the analogy to the baker's transformation construction is exact (although to be sure the boxes have different sizes, and the orientation reversal of the map causes the boxes to be reflected in the $y-$axis.) However now there is a big difference: the cutting-and-stacking is *also* given by an identification by the lattice Λ, and so in $\mathbb{R}^2/\Lambda$ we haven't actually done anything! As a result the map is indeed a homeomorphism:

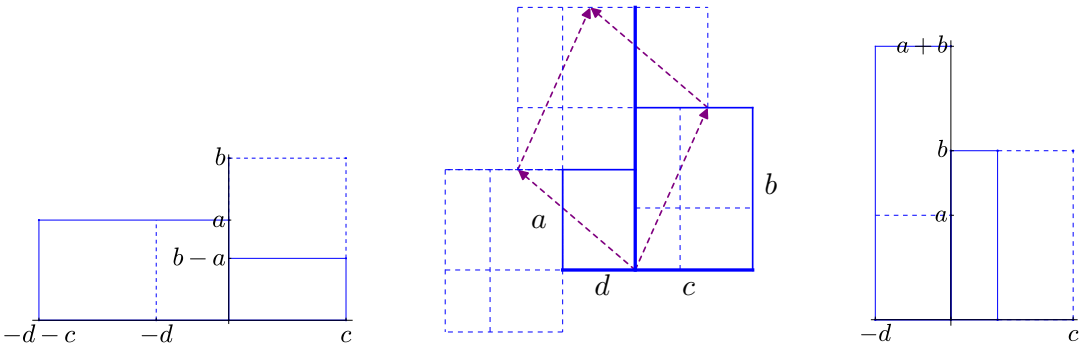


FIGURE 15. Markov partition for the automorphism D of the parallelogram torus $\mathbb{R}^2/\Lambda$: contracting and reflecting along the $x-$axis, expanding along the $y-$axis. On the left is the partition (dotted lines) with its inverse image; in the center, the partition joined with pullbacks of image and preimage; on the right, partition with its image. The partition boundary consists of two line segments which meet at the origin: the upside-down T of the central figure.

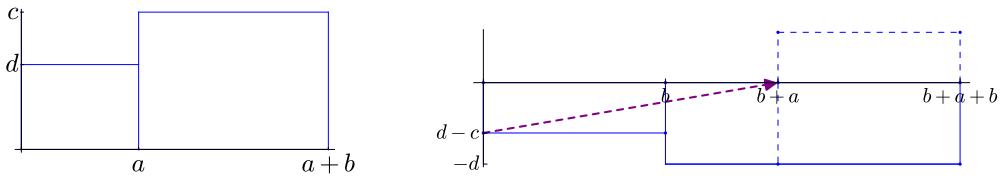


FIGURE 16. Rotating 90° from Fig. 15 , we see the toral automorphism is a continuous version of the baker's transformation. First we expand along the $x-$axis while contracting and reflecting along the $y-$axis; then we cut and stack this image Markov partition to return to the original pair of boxes (modulo the lattice). The smaller box is labelled 1 and the larger 0, giving the transition matrix A for the subshift of finite type.

the space has not been ripped apart (as for the baker's transformation), which for that map resulted in discontinuities along the partition borders.

In this way, our toral automorphism is a continuous (hence "improved") version of the baker's transformation, and so one might wonder if there is an analogue for the doubling map in this case as well, that is, if there is a noninvertible map of the interval which the toral automorphism will have as a homomorphic image.

The answer (Yes!) is shown in Fig. 17. What we do is to consider two line segments $l_1 = [-a, 0]$ and $l_0 = [0, b]$ of the Markov partition of Fig. 16, the lower sides of each of the two boxes; apply the matrix $\widetilde{D}$, and project these vertical segments back down to $l_1 \cup l_0$. This is the map of the interval $[0, b + a]$ given by $f : x \mapsto \lambda^+ x$, where $\lambda^- = b/a > 0$, so it is orientation-preserving, and is **hyperbolic** i.e. $|Df| = b/a > 1$. There is no natural way to decide whether the interval is $[-a, b]$ or $[0, a + b]$, so a better topological model is given by the circle $\mathbb{R}/(a + b)\mathbb{Z}$. But an even more

natural model is given by identifying the endpoints of *both* intervals, to a single point, giving a **bouquet of circles**, the first of length l_0 and the second of length l_1. This topological space is a special case of a **train track** (a concept introduced by Thurston and also by Williams, who called it a **branched one-manifold**, [Wil74]), and this type of Markov map is a **train track map** (see e.g. Bestvina-Handel [BH92] and Los [Los93]). Thus the map of the interval factors onto a map of the circle which in turns factors onto the train track map; this is analogous to the doubling map of the interval factoring onto the doubling map of the circle.

The image of each segment is a union of copies of the l_i, since the image of the partition is a pair of boxes which horizontally are made up of the original ones. In other words, for $f : l_0 \cup l_1 \to l_0 \cup l_1$, the forward image of l_0 is $l_0 \cup l_1$ and of l_1 is l_0, giving the matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ for the combinatorial model Σ_A^+.

Note, however, that Lebesgue measure is not preserved by this map. Due to some quite general and more advanced considerations, there is a unique invariant probability measure which is absolutely continuous with respect to (and indeed is equivalent to) Lebesgue measure. But what is it? In fact the description is easy given our invertible version of the transformation, the toral automorphism depicted in Fig. 16: the measure is $f(x)\mathrm{d}x$ where the graph of f, with the intervals drawn with l_1 on the left, is given on the left-hand side of the figure!

Exercise 4.5. Verify that this is indeed an invariant probability measure for the Markov map.

From the combinatorial perspective of the shift spaces Σ_A and Σ_A^+, both of these invariant measures (for the toral automorphism and the Markov interval map) have remarkable formulas, as the Parry measure, which will be explained in §16.6. The existence of this formula is one of the main reasons why Markov partitions have turned out to be so important.

Remark 4.3. Very nice Markov partitions, with the Adler-Manning property, can be constructed for any orientation-preserving hyperbolic automorphism of $\mathbb{T}^2$, by a similar method. The construction given above, which we learned from Arnoux, will be returned to in §§ 11.4 and 25.5. However one change must be made. In the above example, since matrix entries are $0, 1$, we can use a vertex shift space to represent the combinatorial space. For general hyperbolic automorphisms of $\mathbb{T}^2$, one first shows there exists a conjugacy to an integer matrix with nonnegative entries; one then constructs a Markov partition for this map, such that the transition matrix will be this same matrix, now using the edge shift convention for the *sft*. This discovery of such a nice is due (in different contexts) to Adler [Adl98], Manning [Man02], and also to [AF05].

Let us for example consider the action of

$$M = A^2 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

This matrix defines a hyperbolic toral automorphism, which is now orientation-preserving, with the same eigenvectors as A but now with eigenvalues $(\lambda^+)^2, (\lambda^-)^2$.

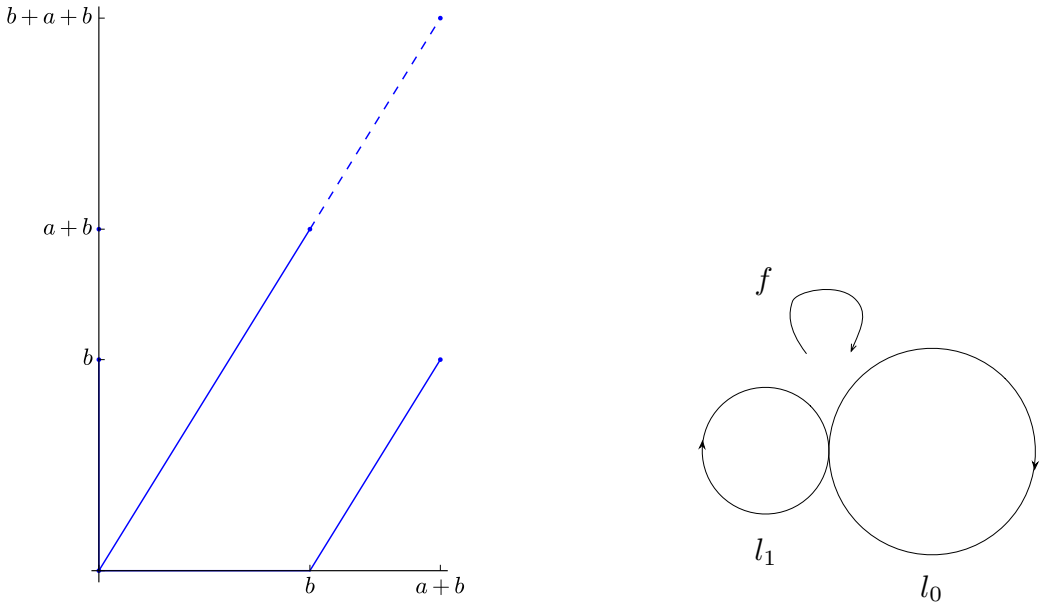


FIGURE 17. Analogue of the doubling map: a Markov map of the interval $l_0 \cup l_1$ by $f : x \mapsto \lambda^+ x$, which factors on to the map of the circle $\mathbb{R}/(a+b)\mathbb{Z}$ and from there onto the *train-track* map where the endpoints of l_0 and l_1 are identified. This is a factor of the toral automorphism of Fig. 16, via projection to the $x-$axis.

By the Adler-Manning theorem, there is a Markov partition for M which represents it as an edge shift space with exactly the same matrix M. Therefore, the edge alphabet and partition have 5 elements, and one can see exactly these five rectangles in the center drawing of Fig. 15, which depicts a generating Markov partition $\mathcal{Q} = A(\mathcal{P}) \vee \mathcal{P} \vee A^{-1}(\mathcal{P})$ for the map M. Another choice which will work is $\mathcal{P} \vee A^{-1}(\mathcal{P}) \vee A^{-2}(\mathcal{P})$; the reader can sketch this partition as an exercise.

5. Recurrence: measure-theoretic and topological

5.1. **Four proofs of the Poincaré recurrence theorem.** Already one can quite easily prove a rather amazing result. This is so important that we shall encounter a number of proofs; the first three we learned from the books of Walters, Halmos and Furstenberg; each has its own points of interest, it is well worth studying them all. The fourth proof isolates some key ideas that remain valid in the setting of infinite measure transformations.

Theorem 5.1. *(Poincaré) Let $T : X \to X$ be a measure-preserving transformation of a probability measure space $(X, \mathcal{A}, \mu)$. Let $A \subseteq X$ with $\mu(A) > 0$. Then $\mu-$almost every point in A returns to A infinitely often.*

Proof. (I)([Wal75]) A point $x \in X$ is in A at time k iff $T^k(x) \in A$ iff $x \in T^{-k}(A)$. So the point enters A after time N iff $x \in \cup_{k=N}^{+\infty} T^{-k}(A)$. It is in A infinitely often iff it is in A for some time $\geq N$ for all $N > 0$. Thus the set of "good" points G, which

begin in A and return to it infinitely often, is

$$G = A \cap \bigcap_{N=0}^{+\infty} \bigcup_{k=N}^{+\infty} T^{-k}(A).$$

Write A_N for $\cup_{k=N}^{+\infty} T^{-k}(A)$, so $G = A \cap \bigcap_{N=0}^{+\infty} A_N$. Note that

$$A \subseteq A_0 \supseteq A_1 \supseteq \dots$$

and that $T^{-1}(A_i) = A_{i+1}$. Since T preserves the measure, $\mu(A_1) = \mu(A_0)$, yet $A_0 \supseteq A_1$; since $\mu(X) < \infty$, therefore $\mu(A_0 \Delta A_1) = 0$, and similarly, $\mu(A_0 \Delta \cap_{N=0}^{+\infty} A_N) = 0$. Thus $\mu(G) = \mu(A \cap \bigcap_{N=0}^{+\infty} A_N) = \mu(A \cap A_0) = \mu(A)$ since $A \subseteq A_0$. □

Proof. (II)([Hal60]) For this approach we look at the set of "bad"points B, those points in A that *never* return. This is

$$B = A \cap T^{-1}(A^c) \cap T^{-2}(A^c) \cap \dots.$$

Now if $x \in B$, it never returns to B, since $B \subseteq A$. Thus $B \cap T^{-l}(B) = \emptyset$ for any $l \geq 1$. Hence for any $k \geq 0$,

$$T^{-k}(B \cap T^{-l}(B)) = \emptyset$$

but this is $T^{-k}B \cap T^{-(l+k)}(B) = T^{-k}B \cap T^{-n}B$ for $n = l+k$. So the sets $B, T^{-1}(B), T^{-2}(B), \dots$ are all disjoint. Now they have the same measure; $\mu(X) < \infty$ then forces $\mu(B) = 0$. What we have shown so far is that a.e. $x \in A$ returns *at least once*. The remaining step is:

Lemma 5.2. *Suppose we know that with the hypotheses of Theorem 5.1, a.e. $x \in A$ returns to A at least once. Then a.e. x returns infinitely often.*

Proof. We apply the proof just given to each of the transformations (X, T^k, μ) for $k \geq 1$. This produces sets $A_k \subseteq A$ of measure $= \mu A$ whose points return at least once for the map T^k. Therefore $x \in A_k$ returns for the map T for some time $> k$. So each $x \in \cap_{k=1}^{+\infty} A_k$ returns under the map T for some time $> n$ for each n, hence returns to A infinitely often. □

□

Each of these arguments has its interesting aspects: Walters' brings in the important notion of the lim sup of a sequence of sets (see Exercise 5.2 below; this notion also occurs in the proof of Lemma 12.4 of Borel-Cantelli and Proposition 5.7 on transitive points; see also [Bar66], [Loè77]), while Halmos' use of the finiteness of the measure space is more transparent, as he constructs directly a sequence of disjoint sets with the same measure (and hence of measure 0). But my favorite is this beautiful little argument due to Furstenberg [Fur81]:

Proof. (III) We shall show that a.e. point in A returns at least once, then apply the lemma given above. Now $\mu(A) > 0$ and the sets $A, T^{-1}A, T^{-2}A, \dots$ each have the same measure, hence cannot all be disjoint. So there exist $i < j = i + k$ and a point x which belongs to both $T^{-i}A$ and $T^{-j}A$. Thus $T^i(x)$ and $T^k(T^i(x))$ are in A. Calling $y = T^i(x)$, we have found a single point $y \in A$ which returns to A. Rather amazingly, this is enough to finish the proof!

Let $B \subset A$ be the points which don't return. But if B has positive measure, then there exists $x \in B$ which returns to B, by the previous argument, hence to A ! So a.e. point in A does return. □

Definition 5.1. Let $(X, \mathcal{A}, \mu)$ be a (possibly infinite) measure space and $T : X \to X$ a measure-preserving transformation. A measurable subset $A \subseteq X$ is **invariant** iff $T^{-1}(A) = A$. The opposite case is that the subset A is **wandering**: that $\{T^{-i}(A)\}_{i \geq 0}$ are all disjoint. A is **compressible** iff $A \supseteq T^{-1}A$ and $\mu(A \setminus T^{-1}A) > 0$.

One says $A \subseteq X$ is **trivial** iff either $\mu(A) = 0$ or $\mu(A^c) = 0$. A measurable transformation is **ergodic** iff any invariant set A is trivial; it is **conservative** iff any wandering set has measure zero. It is **recurrent** iff for any set A of positive measure, for a.e. $x \in A$ there exist infinitely many $n > 0$ such that $T^n(x) \in A$. A set is invariant for a (semi)group action if that holds for each element individually, and is wandering if all the inverse images are disjoint. Ergodic and conservative actions are then defined in the same way.

Exercise 5.1. given a map $T : X \to X$, let us say a set A is **forward–invariant** iff $A = T(A)$.
(i) Show that an invariant set is forward invariant.
(ii) Find an example of a map with a subset A that is forward-invariant but not invariant.
(iii) If T is **invertibl**e, meaning by definition that it is invertible as a measurable transformation; i.e. it is a bijection and is bimeasurable, then T^{-1} is also a measure-preserving map. Then T is ergodic iff T^{-1} is.
(iv) Let us say $T : X \to X$ is **locally invertible** if there exists a countable partition $\{P_i\}_{i \in \mathbb{N}}$ into measurable sets such that each image $T(P_i)$ is measurable, each restriction $T|_{P_i}$ is measurable and and is invertible onto its image. In this case the forward image of any measurable set is measurable.
(v) Find a measure space $(X, \mathcal{A}, \mu)$ and $f : X \to X$ measurable with a measurable $A \subseteq X$ such that the forward image is nonmeasurable. (Hint: in a complete sigma-algebra, all subsets of a set of measure zero are measurable; let A be a subset of a Cantor subset of the interval which maps forward to a nonmeasurable subset of the interval).

We remark that by a deep result of Rochlin, a bijective map on a Lebesgue space is in fact bimeasurable, so it is invertible in the above stronger sense.

Proposition 5.3. *If $\mu(X) < \infty$ then for a measure-preserving transformation T on X:*
(i) There are no wandering sets of positive measure;
(ii) There are no compressible sets;
(iii) T is recurrent.

Proof. Parts (i) and (ii) are immediate; part (iii) is the Poincaré recurrence theorem. □

In the infinite measure setting we can take these conclusions as (desirable) properties; then the first question will be to understand how they are related.

We have:

Proposition 5.4. *There exists a nontrivial wandering set iff there exists a nontrivial compressible set.*

Proof. Given A wandering we take $B = \cup_{i=0}^{\infty} T^{-i}A$; if $\mu A > 0$, this is compressible.
Conversely, given A compressible, define $B = A \setminus T^{-1}A$; this is wandering. □

(Draw the pictures!)

Proposition 5.5. *T is conservative iff it is recurrent.*

Proof. First we prove the easy part: that recurrent implies conservative.

Suppose T is *not* conservative. Then there exists a wandering set A. Hence $\mu(A) > 0$ and for all $k > 0$, $A \cap T^{-k}A = \emptyset$. Thus $x \in A \implies x \notin T^{-k}A$; equivalently, $T^k x \notin A$. So x *never* returns.

Now assume T is conservative. We imitate the proof of Walters for the Poincaré recurrence theorem given above. Thus, given A of measure > 0, we set

$$G = A \cap \bigcap_{N=0}^{+\infty} \bigcup_{k=N}^{+\infty} T^{-k}(A) = A \cap \bigcap_{N=0}^{+\infty} A_N,$$

and wish to show that $\mu(A \setminus G) = 0$.

As before,

$$A \subseteq A_0 \supseteq A_1 \supseteq \dots$$

and $T^{-1}(A_i) = A_{i+1}$. Since T is conservative, there exists no compressible set; hence $\mu(A_i \setminus A_{i+1}) = 0$. Therefore just as before, $\mu(A_0 \Delta \cap_{N=0}^{+\infty} A_N) = 0$, $\mu(G) = \mu(A \cap \bigcap_{N=0}^{+\infty} A_N) = \mu(A \cap A_0) = \mu(A)$ since $A \subseteq A_0$. Hence up to a set of measure 0, A equals G. □

Now since finite measure implies conservative, we have as a corollary yet another proof of the Poincaré recurrence theorem.

We remark that the arguments of Halmos and of Furstenberg also work in this setting, showing that if T is conservative, then for A with measure > 0, a.e. point returns *once*. However to prove that it returns infinitely often, we would need to answer:

Question 1. Does T conservative imply T^n conservative for all $n > 0$?

Exercise 5.2. Given a measure space $(X, \mathcal{A}, \mu)$ we consider the map from $\mathcal{A}$ to L^∞ given by $A \mapsto \chi_A$. Find function-space interpretations for these operations on sets: intersection, union, symmetric difference. Find set interpretations for these operations on functions: product, $|f - g|$, $\int_X |f - g| \mathrm{d}\mu$, $\sup f_i$, $\inf f_i$, $\limsup f_i$, $\liminf f_i$.

5.2. **Transitive points and Baire category.** The usual understanding of a property holding for "almost all" points of a space is measure theoretical, that the complement of the subset where this is valid have measure zero. A complementary topological notion is provided by sets of second Baire category. Given the remarkable recurrence theorem just discussed, it is natural to wonder if there is a purely topological version, based on the notion of Baire category.

Given a homeomorphism T of a complete separable metric space X, a point $x \in X$ is **transitive** iff it has a dense orbit. The aim of this section shall be to show that if there exists one transitive point, the set E of transitive points is residual.

The fascinating relationship between these two very different but in many ways parallel ideas of measure and category is explored in Oxtoby's wonderful little book [Oxt80] of exactly that title (which cannot be recommended too highly, e.g. for graduate students refining their knowledge for analysis qualfying exams, for professors preparing for a lecture course, or for anyone with the time to pursue beautifully presented ideas for their own sake). Here we bring in some basic definitions and one result, specifically related to dynamics. And indeed, the methods involved may remind one of Poincaré recurrence, specifically of Walters' proof of the previous section.

The statement, moreover, might recall Furstenberg's proof of Poincaré recurrence: there, proving first that a single point recurs to a positive measure set A, we then used this to show it holds for (measure theoretically) a.e. point in A, and moreover that a.e. point returns infinitely often.

This always surprising type of logical argument, where an apparently much weaker statement is used to draw a stronger conclusion, is called (*bootstrapping*, as it seems like the genuinely impossible act of "pulling oneself up by one's own bootstraps".

Definition 5.2. A set X with topology $\mathcal{T}$ is a **Polish space** iff there exists a compatible metric $d(\cdot, \cdot)$ which makes X a complete separable metric space.

Exercise 5.3. Show that the following spaces are Polish: (i)a compact metric space; (ii) the real line; (iii) the open interval $(0, 1)$ with the usual topology; (iv) an open disk minus one point; (v) $\Pi_{-\infty}^{\infty}\mathbb{R}$ with the product topology; (vi) a countable product of Polish spaces; $(vii)$the space of continuous functions from $\mathbb{R}$ to $\mathbb{R}$ with the topology of uniform convergence on compact sets.

Definition 5.3. Let $(X, \mathcal{T})$ be a topological space. A A subset A is **nowhere dense** if there is no nonempty open set in which it is dense. It is **meagre** or **of first (Baire) category** iff it is a subset of a countable union of nowhere dense sets. It is **residual** or **of second category** iff it contains a countable intersection of dense open sets.

Thus, the complement of a meagre set is residual and vice-versa.

We recall that a countable union of closed sets is called an $F_\sigma-$set, while its complement, a countable intersection of open sets, is a G_δ. Thus a meagre set is contained in an F_σ of a special type, while a residual set contains a G_δ. Much of the utility of this notion comes from the *Baire Category Theorem*, which guarantees in fact a *dense* G_δ subset:

Theorem 5.6. *Let X be a Polish space. Then a residual set is dense.*

Proof. Let G_i be open and dense, for $i = 1, 2, \ldots$. We shall show that

$$E = \cap_{i=1}^{\infty} G_i$$

is dense. Let $\mathcal{U}$ be an open subset of X, and assume that $d(\cdot, \cdot)$ is a metric compatible with the topology of X, for which X is complete. Since G_1 is dense, there exists $x_1 \in \mathcal{U} \cap G_1$, and there exists $\delta_1 > 0$ such that for the ball of that radius, $\overline{B_{\delta_1}(x_1)} \subseteq \mathcal{U} \cap G_1$ where $\overline{B}$ indicates the closure of B. Now there exists $x_2 \in B_{\delta_1}(x_1) \cap G_2$ and $\delta_2 > 0$

such that $\overline{B_{\delta_2}(x_2)} \subseteq B_{\delta_1}(x_1) \cap G_2$. Continuing in this manner, and choosing δ_n so as to decrease to 0, the sequence $(x_i)_{i\geq 1}$ is Cauchy; by completeness this sequence has a limit point x, and the fact that the closure of each ball is contained in the previous set guarantees that $x \in \mathcal{U} \cap G_k$ for all k. □

Exercise 5.4. Show that the $\mathbb{Q}$, the set of rational numbers, is not a Polish space.

Definition 5.4. Let T be a homeomorphism of X. A point $x \in X$ is **transitive** iff it has a dense orbit. The map T is transitive iff there exists a transitive point.

If T is continuous but not necessarily invertible, we say a point is **forward transitive** iff it has a dense forward orbit, and the map is forward transitive iff there exists a forward transitive point.

Proposition 5.7. *Let $(X, \mathcal{T})$ be a Polish space with no isolated points.*
(i)Let T be a homeomorphism. Then if T is transitive, the set E of forward transitive points is residual.
(ii)Let T be a continuous map. Then if T is forward transitive, the set E of forward transitive points is residual.

We note that in (i), by having biinfinite orbits in the hypothesis and forward orbits in the conclusion, the statement is stronger in both respects. That is, the existence of a single biinfinitely transitive point implies existence of (many) forward transitive points: a residual set hence (by the Baire Category Theorem) a dense G_δ of them. Without the assumption of no isolated points this can fail, as shown by a simple example on p. 129 of [Wal82], of a homeomorphism with a dense biinfinite orbit but no dense forward orbit (imagine the left shift map $n \mapsto n-1$ on $\mathbb{Z}$, extended continuously to its two-point compactification $\mathbb{Z} \cup \{-\infty, +\infty\}$ by declaring $\pm\infty$ to be fixed points).

Proof. With metric d as above, since X is a separable metric space there exists a countable base $\{\mathcal{U}_i\}_{i\geq 1}$ for the topology. Then E is the set of points x such that for each $j \geq 1$, the forward orbit of x meets $\mathcal{U}_j$. That is, for each j, $E \subseteq G_j \equiv \cup_{n\geq 0} T^{-n}(\mathcal{U}_j)$, so we can write:

$$E = \inf G_j = \limsup_{j\geq 1} \sup_{n\geq 0} T^{-n}(\mathcal{U}_j) = \cap_{j\geq 1} \cup_{n\geq 0} T^{-n}(\mathcal{U}_j).$$

We claim that each of the open sets G_j is dense. We wish to show that for each $i \geq 1$, $\mathcal{U}_i$ meets G_j. Now there exists a transitive point w; that is, for (i), the biinfinite orbit $(T^n(w))_{n\in\mathbb{Z}}$ is dense; for (ii) we know this for the forward orbit. Furthermore, since X has no isolated points this collection of points must be infinite. Now any dense infinite sequence of distinct points must meet an open set $\mathcal{U}$ infinitely often: singletons are closed sets in a metric space, so $\mathcal{U} \setminus \{x\}$ is again nonempty open and we can find the next such element. Given $i, j \geq 0$, therefore, in either case, the orbit of w enters both $\mathcal{U}_i$ and $\mathcal{U}_j$ infinitely often, one of them first. If we know w is forward transitive, then from this we know there is a pair of times such that $\mathcal{U}_i$ occurs first. That is, there exists a point x and an $k > 0$ such that $x \in \mathcal{U}_i$ and $T^k(x) \in \mathcal{U}_j$, equivalently, $x \in \mathcal{U}_i \cap T^{-k}(\mathcal{U}_j) \subseteq \mathcal{U}_i \cap G_j$. Thus $\mathcal{U}_i \cap G_j$ is nonempty and hence G_j is dense as claimed, so E is a countable intersection of open dense sets and hence is residual.

If we only know w is biinfinitely transitive, we have to be slightly more careful. Now if $\mathcal{U}_i$ occurs first, the rest of the argument is the same. But for a general argument, we know there exists $n \geq 0$ such that either $\mathcal{U}_i \cap T^n(\mathcal{U}_j)$ or $\mathcal{U}_j \cap T^n(\mathcal{U}_i)$ is nonempty. Call this set $\mathcal{U}$ in either case. We claim that there exists $k > 0$ such that $\mathcal{U}_i \cap T^k(\mathcal{U}_j)$ is nonempty, and then will proceed as before. But the transitive point enters $\mathcal{U}$ infinitely many times, so there exists $m > 0$ and x such that $x \in \mathcal{U}$ and $T^m(x) \in \mathcal{U}$, and we take simply $k = m$ in the first case, or $k = m - n$ in the second and then proceed as before. □

We remark that something similar can be proved for group actions; see §51.

6. Analysis background I: Dual spaces give coordinates; L^p spaces

In this section we first develop some analysis tools which will be needed throuhout the notes. This includes reviewing some basics on Fourier series and transforms, and on dual spaces.

6.1. **Duality: Why "functional" analysis?** *Functional Analysis* could be defined to be the study of analysis on infinite-dimensional vector spaces. Generally, these are *function spaces*, Now the study of functions brings in all the richness of calculus and analysis: derivatives, integrals, measures, series. And since all of that involves limiting operations, we shall need first of all a topology on the space. (Indeed, there are texts with the alternate title *Topological Vector Spaces* (TVS) [Bou13], or *Linear Topological Spaces* [KN$^+$63]).

What makes this study so fascinating and useful is reflected in the fact that, unlike for finite dimensions, there can be various possible choices for this topology, and moreover it will be important to study several topologies on the same space.

But then why is the subject called *Functional* Analysis rather than perhaps Function Space Analysis? That is because of the key role played by the dual V^* of the vector space V. In infinite dimensions we define V^* not to be all linear functionals on V, but instead all *continuous* linear functionals. The role of V^* is then to *coordinatize* V: the V^*-coordinates of $\mathbf{v} \in V$ are all the values $\lambda(\mathbf{v})$ such that $\lambda \in V^*$. If we think of the finite-dimensional case, then choice of a basis, for example the standard basis $(\mathbf{e}_i)_{i=1}^n$ in Euclidean n-dimensional space defines three dual basis vectors $(\lambda_i)_{i=1}^n$ via the inner product, $\lambda_i(\mathbf{v}) = \mathbf{e}_i \cdot \mathbf{v} \equiv v_i$, and this defines an isomorphism from V to the *coordinate space* $\mathbb{R}^n$ via $\mathbf{v} \mapsto (v_1, \ldots, v_n)$. This choice of basis vectors has coordinatized the space. Now we can transport all we know about analysis on $\mathbb{R}$ and $\mathbb{R}^n$ to V by this correspondence. In the same way, for the infinite dimensional space V we do analysis on V via analysis on the coordinate values given by the functionals. However now it is easier to simply take all of V^*, rather than trying to select an efficient subset, i.e. a basis, as this step not only is not necessary but may indeed not be possible (!).

This is a rough sketch, which becomes more concrete when we encounter the norm, weak and weak-star topologies on a given space. Identifying the dual space of some given TVS occupies a central part of the classic [DS57] which includes a large table of spaces and their duals; here "identify" may mean describing the dual in terms of

some other space of functions or measures. Proving these theorems always goes deep and tells us a lot about the spaces involved.

Functional Analysis thus provides both an essential and powerful collection of tools and a clarifying viewpoint on many questions in Analysis, with key applications in geometry, physics, probability and differential equations.

So to look at Functional Analysis we first need to recall some basics of the finite dimensional case, before we can dip into the even more fascinating and richer world of infinite dimensions.

We consider a vector space V; for our purposes, the scalar field will be $\mathbb{C}$ or $\mathbb{R}$. For those used to real vector spaces, the main difference is the formulas is that an inner product has to take account of complex conjugates, as we shall explain.

Let us recall:

Definition 6.1. A *norm* $||\cdot||$ on V is a function with values in $\mathbb{R}$ which satisfies:
(i) $||a\mathbf{v}|| = |a| \cdot ||\mathbf{v}||$ (homogeneity);
(ii) $||\mathbf{v} + \mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}||$ (triangle inequality);
(iii) $||\mathbf{v}|| \geq 0$, and $||\mathbf{v}|| = 0$ iff $\mathbf{v} = \mathbf{0}$. (positive definiteness).

Having a norm of course allows us to define a metric space structure on V, with the distance between points defined by $d(\mathbf{v}, \mathbf{w}) = ||\mathbf{w} - \mathbf{v}||$.

Definition 6.2. When the field is $\mathbb{R}$, an *inner product* is a function from $V \times V$ to $\mathbb{R}$, written $\mathbf{v} \cdot \mathbf{w}$ or $\langle \mathbf{v}, \mathbf{w} \rangle$, satisfying the following;
(1) $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$ (commutative law);
(2) $(a\mathbf{v}) \cdot \mathbf{w} = a(\mathbf{v} \cdot \mathbf{w})$ (associativity of scalar multiplication)
(3) $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$ (distributive law)
$(4a)$ $\mathbf{v} \cdot \mathbf{v} \geq 0$ and
$(4b)$ If $\mathbf{v} \cdot \mathbf{v} = 0$ then $\mathbf{v} = \mathbf{0}$.

These imply that also:
$(2')$ $\mathbf{v} \cdot (a\mathbf{w}) = a(\mathbf{v} \cdot \mathbf{w})$.
$(3')$ $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$.

Properties $(2, 2'), (3, 3')$ tell us that this is a *bilinear form*; (1), $(4a)$, and $(4b)$ add that the form is *symmetric*, *positive* and *positive definite*. See §35.6 below. Note that a positive definite bilinear form defines a norm, via

$$||\mathbf{v}|| \equiv (\mathbf{v} \cdot \mathbf{v})^{1/2}.$$

When the field is $\mathbb{C}$, one replaces that with the notion of an *Hermitian inner product*: all axioms are as above except now (1) are exchanged for:
$(\overline{1})$ $\mathbf{v} \cdot \mathbf{w} = \overline{\mathbf{w} \cdot \mathbf{v}}$ (conjugate-symmetry). This implies that $(2a, b)$ are replaced by:
$(\overline{2a})$ $(a\mathbf{v}) \cdot \mathbf{w} = a(\mathbf{v} \cdot \mathbf{w})$ (just like $(2a)$) but now
$(\overline{2b})$ $\mathbf{v} \cdot (a\mathbf{w}) = \overline{a}(\mathbf{v} \cdot \mathbf{w})$.

Since the Hermitian definition reduces to the real one when the field is $\mathbb{R}$, sometimes in the literature one begins by defining an inner product via the Hermitian axioms for both $\mathbb{C}$ and $\mathbb{R}$; see [Axl97].

We note that from $(\overline{1})$ $\mathbf{v} \cdot \mathbf{v} \in \mathbb{R}$; from $(4a, b)$ we have as before that we have a norm, with $||\mathbf{v}||^2 = \mathbf{v} \cdot \mathbf{v}$.

An example is $\mathbb{C}^n$, where the standard Hermitian inner product of $\mathbf{v} = (v_1, \cdots, v_n)$ and $\mathbf{w} = (w_1, \cdots, w_n)$ is defined to be

$$\mathbf{v} \cdot \mathbf{w} = \sum v_i \overline{w}_i. \tag{3}$$

To verify $(4a, b)$ note that for $n = 1$ then with $\mathbf{v} = z \in \mathbb{C}$ we have $||\mathbf{v}|| = (z\overline{z})^{\frac{1}{2}} = |z|$, while for general n,

$$||\mathbf{v}|| = \left(\sum v_i \overline{v}_i \right)^{\frac{1}{2}} = \left(\sum |v_i|^2 \right)^{\frac{1}{2}}$$

.

In a real vector space one defines the angle θ between two vectors $\mathbf{v}$ and $\mathbf{w}$ via the equation

$$\mathbf{v} \cdot \mathbf{w} = ||\mathbf{v}||||\mathbf{w}|| \cos\theta.$$

Note that since $\cos(\theta) = \cos(-\theta) = \cos(2\pi - \theta)$ this does not depend on the order in which we choose the vectors with respect to some orientation chosen on the plane containing $\mathbf{v}, \mathbf{w}$.

When the field is $\mathbb{C}$, we define orthogonality in the same way: $\mathbf{v}, \mathbf{w}$ are orthogonal iff $\mathbf{v} \cdot \mathbf{w} = 0$. More generally one defines the *Hermitian angle* between two vectors by the equation

$$|\mathbf{v} \cdot \mathbf{w}| = ||\mathbf{v}||||\mathbf{w}|| \cos\theta$$

so $-\pi/2 \leq \theta \leq \pi/2$.

One could also consider the angle between the vectors as elements of the real vector space $\mathbb{R}^{2n}$, but this will in general give a different number. That is already true for $\mathbb{C}^1$, since e.g. $z = a + bi$ and $w = -b + ai$ are orthogonal as vectors in $\mathbb{R}^2$, yet $z\overline{w} \neq 0$. (Indeed, in any one-dimensional vector space with a Hermitian inner product, no two nonzero vectors can be orthogonal.) So, we have to be a bit careful with our intuition of what angle means here!

Writing K for our field, for finite dimensions, V is isomorphic to K^n. Now if we have a norm on V, for instance the Euclidean norm, the L^1 or L^∞ norm, then a key point is that all norms are equivalent, meaning that they all give the same topology. See Lemma 35.41.

That is decidedly *not* the case for infinite dimensions, which in fact is exactly what makes functional analysis such a useful and fascinating subject. Here we sketch an explanation, of how one can visualize infinite dimensions, and of how one can approach these different possible topologies.

Now for an infinite dimensional space, one proves (by the Axiom of Choice, or equivalently by Zorn's Lemm–exercise!!) that there always exists an algebraic basis, called a *Hamel basis* B. This makes V isomorphic to an abstract space of functions, since one can identify $\mathbf{v}$ with its coordinates $(\lambda(\mathbf{v}))_{\lambda \in B}$.

The subject of Functional Analysis is the study of infinite dimensional spaces, hence spaces of functions. But rather than algebraically via the Hamel basis, we impose some interesting, useful or natural topology on V. Thus a *topological vector space* V is a vector space with a topology $\mathcal{T}$ such that the operations $+$ and $\cdot$ (multiplication by a scalar) are continuous. The key to the study of $(V, \mathcal{T})$ is to replace the Hamel

basis by the *dual space* V^* of $(V, \mathcal{T})$, by definition the set of all continuous linear functionals on V.

For our function space V, there are many possible choices of topology, some given by (now nonequivalent!) norms, hence many possible dual spaces. This leads to many of the different notions of convergence of sequences of functions, which one encounters in analysis.

Choice of a basis in finite dimensions gives coordinates on V, as it defines an isomorphism to K^n, which can then be thought of as a space of coordinates for V.

For infinite dimensions, there may not exist a good choice of basis. So instead one takes all of V^*, and considers for $\mathbf{v} \in V$, the entire set of coordinates

$$\{\lambda(\mathbf{v}) : \lambda \in V^*\}.$$

This is overkill, but works out very well.

The idea of *Functional Analysis* is to do analysis with these functionals. In other words, properties we want to study are examined by way of the coordinates defined by V^*.

Recalling that an inner product is denoted by $\langle \mathbf{v}, \mathbf{w} \rangle$, we more generally write

$$\langle \lambda, \mathbf{v} \rangle = \lambda(\mathbf{v}).$$

This is a *pairing* of V and V^*, as it is a bilinear function on $V^* \times V$.

In finite dimensions, V is isomorphic to V^*. This isomorphism is not *natural*, as it is not unique: it depends on the choice of an inner product on V, as follows Given such an inner product, written $\langle \mathbf{v}, \mathbf{w} \rangle$ or $\mathbf{v} \cdot \mathbf{w}$, then we define the isomorphism from V to V^* by

$$\mathbf{v} \mapsto \langle \mathbf{v}, \cdot \rangle.$$

Then, choosing a basis $(\mathbf{e}_1, \ldots \mathbf{e}_n)$ for V, we consider the *dual basis* $(\langle \mathbf{e}_1, \cdot \rangle, \ldots, \langle \mathbf{e}_n, \cdot \rangle,)$ for V^*. Then the map $\mathbf{v} \mapsto (\mathbf{e}_1 \cdot \mathbf{v}, \ldots, \mathbf{e}_n \cdot \mathbf{v})$ defines an isomorphism from V to K^n.

This gives the coordinates for $\mathbf{v}$ in terms of the basis of V, or equivalently in terms of the dual basis for V^*.

6.2. L^p **spaces and Fourier series.** Some of the principal examples are the L^p and l^p spaces.

Given a measure space $(X, \mathcal{A}, \mu)$ then for $0 < p < \infty$, the space $L^p = L^p(X, \mathcal{A}, \mu)$ is defined to be the vector space of all $f : X \to K$ such that $\int_X |f|^p \mathrm{d}\mu < \infty$. We set

$$||f||_p = \left(\int_X |f|^p \mathrm{d}\mu \right)^p.$$

For $p = \infty$ we make the special definition that

$$||f||_\infty = \sup |f(x)|.$$

These are norms, and make L^p into a *Banach space*, by definition a complete normed topological vector space.

If q is such that

$$\frac{1}{p} + \frac{1}{q} = 1,$$

then p, q are called *conjugate exponents*.

Particularly important examples are where X is $[0,1]$ or $[0,2\pi]$ or $\mathbb{R}^n$, with Lebesgue measure, or where X is a subset of the integers $\mathbb{Z}$ and μ is counting measure. In this last case one writes l^p for the space of sequences $L^p(\mathbb{Z},\mu)$. Thus for an index set $\mathcal{I} \subseteq \mathbb{Z}$, $l^2(\mathcal{I}) = \{\underline{a} = (a_0, a_1, \dots) : ||\underline{a}|| < \infty\} \subseteq \mathbb{R}^{\mathcal{I}}$ where $\langle \underline{a}, \underline{b}\rangle = \sum_{\mathcal{I}} a_i \overline{b}_i$ and so $||\underline{a}||_2 = \left(\sum_{\mathcal{I}} a_i^2\right)^{\frac{1}{2}}$.

Note that for index set $\mathcal{I} = \{1, 2, \dots, d\}$ this is the usual Euclidean norm on $\mathbb{R}^d$.

We note that in probability terms, $\mathbb{E}(|f|) = ||f||_1$ while $\mathrm{var}(f) = ||f - \mathbb{E}(f)||_2$. More generally, f has finite p^{th}moment iff it is in L^p, since $\mathbb{E}(f^p) = (||f||_p)^p$.

Theorem 6.1.
(i) For $0 < p < \infty$, the dual space of the Banach space L^p is L^q. For L^1 the dual space is L^∞. As above, for $f \in L^p$ and $g \in L^q$, we define the pairing $\langle f, g\rangle = \int_X f\overline{g}\, d\mu$.
(ii) Since L^2 is self-dual, $\langle f, g\rangle$ defines a (Hermitian) inner product. This makes L^2 into a Hilbert space*, i.e. a complete inner product space, with norm*

$$||f|| = \langle f, f\rangle^{\frac{1}{2}} = \left(\int_X f\overline{f}\, d\mu\right)^{\frac{1}{2}}.$$

(iii) The dual space of $L^\infty(X)$ where X is a finite measure space is $bca(X)$, the space of bounded countably additive signed measures. For X an infinite measure space, it is $ba(\mathbb{R})$, the bounded finitely additive signed measures. Note that L^1 embeds naturally in this dual space in both cases.

For the most classical example of an infinite-dimensional Hilbert space we take the measure space X to be the one-dimensional torus (i.e. the circle) $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$. We let m denote normalized Lebesgue measure so $\mu(\mathbb{T}) = \mu[0, 2\pi) = 1$; that is, $\mathrm{d}m = 1/2\pi \mathrm{d}x$, and consider the complex or real field K. Then, given $f : \mathbb{T} \to K$, we identify this with the corresponding 2π-periodic function on $\mathbb{R}$, also written f.

For the space $L^2(\mathbb{T})$ we then have the Hermitian inner product $\langle f, g\rangle = \int_0^{2\pi} f\overline{g}\mathrm{d}m$. We define $e_n(x) = e^{2\pi i n x}$ for $n \in \mathbb{Z}$.

Lemma 6.2. *For the Hilbert space $L^2(\mathbb{T})$, the complex exponentials $(e_n)_{n\in\mathbb{Z}}$ provide an orthonormal basis.*

We write $a_n = \widehat{f}(n) = \langle e_n, f\rangle$, the *Fourier coefficients* of f.

Now for $f \in L^2$, the *Fourier series*

$$\sum_{n\in\mathbb{Z}} \widehat{f}(n) e_n \tag{4}$$

converges and equals f.

More generally, for other function spaces (e.g. L^1) one wishes to know whether and in what sense the Fourier series converges to f.

A further question is: considering a given Banach space of functions $\mathcal{L}$, what is the corresponding space $\widehat{\mathcal{L}}$ of coefficients, and vice-versa: given a space of sequences, what is its image by formula (4)?

The Fourier series ia also called the Fourier *expansion* of f. One thinks of this in two quite different ways: *geometrically,* as the expression of a vector in terms of an orthonormal basis in Hilbert space; and *physically*, as providing a *harmonic analysis*

of the function. Thus in (4) we have written a periodic signal or vibration f as a sum of waves which are harmonically related (as the frequencies n are integer multiples of the basic frequency 1). These are *complex* waves, i.e. spirals (or more accurately, helices) but can be decomposed as *real* waves as we explain.

We set

$$\widehat{f}(n) = a_n = \int_0^{2\pi} e_n f \mathrm{d}m = \int_0^{2\pi} e^{-2\pi i n x} f(x) \mathrm{d}m.$$

This is the n^{th} *Fourier coefficient* of f.

Given $\underline{a} = (a_n)_{n\in\mathbb{Z}} \in l^2$, we define $f : \mathbb{T} \to \mathbb{C}$ by $\check{T}(\underline{a}) = f$ where

$$f(x) = \sum_{n\in\mathbb{Z}} a_n e_n.$$

Theorem 6.3. *The map $T : f \mapsto \widehat{f} = \underline{a} = (a_n)_{n\in\mathbb{Z}}$ is an isometry from $L^2(\mathbb{T})$ to $l^2(\mathbb{Z})$. That is, $\langle f, g\rangle = \langle \widehat{f}, \widehat{g}\rangle$. The inverse map is given by $\underline{a} \mapsto \check{T}(\underline{a}) = \sum_{n\in\mathbb{Z}} a_n e_n$.*

The connection with (real-valued) waves is seen via *Euler's formula*

$$e^{i\theta} = \cos\theta + i \sin\theta,$$

from which it follows that
$\frac{1}{2}(e^{i\theta} + e^{-i\theta}) = \cos\theta$; $\frac{1}{2}(e^{i\theta} - e^{-i\theta}) = \sin\theta$ whence if $a_n = a_{-n}$ then

$$\frac{1}{2}\sum_{-N}^{N} a_n e_n = \sum_{0}^{N} a_n \cos(n\theta)$$

while if $b_n = -b_{-n}$ then

$$-i \cdot \frac{1}{2}\sum_{N}^{N} b_n e_n = \sum_{0}^{N} b_n \sin(n\theta).$$

Thus given $(a_n)_{n\geq 0}$, $(b_n)_{n\geq 0}$ and defining a_n, b_n for $-n$ by $a_{-n} = a_n$ and $b_{-n} = -b_n$, and setting for $n \in \mathbb{Z}$, $c_n = a_n - ib_n$, we have that

$$\sum_{0}^{N} a_n \cos(n\theta) + \sum_{0}^{N} b_n \sin(n\theta) = \frac{1}{2}\sum_{-N}^{N} c_n e_n.$$

In this way, any series in *cos* and *sin* can be realized as a Fourier series, in a unique way.

This is for a_n, b_n complex, but includes the case of real Fourier series, which by definition is such a series with $(a_n)_{n\geq 0}$, $(b_n)_{n\geq 0}$ real.

Conversely, any $\mathbb{C}$-valued fuction f can be uniquely decomposed into its real and imaginary parts $f^{\mathbb{R}} = Re(f), f^{Im} = Im(f) : \mathbb{R} \to \mathbb{R}$, with $f = f^{\mathbb{R}} + if^{Im}$. Consider $(d_n)_{n\in\mathbb{Z}}$ with $d_n \in \mathbb{C}$, and decompose it as $d_n = d_n^{\mathbb{R}} + id_n^{Im}$.

Let c_n be one of these. Now since (exercise!) since any function f on $\mathbb{R}$ or Z can be decomposed uniquely into its even and odd parts, $f = f_e + f_o$, given $(c_n)_{n\in\mathbb{Z}}$ with $c_n \in \mathbb{R}$, uniquely write $a_n = b_n + c_n$.

Thus we can separate a series

$$\frac{1}{2}\sum_{-N}^{N} c_n e_n$$

into four series in sin and cos with real coefficients, in a unique way. $\sum_N^N a_n e_n = \sum_0^N b_n \cos(n\theta) + \sum_0^N c_n \sin(n\theta)$.

We define μ to be counting measure on $\mathbb{Z}$, and $l_2 = L^2(\mathbb{Z}, \mu)$ to be all $\mathbb{C}$-valued sequences $(a_n)_{n\in\mathbb{Z}}$ such that $\sum |a_n|^2 < \infty$. This is a Hilbert space with Hermitian inner product $\langle a_n, b_n \rangle = \sum a_n \bar{b}_n$. (See (129)). Thus the norm of this Banach space is $||a(\cdot)||_2 = (\sum |a_n|^2)^{1/2}$.

Theorem 6.4.

6.3. **Taylor series, Fourier Series and Laurent Series.** We just want to touch here on the beautiful relationship between these notions.

A *real power series* is

the radius of convergence is

MacLaurin's Series, or the Taylor series at 0, is

Examples are

7. Analysis background II: signed measures as dual spaces; Riesz representation, Krein-Milman and Choquet; existence of invariant measures, the ergodic theorem, generic points, mixing

At this point it will be healthy to recall a bit more of fundamental measure theory, by way of functional analysis.

Let (X, d) be a compact metric space, and $\mathcal{C}(X) = \mathcal{C}(X, \mathbb{R})$ the space of continuous real-valued functions on X. We give $\mathcal{C}(X)$ the sup norm, $||f||_\infty = \sup_{x\in X} |f(x)|$, with respect to which it is a Banach space. Then ....

Radon-Nikodym theorem

product spaces (top and measure) : Tychonoff theorem

-product measure

separability of space of continuous functions

dual of continuous functions on interval and reals (Riesz (Markov Kakutani) representation) Rudin R and C, D and S for ftly add...

Stone-Weir: Kelley Top

used by Nelson for Brownian M

Weak* topology

Banach-Alaoglu: X compact metric then prob measures are weak* compact Walters82 p. 150).

dual space of uniformly bdd cts functions on Polish space;

tightness

Banach-Alaoglu and Krein-Milman;

Choquet theorem

7.1. **Existence of invariant measures.**

Theorem 7.1. *Let $(X, \mathcal{A})$ be a measurable space (a set X with a $\sigma-$algebra $\mathcal{A}$).*
Let $T : X \to X$ be measurable.
The transformation is ergodic iff each invariant measurable real-valued function is a.s. constant. [Bil65] *p. 13. (This statement is true also for $\mathbb{C}$-valued functions, since $f = u + iv$ is invariant iff u and v are.)*
We call μ an ergodic measure *iff the transformation $(X, \mathcal{A}, T)$ is ergodic.*
(i) (see Billingsley [Bil65] *p. 38 ff.) Two ergodic probability measures $\mu, \widehat{\mu}$ are either identical or mutually singular.*
(ii) the collection $\mathcal{P}_{\mathcal{A}}$ of invariant probability measures is a convex set, whose extreme points are exactly the ergodic invariant measures.

If $(X, \mathcal{T})$ is a topological space and $T : X \to X$ is continuous, then we write $\mathcal{M}(X, T)$ for the collection of all probability Borel measures, and give this the weak- topology. Then:*
(iii) (Krylov-Bogliubov) If X is compact metric, then $\mathcal{M}(X, T)$ is a nonempty convex compact set. (See e.g. of [Wal82] *p. 152).*
(iv) (Fomin) [**?**] *In fact the statement of (ii) holds if $(X, \mathcal{T})$ is a Polish space (Def. 5.2).*
(v) (Krein-Milman) Any element of $\mathcal{M}(X, T)$ can be expressed in a unique way as an (integral) convex combination of extremem points. That is,....
(v) (Choquet), see [Phe01]

Proof. □

Definition 7.1. A topological transformation (or flow) is **uniquely ergodic** iff there exists a unique invariant probability measure.

generic points

Definition 7.2. Recall that when $(X, \mathcal{T})$ is a topological space with Borel probability measure μ and T a measure-preserving transformation on X, then $x \in X$ is a *generic point* for T iff for each $F \in \mathcal{C}(X)$, the ergodic averages converge to the expected value:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} F \circ T^n(x) = \int_X F d\mu. \tag{5}$$

One has the following important, easy to prove but at first surprising result:

Theorem 7.2. *(Krylov-Bogliubov) If $(X, \mathcal{T})$ is a compact metric space, with probability measure μ and continuous, ergodic measure-preserving map T, then μ-a.e. point is a generic point.*

Proof. By the Birkhoff ergodic theorem, for every continuous function $F \in C(X)$ there is a full measure set $E_F \subseteq X$ such that for every $x \in E_F$ (116) holds true. Since X is compact, $C(X)$ is separable (for the sup norm). Let $\mathcal{D} = \{F_i\}_{i\in\mathbb{N}}$ be a dense subset of $C(X)$. Then (116) holds for every x in $E = \cap_{i\in\mathbb{N}} E_{F_i}$, simultaneously for every $F \in \mathcal{D}$. Now every $F \in C(X)$ is ε-uniformly approximated by $F \in \mathcal{D}$. So

both sides of (116) are within ε hence the equation holds for every $F \in C(X)$; that is, μ-almost every x is a generic point. □

Theorem 7.3. *(Fomin) The same holds if $X, \mathcal{T}$) is a Polish space (a topological space such that htere exists an equivalent metric which makes this a complete separable metric space).*

This is [Fom43], see below (???)

Exercise 7.1. *Show that for the rotation $R_{\mathbf{v}}$, minimality is equivalent to unique ergodicity.*

TO DO...
existence of invariant measure for compact space
Furst defn of amenable
ergodicity and extreme points, ergodic decomposition
Lebesgue spaces; example: Polish spaces.
finitely additive measures and compactness
unique ergodicity
Hint: For the flow part you may need .... (flow cross-section)
Extend these results to the $d-$torus.

Theorem 7.4. *(Birkhoff 1931) Let T be a measure-preserving transformation of a probability space $(X, \mathcal{A}, \mu)$, and let $f \in L^1(X, \mu)$. Then there exists an invariant L^1 function $\overline{f}$ and an invariant set of full measure $X_1 \subseteq X$ such that*

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) \to \overline{f}$$

for all $x \in X_1$ and $\int_X f\, d\mu = \int_X \overline{f}\, d\mu$.

In particular, if T is ergodic then $\overline{f} = \int_X f \mathrm{d}\mu$ is constant, so we have the famous statement "time average= space average":

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) \to \int_X f \mathrm{d}\mu \text{ almost surely.}$$

Given $(X, \mathcal{A}, \mu, T)$ as above, we define a linear operator U on $L^2(A, \mu)$ (with complex values) by $U(f) = f \circ T$; this is the **Koopman operator**. Since T preserves μ, $\langle Uf, Ug\rangle = \langle f, g\rangle$ i.e. U is a unitary operator.

As a corollary of Birkhoff's theorem one has von Neumann's L^2 (mean) ergodic theorem (since convergence a.s. implies convergence in L^2):

Theorem 7.5. *(von Neumann 1932) Let T be a measure-preserving transformation of a probability space $(X, \mathcal{A}, \mu)$, and let $f \in L^2(X, \mu)$. Then for $\overline{f}$ the projection of f to the subspace of invariant functions, we have*

$$\left|\left|\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) - \overline{f}\right|\right|_2 \to 0.$$

Remark 7.1. In fact despite the dates of the articles, von Neumann's theorem preceded Birkhoff's and inspired it- Birkhoff was the editor at the *Annals* to whom von Neumann sent his article (!). See [Bir31], [Neu32b], [Neu32a].

–choosing a point from circle by dynamics
Krylov- Bougliobov

8. Mixing, weak mixing and ergodicity

Definition 8.1. Given a measure-preserving transformation T of a probability space $(X, \mathcal{A}, \mu)$, T is **mixing** iff for every $A, B \in \mathcal{A}$ we have
(i)

$$\mu(A \cap T^{-n}B) \to \mu A \mu B \text{ as } n \to \infty.$$

It is **weak mixing** iff
(ii)

$$\frac{1}{N}\sum_{n=0}^{N-1} |\mu(A \cap T^{-n}B) - \mu A \mu B| \to 0 \text{ as } N \to \infty.$$

A third related condition, which we come back to, is
(iii)

$$\frac{1}{N}\sum_{n=0}^{N-1} \mu(A \cap T^{-n}B) \to \mu A \mu B \text{ as } N \to \infty.$$

Remark 8.1. If T is invertible, the apparent time asymmetry is illusory, since the above statements with limit at $+\infty$ can be replaced equivalently by limits at $\pm\infty$, since $\mu(A \cap T^{-n}B) = \mu(T^n(A \cap T^{-n}B)) = \mu(T^n(A) \cap B)$.

Proposition 8.1. *Each of* (i), (ii), (iii) *implies ergodicity.*

Proof. Suppose A is invariant. Then taking $B = A$ in (i), we have $\mu(A) = \mu(A \cap T^{-n}A) \to (\mu A)^2$ as $n \to \infty$, so for $x = \mu A$, $x = x^2$, $x^2 - x = x(x-1) = 0$ leaving $x = 0$ or $x = 1$, and A is trivial. The same proof works assuming (ii) or (iii). □

Lemma 8.2. *Let* $(X, \mathcal{A}, \mu, T)$ *be a measure-preserving transformation of a probability space. Then for any* $A, B \in \mathcal{A}$, *we have*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \mu(A \cap T^{-n}B) = \int_X \chi_A \overline{\chi}_B \, d\mu$$

where $\overline{\chi}_B$ *is the projection of* χ_B *to the space of invariant functions.*

Proof. By the Birkhoff ergodic theorem,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{0}^{N-1} \chi_B(T^n(x)) \to \overline{\chi}_B$$

for a.e. x, where by the projection to the space of invariant functions we mean that for any invariant subset E, $\int_E \overline{\chi}_B \mathrm{d}\mu = \mu(B \cap E)$. Multiplying both sides of the equation

by the function χ_A, we still have almost-sure convergence:

$$\chi_A\left(\frac{1}{N}\sum_0^{N-1}\chi_B\circ T^n\right)\to\chi_A\overline{\chi}_B$$

Making use of the Lebesgue dominated convergence theorem (i.e. that a.s. convergence for bounded functions implies the integrals converge), while noting that $\chi_A(\chi_B\circ T^n)=\chi_A\chi_{T^{-n}B}=\chi_{A\cap T^{-n}B}$, yields

$$\frac{1}{N}\sum_0^{N-1}\mu(A\cap T^{-n}B)\to\int_X\chi_A\overline{\chi}_B\mathrm{d}\mu.$$

□

Proposition 8.3. *In fact (iii) of Definition 41.1 is equivalent to ergodicity.*

Proof. We have already proved that (*iii*) implies ergodicity. Now assuming T is ergodic, then in the proof of Lemma 8.2, $\overline{\chi}_B$ is the constant function $\mu(B)$, so by the lemma, $\lim\frac{1}{N}\sum_{n=0}^{N-1}\mu(A\cap T^{-n}B)=\mu(A)\mu(B)$. □

Exercise 8.1. Show that the odometer transformation is not weak mixing. Show the same for an irrational circle rotation.

We have already shown that, separately, mixing and weak mixing imply ergodic, but it is worth noting that one now has a different argument:

Proposition 8.4. *Mixing implies weak mixing implies ergodic.*

Proof. For a sequence $(a_k)_{k=0}^{\infty}$, then $a_k\to 0$ iff $|a_k|\to 0$, which certainly implies that $\frac{1}{n}\sum_{k=0}^{n-1}|a_k|\to 0$. Furthermore $|\frac{1}{n}\sum_{k=0}^{n-1}a_k|\leq\frac{1}{n}\sum_{k=0}^{n-1}|a_k|$. Therefore in Def. 41.1, taking $a_k=\mu(A\cap T^{-k}B)-\mu A\mu B$ we have $(i)\implies(ii)\implies(iii)$. □

Exercise 8.2. Show, directly from the definition, that the odometer transformation is ergodic but is not weak mixing. Show the same for an irrational circle rotation.

Lemma 8.5. *For a bounded complex-valued sequence $(a_k)_{k\geq 0}$, these are equivalent:*
(a)

$$\frac{1}{n}\sum_{k=0}^{n-1}|a_k|\to 0$$

(b)

$$\frac{1}{n}\sum_{k=0}^{n-1}|a_k|^2\to 0$$

(c)

$$a_k\to 0 \text{ in density },$$

i.e. along a set of times of density one.

Proof. As noted above, clearly $a_k \to 0 \implies \frac{1}{n}\sum_{k=0}^{n-1} a_k \to 0$. It follows that for a_k bounded, $(c) \implies (a)$. To show $(a) \implies (c)$, we shall construct a set $J \subseteq \mathbb{N}$ of density zero such that $a_k \to 0$ for $n \notin J$. A naive approach would be the following. We know that for each $m \in \mathbb{N}$ the set $J_m = \{n : |a_n| > 1/m\}$ has density zero; we could hope that $J = \cup_{m\geq 1} J_m$ will work. The problem with this argument is that $\cup J_m$ may not have density zero (one way of looking at this is that density only leads to a *finitely* additive measure on the integers; see §33.1). Instead, noting that $J_1 \subseteq J_2 \subseteq ...$, we define J to be J_1 on a first interval $[0, n_1] \cap \mathbb{N}$, then J_2 on $[n_1 + 1, n_2] \cap \mathbb{N}$, and so on. We choose this increasing sequence n_m so that a density estimate of $1/m$ has kicked in for the set at the next stage, J_{m+1}, at that point. That is, for all m, for all $L \geq n_m$, $\frac{1}{L}\#\{J_{m+1} \cap [0, L]\} < 1/m$. Then for $n_m < L \leq n_{m+1}$ we have $J \cap [0, L] \subseteq J_{m+1} \cap [0, L]$ so $\frac{1}{L}\#\{J \cap [0, L]\} \leq \frac{1}{L}\#\{J_{m+1} \cap [0, L]\} \leq 1/m$, so J does have density zero, and $a_n \to 0$ off of J as desired.

Lastly, from the equivalence of (a) and (c), applied instead to the sequence $|a_n|^2$, we have that $(b) \iff a_n^2 \to 0$ in density, but this last statement is certainly equivalent to (c).

□

Proposition 8.6. *These are equivalent, for a measure-preserving transformation T of a probability space $(X, \mathcal{A}, \mu)$:*
(i) T is weak mixing;
(ii) Given $A, B \in \mathcal{A}$, there is a subset $K \subseteq \mathbb{N}$ of density one such that T is mixing along K; that is, if $K = \{k_1, k_2, \dots\}$ then $\lim_{i\to\infty} \mu(A \cap T^{-k_i} B) = \mu A \mu B$.

Proof. Taking as above $a_n = \mu(A \cap T^{-n} B) - \mu A \mu B$, then this follows from $(a) \iff (c)$ in the lemma. □

Exercise 8.3. Use part (ii) above to give another proof that the odometer and the irrational rotation are not weak mixing.

Remark 8.2. Now mixing states that for all A, B,

$$\mu(A \cap T^{-n} B) \to \mu A \mu B.$$

An interpretation of this statement is that the events A and $T^{-n}(B)$ are *asymptotically independent.*

From Prop. 8.6, weak mixing then has the interpretation of the sets becoming asymptotically independent except for a rare set of times.

Here are two further interpretations of mixing, via relative measures. First, the above has this equivalent phrasing: that for each fixed B, then for all chosen A,

$$\mu_{T^{-n}B}(A) \to \mu A;$$

this expresses that $T^{-n}B$ is getting thoroughly dispersed throughout the space, as

$$\mu_{T^{-n}B} \to \mu \text{ as } n \to \infty$$

in this natural sense.

Secondly, for each fixed A, then for all B,

$$\mu_A(T^{-n} B) \to \mu B \text{ as } n \to \infty,$$

which says that each set B is getting evenly dispersed throughout A as $n \to \infty$.

Again, from Prop. 8.6, weak mixing has a corresponding version for both of these.

As pointed out by [EW10] p. 70, the key Lemma 8.5 (which we learned from Walters [Wal82]) is due to Koopman and von Neumann [KN32]. Part (ii) of Prop. 8.6 will bear further fruit when we return to further properties of weak mixing in §14 below.

Remark 8.3. It is important in the definitions that we are working with a *probability* space. (See Exercise 3.3). Indeed, suppose μ is a probability measure and $\nu = c\mu$ for $c > 0$, so $||\nu|| = \nu(X) = c$; then A, B are independent iff $\mu(A \cap B) = \mu(A)\mu(B)$ and so iff $\nu(A \cap B) = \nu(A)\nu(B)/||\nu||$, so this is the natural definition of independence if the measure is finite; the definitions of mixing and weak mixing in 41.1 should be changed accordingly.

We mention that as a consequence one has a possible definition of mixing in the infinite measure setting: the limit, taken as we induce on ever-larger sets of finite measure, yields

$$\nu(A \cap T^{-n}B) \to 0 \text{ as } n \to \infty.$$

.....

Reason for proving with squares wil be seen below in ...

....

Lemma: enuf to chack any on generating collection of sets

example: Bernoulli shift, golden toral aut. are mixing hence ergodic

9. Information and entropy.

-20 questions, information content of a subset -information of a partition: expected value of information function -information and independence -information of a transformation -ergodic theorem of information.

10. Basic constructions

We begin with definitions for the set category, whose objects are a set X together with a map $f : X \to X$, i.e. for sets and functions with no additional measure-theoretic or topological structure. For a category of dynamical systems with more structure, e.g. the topological category, with continuous maps on compact metric spaces, or the smooth or the measurable and measure- preserving category, the natural changes in the definitions are made. We learned much of this material in an excellent course in Ergodic Theory given by Doug Lind.

10.1. **Products.**

Definition 10.1. The **product** of two transformations (X, T) and (Y, S) is the map $T \times S$ on the product space $X \times Y$ where by definition $(T \times S)(x, y) = (Tx, Sy)$.

This provides the simplest example of a factor of a dynamical system, as both (X, T) and (Y, S) are homomorphic images of $(X \times Y, T \times S)$.

Exercise 10.1. *Show that the product of two circle rotations* $(\mathbb{T}, R_\theta)$ *and* $(\mathbb{T}, R_\phi)$ *is isomorphic (via a measure-preserving homeomorphism) to a rotation on the two-torus; the measures are the product of Lebesgue measure on the circles, and Lebesgue measure on the torus, respectively. Show that* $R_\theta \times R_\theta$ *is not ergodic.*

10.2. **Natural extension.** We have already seen an example of the projection of an invertible onto a non-invertible transformation: a two-sided shift factoring onto a one-sided shift (Exercise 4.5). A related example is the projection of the baker's transformation onto the doubling map of the interval, by considering only the x coordinate, see Figs. 7, 5, and the projection of a hyperbolic toral automorphism onto a Markov map of the interval in Figs. 16, 17.

But is this a general phenomenon, that is, does *every* non-invertible map arise as a factor of an invertible one? In other words, can one always find an invertible extension of any map? And, if this can be done, is there a natural choice for this covering transformation?

Now such an extension is certainly not unique: given one invertible extension, another, in some sense larger, one can be produced simply by taking the product with any invertible map. To be more precise, we consider dynamical systems in a given category: set, topological, and measure-theoretic, and make this

Definition 10.2. We define a partial order on the collection of all dynamical systems in one of the above categories, with $(X, T) \leq (Y, S)$ iff there is a factor map from (Y, S) to (X, T). (Recall that by definition, this is a *surjective* map which semiconjugates the dynamics).

The natural thing to do is, then, to look for a smallest invertible extension, with respect to this order. There is a general mathematical procedure to try in such a situation, known as an **inverse limit** construction. We describe how to carry this out in each of the above categories.

Let $T : X \to X$, perhaps not invertible. Writing $\Pi = \Pi_{i=-\infty}^{+\infty} X$, we give this space the dynamics of the left shift map σ. We then define $\widehat{X} \subset \Pi$ to be the set of all biinfinite sequences of points $\underline{x} \equiv (\dots x_{-1}.x_0 x_1 \dots)$ such that $x_{i+1} = T(x_i)$ for all $i \in \mathbb{Z}$. We write $\widehat{T}$ for the shift restricted to $\widehat{X}$, and define $\pi : \widehat{X} \to X$ by $\pi(\underline{x}) = x_0$.

In what follows we note that if T is not surjective, then we can replace (X, T) by the restriction to the eventual range (Proposition 4.2).

Proposition 10.1. $\widehat{T} : \widehat{X} \to \widehat{X}$ *is an invertible map. satisfying* $(\widehat{X}, \widehat{T}) \geq (X, T)$. *We have* $T \circ \pi = \pi \circ \widehat{T}$, *and* π *is surjective if and only if* T *is.*

If $S : Y \to Y$ *is an invertible map and* $\varphi : Y \to X$ *satisfies* $\varphi \circ S = S \circ \varphi$, *then there exists a unique* $\widehat{\varphi} : Y \to \widehat{X}$ *such that* $\pi \circ \widehat{\varphi} = \varphi$, *and* $\widehat{\varphi}$ *is surjective iff* φ *is.*

If T *is a continuous map of the topological space* $(X, \mathcal{T})$, *let* $\widehat{\mathcal{T}}$ *be the relative topology on* $\widehat{X}$ *induced from the product topology on* Π. *Then* $\widehat{T}$ *is a homeomorphism of* $\widehat{X}$ *and* $\widehat{\pi} : \widehat{X} \to X$ *is continuous. If* S *is a homeomorphism of the topological space* $(Y, \mathcal{S})$ *such that the map* $\varphi : Y \to X$ *as above is continuous, then* $\widehat{\varphi}$ *is continuous.*

If T *is a measure-preserving map of the measure space* $(X, \mathcal{A}, \mu)$, *then defining* $\widehat{\mathcal{A}}$ *to be the relative sigma-algebra on* $\widehat{X}$ *induced from the product sigma-algebra on* Π,

there exists a unique measure $\widehat{\mu}$ on $(\widehat{X}, \widehat{\mathcal{A}})$ such that $\pi_*(\widehat{\mu}) = \mu$; $\widehat{\mu}$ is $\widehat{T}$-invariant, and if S is an invertible measure-preserving of a measure space $(Y, \mathcal{B}, \nu)$ such that the map $\varphi : Y \to X$ takes ν to μ, then $\widehat{\varphi} : Y \to \widehat{X}$ as above is a measure-preserving homomorphism.

Proof. Beginning in the set category, supposing T is onto, then given $x_0 \in X$ there exists a sequence of preimages, $(\dots x_{-2}, x_{-1})$ such that $x_{i+1} = T(x_i)$ for all $i \leq -1$. Defining the biinfinite string $\underline{x} \equiv (\dots x_{-1}.x_0 x_1 \dots)$ so this holds for all $i \geq 0$ as well, we have found $\underline{x} \in \widehat{X}$ such that $\pi(\underline{x}) = x_0$, whence π is onto. Conversely, given $x_0 \in X$, then if π is onto we have some $\underline{x} \in \widehat{X}$ such that $\pi(\underline{x}) = x_0$. Here $\underline{x} \equiv (\dots x_{-1}.x_0 x_1 \dots)$, whence $T(x_{-1}) = x_0$ and T is onto.

Now given $S : Y \to Y$ invertible with a semiconjugacy $\varphi : Y \to X$, and given $y_0 \in Y$, let $\underline{y} = (\dots y_{-1}.y_0 y_1 \dots)$ be the (unique) sequence such that $S(y_i) = y_{i+1}$. Define $\widehat{\varphi} : Y \to \widehat{X}$ by $\widehat{\varphi}(y_0) = \underline{x} \equiv (\dots \varphi(y_{-1}).\varphi(y_0), \varphi(y_1) \dots)$; that $\underline{x} \in \widehat{X}$ follows from the fact that $\varphi \circ S = S \circ \varphi$. Then $\widehat{\varphi} \circ S = S \circ \widehat{\varphi}$; and $\widehat{\varphi}$ is the unique such map; moreover $\widehat{\varphi}$ is indeed surjective iff φ is.

Moving to the topological category, for $i \in \mathbb{Z}$, let $\pi_i : \Pi \to X$ denote the i^{th} coordinate projection; thus $\pi = \pi_0$. The product topology on Π is the smallest topology to make each coordinate projection π_i continuous, so a fortiori $\pi : \widehat{X} \to X$ is continuous. More concretely, for $\mathcal{U}$ open in X, then $\widehat{\mathcal{U}} \equiv \pi^{-1}(\mathcal{U}) = (\dots \times X \times X \times \mathcal{U} \times X \times \dots) \cap \widehat{X}$ which is open in the relative topology on $\widehat{X}$.

Now the product topology is generated by sets of the form $\pi_i^{-1}(\mathcal{U})$. And for any $k \in \mathbb{Z}$, $\widehat{T}^k(\pi_i^{-1}(\mathcal{U})) = \pi_{i+k}^{-1}(\mathcal{U})$ which is open; in particular this holds for $k = \pm 1$, proving that $\widehat{T}$ is a homeomorphism.

Lastly we consider the category of measure-preserving mappings.

For $n \leq k \leq m \in \mathbb{Z}$, consider $A_n, \dots, A_m \in \mathcal{A}$ and set $\widehat{A}_k = \pi_k^{-1}(A_k)$. The sigma-algebra $\widehat{\mathcal{A}}$ is generated by sets of the form $\widehat{A}_n \cap \dots \cap \widehat{A}_m = (\dots \times X \times X \times A_n \times \dots \times A_m \times X \times \dots) \cap \widehat{X}$. We define $\widehat{\mu}$ on such a set by $\widehat{\mu}(\widehat{A}_n \cap \dots \cap \widehat{A}_m) = \mu(\pi_n(\widehat{A}_n \cap \dots \cap \widehat{A}_m)) = \mu(\pi_0(T^{-n}(\widehat{A}_n \cap \dots \cap \widehat{A}_m))$. This is additive since for two disjoint such sets, $(\widehat{A}_n \cap \dots \cap \widehat{A}_m)$ and $(\widehat{B}_n \cap \dots \cap \widehat{B}_m)$, the image by T^{-n} and then by π is disjoint and μ is additive. Since μ is T-invariant, $\widehat{\mu}$ is $\widehat{T}$-invariant. We show that for the map $\varphi : Y \to X$, given that $\varphi_*(\nu) = \mu$, then $\widehat{\varphi}_*(\nu) = \widehat{\mu}$. Now for $A \in \mathcal{A}$, $\mu(A) = \widehat{\mu}(\pi^{-1}(A))$ but also $\mu(A) = \nu(\varphi^{-1}(A))$, and since $\widehat{\varphi}^{-1}(\pi^{-1}(A)) = \varphi^{-1}(A)$ we are done.

□

Now we see that the natural extension satisfies a universal property.

Corollary 10.2. *Given a dynamical system (X, T) with T surjective, in the set, topological or measure-preserving categories, then the the natural extension $(\widehat{X}, \widehat{T})$ is the unique (up to isomorphism) smallest extension of (X, T) which is an invertible transformation.*

Proof. If there is another minimum extension (Y, S) then there exists $\varphi : Y \to X$ and $\Phi : \widehat{X} \to Y$ both surjective such that $\pi = \varphi \circ \Phi$. We also know than since $\widehat{X}$ is

minumum, for this same φ there exists $\widehat{\varphi} : Y \to \widehat{X}$ with $\widehat{\varphi} \circ \pi = \varphi$. We claim that $\Phi \circ \widehat{\varphi}$ is the identity id_Y. Suppose that for some $y \in Y$, $\Phi \circ \widehat{\varphi}(y) = \widetilde{y}$. Let $\underline{x} = \widehat{\varphi}(y)$. Then we can identify $\underline{x} = (\dots x_{-1}.x_0x_1\dots)$, since $x_0 = \pi(\underline{x}) = \pi \circ \widehat{\varphi}(y) = \varphi(y)$, and similarly, $x_k = \pi(T^k(\underline{x})) = \varphi(S^k(y))$.

Now we are given that $\Phi(\underline{x}) = \widetilde{y}$. We claim that $y = \widetilde{y}$. Since Φ is surjective, there is some $\widetilde{\underline{x}}$ which maps to y. But we know the sequence $\widetilde{\underline{x}}$, since $\varphi \circ \Phi(\widetilde{\underline{x}}) = \varphi(y) = \pi(\widetilde{\underline{x}}) = \widetilde{x}_0$ but this is also $\varphi(y) = x_0$; similarly $\widetilde{x}_k = \underline{x}_k$ for all $k \in \mathbb{Z}$. Thus $y = \Phi(\widetilde{\underline{x}}) = \Phi(\underline{x}) = \widetilde{y}$.

Thus $\widehat{\varphi}$ is an isomorphism from (Y, S) to $(\widehat{X}, \widehat{T})$. It follows that in the topological and measure categories this is a topological, respectively measure, isomorphism. □

Remark 10.1. REMARK: (no dynamical Schröder-Bernstein theorem).??? examples: solenoid/shift space

10.3. **Towers.** Beginning in the set category, given X with a map $T : X \to X$ and a function $r : X \to \mathbb{N}^* = \{1, 2, \dots\}$ (for the time being, no measure or $\sigma-$algebra is involved) we define a countable partition $\mathcal{P} = \{X_1, X_2, \dots,\}$ of X by the values taken, with $X_n = \{x : r(x) = n\}$.

We define subsets of the product space $X \times \mathbb{Z}$ as follows. For $n \in \mathbb{Z}$ we write $X_{n,k} = X_n \times \{k\}$. The n^{th} **column** over X_n is $C_n = \cup_{k=0}^{n-1} X_{n,k}$. The k^{th} **level** of this column is $X_{n,k}$, so C_n has n levels. We define the **tower space** to be the union of the columns, $\widehat{X} = \cup_{n\in\mathbb{N}^*} C_n$; thus $\widehat{X} = \{(x, k) : 0 \le k < r(x)\}$. See Fig. 18.

The k^{th} level of the tower is the union of the column levels, so $L_k = \{(x, k) \in \widehat{X}\}$. The zeroth level L_0 is called the **base** of the tower, and is naturally identified with X via $(x, 0) \mapsto x$. The **top** $\equiv \cup_{n=1}^{\infty} X_{n,n-1}$ is the union of all the highest levels of the columns; a point $(x, 0)$ in the base corresponds to the point above it in the top, $(x, r(x) - 1)$.

We define a map $\widehat{T}$ on $\widehat{X}$ as follows: in each column (excluding its top level) we ascend like an elevator, sending $X_{k,n}$ to $X_{k+1,n}$ by $(x, k) \mapsto (x, k + 1)$, and on the top define:

$$\widehat{T}(x, m) = (Tx, 0).$$

The set $\widehat{X}$ is known as the **tower** or **Kakutani skyscraper** over X with **return-time function** or **height function** r, since the time it takes for a point $(x, 0)$ in L_0 to return to the base is $r(x)$, which equals the height of the tower. We write $(\widehat{X}, \widehat{T}, r)$ for the tower space and map.

So far we have been working with sets; now we bring in measures. Given a measure-preserving transformation $(X, \mathcal{A}, T, \mu)$, with μ finite or infinite $\sigma-$finite, together with a measurable function $r : X \to \mathbb{N}^*$, we extend μ to the tower as follows: letting m denote counting measure on $\mathbb{N}$, we define $\widehat{\mu}$ on $\widehat{X}$ to be the restriction of product measure $\mu \times m$ to $\widehat{X} \subseteq X \times \mathbb{N}$; that is, we copy the measure on the base vertically on each column.

Lemma 10.3. *If $(X, \mathcal{A})$ is a measurable space and $T : X \to X$ and $r : \widehat{X} \to \mathbb{N}^*$ are measurable functions, then the tower map $\widehat{T} : \widehat{X} \to \widehat{X}$ is measurable and $\widehat{\mu}$ is invariant for $\widehat{T}$.*

Proof. For E a measurable suset of a level ≥ 1, $T^{-1}(E)$ is one level down hence is measurable and has the same measure. It remains to check this for $E \in \mathcal{A}$ with $E \subseteq L_0$. Identifying X with L_0, we have $T : L_0 \to L_0$; we compare $T^{-1}(E)$ to $\widehat{T}^{-1}(E)$. For each n, $E_n \equiv X_n \cap T^{-1}(E)$ is a measurable set; we ride this up to the top of the column C_n over X_n to $E'_n = \widehat{T}^{-1}(E) \cap C_n$, which is measurable and of the same measure. But $\widehat{T}^{-1}(E)$ equals $\cup_{n\geq 1} E'_n$, so $\widehat{\mu}(\widehat{T}^{-1}(E)) = \widehat{\mu}(\cup_{n\geq 1} E'_n) = \mu(\cup_{n\geq 1} E_n) = \mu(T^{-1}(E)) = \mu(E)$. □

We remark that in the above, we have not assumed T is an invertible map.

10.4. **Return times and induced maps.** Beginning with $T : X \to X$ measurable, and given $A \subset X$, we set $B = A^c$ and define the **return-time function** $r_A : A \to \widehat{\mathbb{N}}^*$ by

$$r_A(x) = \inf\{n > 0 : T^n(x) \in A\}$$

with $r_A(x) = \infty$ if x never returns to A.

Setting

$$A_k = \{x \in A : r_A(x) = k\}, \tag{6}$$

we call the collection $\{A_k : k \in \widehat{\mathbb{N}}^*\} = \mathbb{N}^* \cup \{\infty\}$ the **return-time partition** of A. Note that $A_1 = A \cap T^{-1}(A)$, and in general: $A_k = A \cap T^{-1}(B) \cap \cdots \cap T^{-(k-1)}(B) \cap T^{-k}(A)$. As a consequence, these are measurable sets, and so r_A is a measurable function.

The **induced map** or **first-return map** of T on $A \setminus A_\infty$ is:

$$T_A(x) = T^{r_A(x)}(x),$$

that is, $T_A = T^n$ on the set A_n.

We write $X_A = \cup_{n\in\mathbb{N}} T^n(A)$. Since these are forward (rather than inverse) images, whether or not this set is measurable is a subtle point; that is true for instance if T is locally invertible (see Exercise 5.1).

The dynamics of T on this subspace of X is indicated in Fig. 19; points move upwards in a bijective way.

Now we restrict attention to invertible maps. We use the return-time data to build the tower of height r_A over the induced map (A, T_A), denoting this by $(\widehat{X}_A, \widehat{T})$. We define a map $\alpha : \widehat{X}_A \to X_A$ by $\alpha(x, k) = T^k(x)$; see Fig. 18.

(We return to consider further the non-invertible case below in §50.1.)

Theorem 10.4. *Given an invertible measure-preserving transformation T of a measure space $(X, \mathcal{A}, \mu)$, and a recurrent subset $A \in \mathcal{A}$ of positive measure, then the induced (first-return) map T_A is measurable and measure-preserving. The tower with height r_A built over $(A, \mathcal{A}|_A, \mu|_A, T_A)$ is isomorphic to the restriction of the original map $(X, \mathcal{A}, \mu, T)$ to the set X_A swept out by A, via the map α. In particular, if T is conservative ergodic then the tower map is isomorphic to the original map.*

Proof. We have already noted that each A_k is a measurable set, with r_A a measurable function.

Since T is invertible as a measurable transformation, as noted in Exercise 5.1, T^{-1} is also a measure-preserving map, whence X_A is a measurable set.

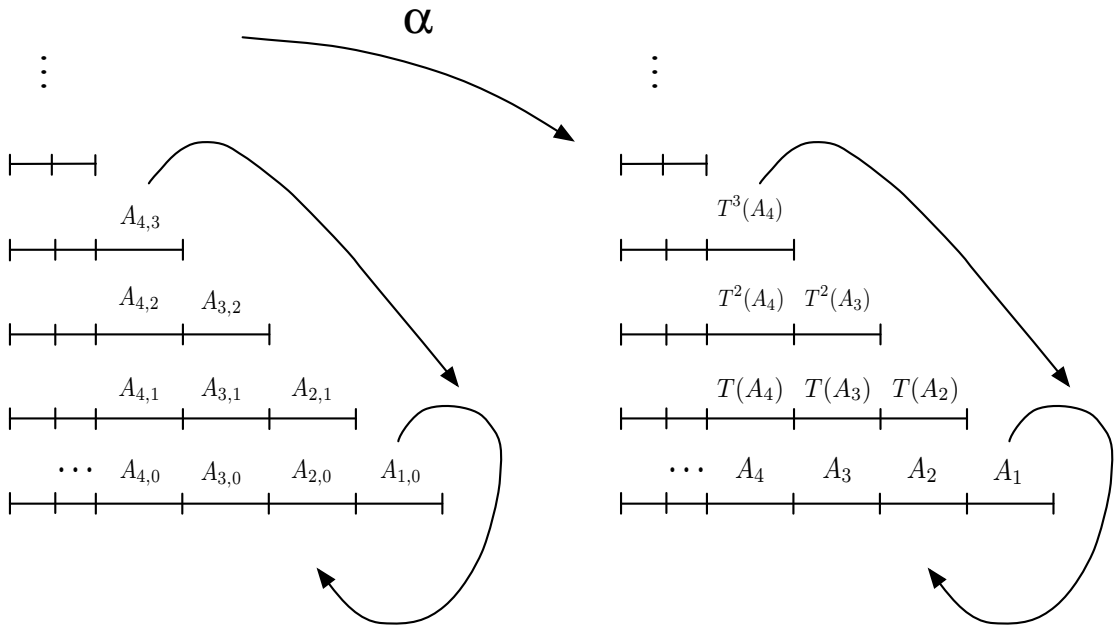


FIGURE 18. The conjugacy from the external to the internal tower.

That A is recurrent means $\mu(A_\infty) = 0$. Given $E \subseteq A$ measurable, $T^{-1}(E)$ is a measurable subset of the top of the tower. As in the proof of Lemma 10.3, we compare $\widetilde{E} \equiv T^{-1}(E)$ to $T_A^{-1}(E)$, except now we use the converse argument. Thus to show the measurability of $T_A^{-1}(E)$, we move each piece on the top of a given column C_n (that is, $C_n \cap \widetilde{E}$) down to the base, getting $E_n \equiv T^{-(n-1)}(C_n \cap \widetilde{E}_n)$. The union of these sets is the inverse image: $T_A^{-1}(E) = \cup E_n$, which is measurable and has the same measure as $T^{-1}(E)$ and hence as E. □

In summary, there is a duality between the operations of inducing and tower-building: the tower built over an induced map is isomorphic to the original map, and the induced map on the base of a tower map is the original map on that base. The figure on the right, since it consists of points from the original space X, can be called an **internal tower**, in contrast to an **external tower** on the left which has been built by adding points to the space.

10.5. **Four applications of the tower construction.** We illustrate the importance and power of Kakutani's tower idea, coupled with that of the the natural extension, with four results.

Poincaré Recurrence via towers.

First we have yet another proof of the Poincaré Recurrence Theorem, which is perhaps the most transparent of all, because of the geometric picture that comes from the idea of representing a transformation as the tower over an induced map.

Theorem 10.5. *Given a measure-preserving transformation $(X, \mathcal{A}, \mu, T)$ and $A \in \mathcal{A}$ with $\mu(A) > 0$, writing $\widetilde{A}$ for the set of points in A which return at least once, then if*

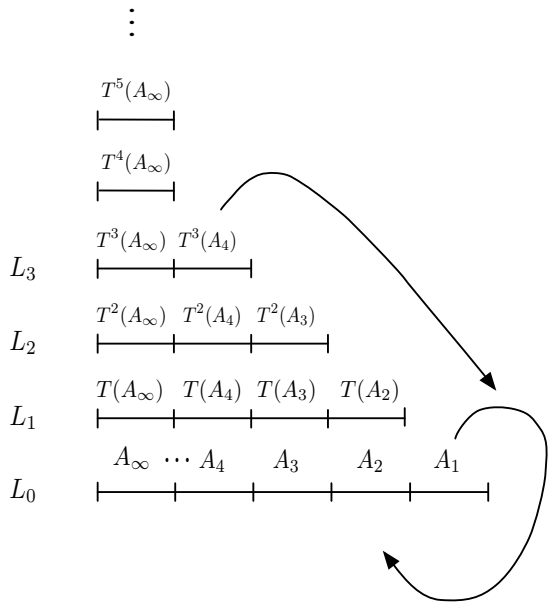


FIGURE 19. The tower proof of Poincaré recurrence.

$\widetilde{A}$ has positive measure, a.e. point of $\widetilde{A}$ returns infinitely many times. If $\mu(X) < \infty$, then this is a.e point of A.

Proof. (invertible case) We build the tower $X_A = \cup_{n\in\mathbb{N}} T^n(A)$ of Fig. 19. We are not assuming r_A is everywhere finite; the points which never return define the set A_∞, with an infinite height column $C_\infty = \cup_{k\in\mathbb{N}} A_{\infty,k}$ above it. We know the points in the complement $A^1 \equiv A \setminus A_\infty = \cup_{i=1}^\infty A_{i,0}$ all return at least once. Since T is invertible, the levels $A_{\infty,k}$ of the column C_∞ are all measurable, with the same measure. If X has finite measure then A_∞ must have measure zero whence $\mu(A \setminus A^1) = 0$.

In Theorem 50.1 we have shown the map $T_A : A \to A$ is measure-preserving. Thus $A^2 \equiv T_A^{-1}(A^1)$ has full measure in A. And $x \in T_A^{-1}(A^1)$ iff $T_A(x) \in A^1$ iff $(T_A)^2(x) \in A$. Continuing in this way, the sets $A^1 \supseteq A^2 \supseteq A^3 \supseteq \dots$ nest down to a set of full measure in A, so indeed a.e. point of A returns to A infinitely many times. □

Proof. (noninvertible case) Supposing now that $(X, \mathcal{A}, \mu, T)$ is noninvertible, we construct its natural extension, denoted now $(Y, \mathcal{B}, \nu, S)$; we write $\pi : Y \to X$ for the natural homomorphism. Let $A \subseteq X$ have positive measure, and consider $C = \pi^{-1}(A)$. Then $\nu(C) = \mu(A)$ since π is measure-preserving. From the proof for the invertible case, there exists a set $G \subseteq C$ of full measure such that every $x \in G$ returns to C infinitely many times. Consider the points $B \subseteq A$ which do not return to A infinitely many times. Then any $w \in \pi^{-1}(B)$ is not in G. Hence $\mu(B) = 0$. □

Note by the above that if $(X, \mathcal{A}, \mu, T)$ is conservative, then given A of positive measure, by the above $\mu(A_\infty) = 0$, and hence a.e. point returns infinitely often. This gives a second proof that conservative implies recurrent, see Proposition 5.5.

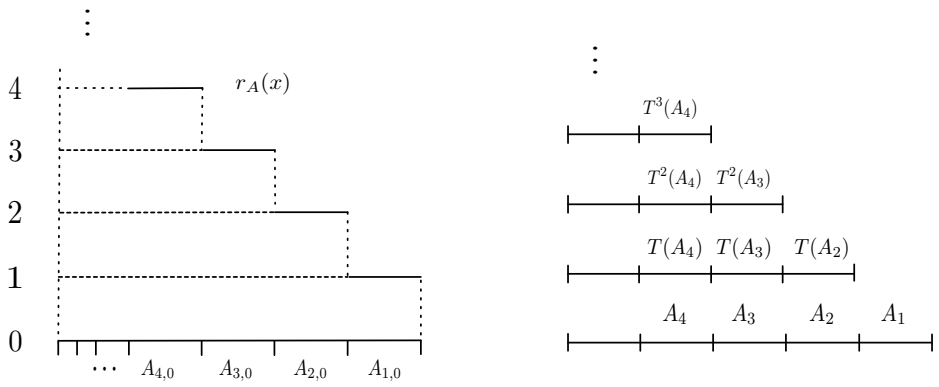


FIGURE 20. The integral equals the total mass.

Kac' theorem via towers.

The Poincaré Recurrence Theorem states that we return to A; a next question will be: how long will it take to return? A precise answer is given by the following famous theorem of Kac:

Theorem 10.6. *Let $(X, \mathcal{A}, \mu, T)$ be a measure-preserving transformation of a probability space. We assume that the future iterates of A sweep out all of X (e.g. if the map is ergodic). Then the expected return time to A is $1/\mu(A)$.*

Proof. (invertible case) Since A sweeps out X, it has positive measure. We draw the tower picture of Fig. 18. We draw also the graph of the return-time function r_A. Now the expected return time is just the expected value (see §3 and Fig. 3) of r_A, and is the integral over the normalized (i.e. the relative) measure μ_A. The integral with respect to the non-normalized restricted measure $\mu|_A$ can be found by summing the mass of each horizontal rectangle on the left, which equals the sum of the mass of each level on the right. And this total mass is exactly $\mu(X) = 1$.

This gives

$$\mathbb{E}(r_A) = \int_A r_A \mathrm{d}\mu_A = \frac{1}{\mu(A)} \int_A r_A \mathrm{d}\mu|_A = \frac{1}{\mu(A)} \cdot 1.$$

Proof (noninvertible case) Given $(X, \mathcal{A}, \mu, T)$ noninvertible, we construct its natural extension $(Y, \mathcal{B}, \nu, S)$, with $\pi : Y \to X$ the natural homomorphism. Now consider $A \subseteq X$ and its lift $B = \pi^{-1}(A)$; these have the same measure. Note that the return-time function lifts:

$$r_B = r_A \circ \pi.$$

Thus

$$\mathbb{E}(r_A) = \mathbb{E}(r_B) = \frac{1}{\mu(B)} = \frac{1}{\mu(A)}.$$

□

Rochlin's Lemma via towers.

We next introduce the following basic tool of ergodic theory:

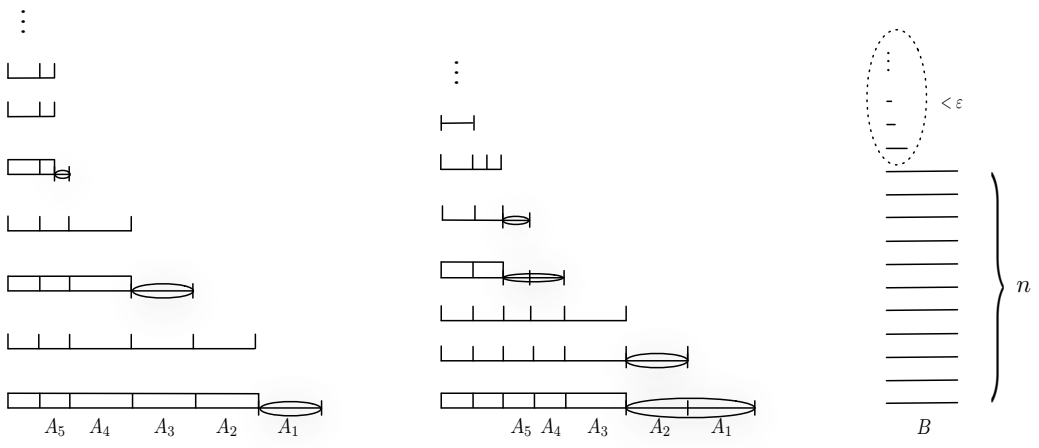


FIGURE 21. Proof of Rochlin Lemma for $n = 2, 3$; Rochlin tower of height n.

Theorem 10.7. *(Rochlin) Given an ergodic invertible measure-preserving transformation $(X, \mathcal{A}, \mu, T)$, an integer $n \geq 1$ and $\varepsilon > 0$, then there exists a set $A \in \mathcal{A}$ such that the return time to A is greater than n and the column of the tower of height n over A has mass greater than $1 - \varepsilon$.*

Proof. For $n = 2$, we set $B_0 = \cup_{k \geq 2} A_k$, $B_2 = T^2(\cup_{k \geq 4} A_k)$, $\ldots$, $B_{2j} = T^{2j}(\cup_{k \geq 2(j+1)} A_k)$ so $B_0 \subseteq L_0$, $B_2 \subseteq L_2, \ldots$, $B_{2n} \subseteq L_{2n}$. We define $B = \cup_{k \geq 0} B_{2k}$, see the left-hand side of Fig. 21, and build the tower of height 2 over B, with two equal measure levels $B, T(B)$. Now we estimate the measure of what is left over, the set $X \setminus B \cup T(B)$; these are the circled sets in the Figure. Pushing these sets down to the base, we get $A_1 \cup A_3 \cup A_5$ which has the same measure and is $\leq \mu(A)$.

Thus if we choose A to have measure $< \varepsilon$, we have found a set B such that $\mu(X \setminus (B \cup T(B)) < \mu(A) < \varepsilon$.

For $n = 3$, we proceed in a similar way, shown in the right-hand side of Fig. 21. Now our "error estimate" is $\mu(X \setminus (B \cup T(B) \cup T^2(B)) < 2\mu(A)$, so if we choose A to have $2\mu(A) < \varepsilon$ we are done.

For height n, beginning with $\mu(A) < \varepsilon/n$, we have a Rochlin tower of height n with error less than ε, as desired.

□

Remark 10.2. From Rochlin's lemma one draws the quite startling conclusion that all transformations are nearly alike, the only difference between them being in that tiny part of arbitrarily small measure at the top of the very high tower. The reason this is surprising is that there are maps that we know have very distinct behavior, for instance with entropy zero, finite or even infinite; maps that have strong independence properties, and maps that have many invariant measures, while others are uniquely ergodic. So what is going on? Part of the answer is that Rochin's Lemma is a purely measure-theoretic statement, while many of these properties involve a rich

mix of categories, topological and geometric. Now measure entropy is purely measure-theoretic; however what happens in the above construction is simply this: one could redefine the metric inside the column so that the motion up the column is an isometry, and so contributing no entropy; but then all we will have done is to concentrate all the entropy inthat last tiny but crucial little bit.

In other words this apparent closeness gives less than at first meets the eye. Nevertheless, in special situations, and in the right hands, the lemma is a powerful tool indeed: a remarkable example is given by the tower constructions (employing a stronger form of Rochlin's Lemma) used in Ornstein's isomorphism theory [Orn73], [Shi73],

Nonmeasurable sets via towers; 2 partitions; forward image nonmeas

In every analysis course one sees the construction of a nonmeasurable set, usually using a coset space and the Axiom of Choice (see e.g. §3.4 of [Roy68], p.69 of [Hal50]). Looking at the same example through the eyes of dynamics (and in particular the tower construction) makes it especially clear what is the essential point here.

Proposition 10.8. *Let $(X, \mathcal{A}, \mu, T)$ be a conservative ergodic invertible measure-preserving transformation, and assume that a.e. orbit is countably infinite. Then a set A formed by choosing a single point from each orbit gives a fundamental domain for the action of $\mathbb{Z}$ generated by the transformation; this set is nonmeasurable, i.e. cannot be in $\mathcal{A}$. The same holds for an action of a countably infinite group or semigroup.*

Proof. Let $N \subseteq X$ be the collection of periodic points, i.e. those with finite orbits; we are given that $\mu(N) = 0$. We consider (see Example 4.2) the orbit equivalence relation on X. This partitions the space; we choose (via the Axiom of Choice!) one point from each partition element; this is the set A. By definition this is a fundamental domain for the associated $\mathbb{Z}$-action on $X \setminus N$. As for any fundamental domain for a group action, all iterates $T^n(A)$ are disjoint, and their union is the whole space. Directly, if $x \in A \cap T^n(A)$, then $T^n(x) \in A$ but $T^n(x) \in \mathcal{O}(x)$ and A contains exactly one point from each orbit. Furthermore, $\cup_{n\in\mathbb{Z}} T^n(A) = X$ for the same reason.

The proof of nonmeasurability will be by contradiction. Suppose, then, that A is measurable. We arrive at the conclusion in three slightly different ways:
(1) Since A is measurable, then if A has positive measure, by recurrence, a.e. point in A returns, but every point in $A \setminus N$ never returns. Thus A has measure zero, but this contradicts that $\cup_{n\in\mathbb{Z}} T^n(A) = X$.
(2) Since A is measurable and its iterates sweep out the space, by Theorem 10.4 the tower of height r_A over the induced map (A, T_A) is isomorphic to the original map. However $r_A(x) = \infty$ for every $x \in A \setminus N$, so this is a wandering set, hence by recurrence must have measure zero, giving a contradiction as before.
(3)The iterates of $A \setminus N$ are all disjoint so the future iterates form a tower over this set, which consists of a single column, as on the left-hand side of the tower in Fig. 19. That is, $A \setminus N = A_\infty$, and if A is measurable it must have measure zero, since otherwise this contradicts conservativity.

Note that for the finite measure case, since the set A is wandering, its measure cannot meaningfully assume any nonnegative real value- even assuming only finite additivity of μ! □

For a concrete example, let $(\mathbb{T}, R_\theta)$ be the circle rotation $x \mapsto x + \theta (\mathrm{mod}\, 1)$. The collection of orbits $\{\mathcal{O}(x) : x \in \mathbb{T}\}$ partition $\mathbb{T}$. From the algebraic point of view, since $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ is a group, letting H denote the (dense) subgroup of $\mathbb{T}$ generated by θ, then the orbits are exactly the cosets of x. As above, by the Axiom of Choice we form a set A which contains exactly one point from each of these equivalence classes. Then each orbit is countably infinite iff θ is irrational, and in this case $r_A(x) = +\infty$ for all $x \in A$, and $(\mathbb{T}, R_\theta)$ is isomorphic to the tower map with a single biinfinite column. If on the other hand θ is a rational number, then the tower is a single column of constant height.

Remark 10.3. In conclusion, from Example 4.2 together with the above, forming the quotient space (factoring out by the orbit equivalence relation) is problematic for any except the simplest dynamics: that which we encounter in geometry when building homogeneous spaces. Here are some related questions:

Exercise 10.1. Does there exist a measure space with a countable collection of nonmeasurable sets whose union is measurable? Can there exist a finite collection of nonmeasurable sets whose union is measurable?

Show that a similar idea works for flows: given an irrational rotation flow on the torus for $n \geq 2$, $\tau_t(\mathbf{u}) = \mathbf{u} + t\mathbf{v} (\mathrm{mod}\, \mathbb{Z}^d)$, let E be a collection of representatives for the orbit equivalence relation. Show that for ε sufficiently small, the set $\{\tau_t(E) : t \in [0, \varepsilon]\}$ is not measurable.

Find two measurable spaces $(X, \mathcal{A})$ and $(Y, \mathcal{B})$, a measurable function $f : X \to Y$ and a set $a \in \mathcal{A}$ such that $f(A)$ is not measurable.

A silly answer to the last exercise is this: take $X = Y = \{0, 1\}$ with f the identity map, $\mathcal{A}$ the power set sigma-algebra (all subsets) and $\mathcal{B}$ the trivial sigma-algebra ($\emptyset$ and X). Here is a hint for a more interesting example: consider $X = Y = I$, with f the Cantor function, and with $\mathcal{A} = \mathcal{B}$ the Lebesgue sigma-algebra (the completion of the Borel sigma-algebra with respect to Lebesgue measure).

10.6. **Flow built under a function and Poincaré cross-section.** The tower idea has a continuous-time version, also due to Kakutani.

We begin with an invertible measure-preserving map $(X, \mathcal{A}, \mu, T)$ and a measurable function $r : X \to [0, \infty)$. For the product space $X \times \mathbb{R}$, with the product measure where $\mathbb{R}$ has Lebesgue measure, we then consider the subset $\{(x, t) : 0 \leq t \leq r(x)\}$ and lastly define X_r to be this subset modulo the identification $(x, r(x)) \sim (Tx, 0)$.

We then define a flow τ_t on X_r by $(x, s) \mapsto (x, s + t)$.

Conversely, given a flow (X, τ_t), a **cross-section** is a subset $A \subseteq X$ such that each orbit meets A in a discrete set of times. This is also referred to as a **transversal** to the flow.

We require A to be **measurable cross-section** in the sense that a small rectangle $\{\tau_t(A) : t \in [0, \varepsilon]\}$ is a measurable subset of X.

We then define a map T_A on A, the **first return** or **Poincaré** map; this preseves the measure μ^A defined by $\mu^A(E) = \frac{1}{\varepsilon}\{\tau_t(E) : t \in [0, \varepsilon]\}$ for $\varepsilon < r_A$ on E.

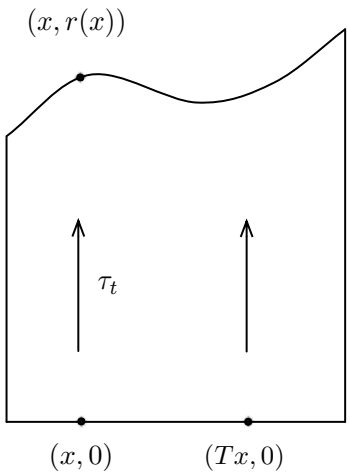


FIGURE 22. The special flow

Theorem 10.9. *(Ambrose-Kakutani) Given a conservative ergodic measure-preserving flow (X, τ_t), there exists a measurable cross-section A. The first return map T_A is ergodic, and the special flow built over (A, T_A, μ^A)*

Given a flow (X, τ_t), one way to discretize it is to consider the time-t_0 map for some fixed time (such as the time-1 map τ_1). A very different way is to consider the return map to a cross-section of the flow; the Ambrose-Kakutani theorem guarantees that this can be done.

We remark that infinite measures can occur for the flow space, the cross-section, or both: we can have a finite measure for the cross-section with a nonintegrable return-time function r; or we can have a finite measure for the flow but a cross-section with infinite measure, and a function r that is so small that it integrates to give that finite number.

Definition 10.3. Given an invertible map $T : X \to X$, in the special case where $r(x) \equiv 1$, the special flow is known as the **suspension flow** of the map, and the flow space X_r is a **suspension space**.

When $T : M \to M$ is a diffeomorphism of a manifold M then the suspension construction can give an interesting way of creating a new manifold M_1. Any manifold which can arise in this way is know as a *manifold which fibers over the circle*. The reason for this name is that the identification of the top and bottom of the flow space projects each vertical segment over $(x, 0)$ in the base horizontally to the circle S^1. Therefore the flow space is a fiber bundle with base this circle and with fibers equal to X. (This is a switch of perspective, since from the dynamical point of view the base of the special flow is the space M !) This manifold is the product space $S^1 \times M$ modulo the identification of the fibers over 1 and 0 via the map T, known as the **holonomy map** for this fiber bundle. In particular, if T is not homotopic to the identity, then the fiber bundle is **nontrivial** in the sense that it is not homeomorphic to the product space $M \times S^1$. We will see examples of this below.

Exercise 10.2. *Show that for the rotation flow on the torus, flow $\tau_{\mathbf{v},t}$, minimality is equivalent to unique ergodicity.*

A quite different, more abstract construction of the special flow is given in §29.1. In fact, this is more general, as it allows for *nonpositive* "return times".

10.7. **A further application of towers: Kakutani equivalence.** Kakutani's idea of induced map with the converse operation of tower-building is remarkably powerful and useful. An excellent illustration is given by the methods of proof presented here, which will be entirely geometric, regarding the basic properties of an equivalence relation on transformations- much weaker than measure-theoretic isomorphism- introduced by Kakutani.

Definition 10.4. We begin by defining an order on conservative ergodic measure-preserving transformations, writing $(Y, \mathcal{B}, \nu, S) \leq (X.\mathcal{A}, \mu, T) \leq$ iff there exists $A \in \mathcal{A}$ with $\mu(A) > 0$ such that the induced map $(A, \mathcal{A}|_A, \mu|_A, T_A)$ is isomorphic to $(Y, \mathcal{B}, \nu, S)$. For short, we write this as $S \leq T$. We then define two relations: $T_1 \sim_\ominus T_2$ iff there exists T such that $T_1 \leq T$ and $T_2 \leq T$; $T_1 \sim_\oplus T_2$ iff there exists S such that $S \leq T_1$ and $S \leq T_2$.

Lemma 10.10. *Given an invertible conservative ergodic measure-preserving transformation $(X, \mathcal{A}, \mu, T)$, then for $A \in \mathcal{A}$ with $\mu(A) > 0$ with $B \equiv T^n(A)$, we have that $(A, \mathcal{A}|_A, \mu|_A, T_A)$ is isomorphic to $(B, \mathcal{A}|_B, \mu|_B, T_B)$.*

Proof. Defining $\Phi : A \to B$ by $\Phi = T^n$, we claim that $r_B \circ \Phi = r_A$. This follows immediately from the definitions. Then T_A □

From the Ambrose-Kakutani theorem an ergodic measure-preserving flow can be modelled by a special flow over a discrete-time transformation, the induced map to some "transversal". But for this to be useful, one would like to know how, given a fixed flow, such return maps might be characterized. This will be a central motivating idea in what follows. We begin with the same question for discrete time:

Definition 10.5. Two transformations $(X, \mathcal{A}, \mu, T)$ and $(Y, \mathcal{B}, \nu, S)$ are said to be **Kakutani equivalent** iff there exists a third transformation $(Z, \mathcal{C}, R, \rho)$ of which each is an induced transformation.

Theorem 10.11. *This defines an equivalence relation.*

Theorem 10.12. *Two transformations are Kakutani equivalent iff there exists a measure-preserving flow with measurable cross-sections for which both are first return maps.*

A key step is:

Lemma 10.13. *Two transformations are Kakutani equivalent iff each has a subset of positive measure such that the induced maps are isomorphic.*

Remark 10.4. A finite measure transformation can be Kakutani equivalent to an infinite one; see Example ??? (invertible renewal shift).

11. Further examples

11.1. **Stationary stochastic processes.**

11.2. **Almost-periodic functions.**

11.3. **Symbolic dyns for rotations, irrational flow on torus.**

11.4. **Continued fractions and the Gauss map; infinite measure version.** We recall some basics about continued fractions. (We return to this topic in §19.1.) We shall use the notation

$$x = [n_0 n_1 \dots] = \cfrac{1}{n_0 + \cfrac{1}{n_1 + \cdots}}$$

As is easy to show, rationals have a finite expansion, while every x irrational in $(0,1)$ has a unique such expansion. It is known that a quadratic irrational (a root of a quadratic equation) has an eventually periodic expansion.

We mention that when we study general Anosov automorphisms of $\mathbb{T}^2$, continued fractions of (all!) other quadratic irrationals will come into play. See §**??**.

11.5. **Double suspension, commutation relation.**

11.6. **The scenery flow of the Cantor set.** The informal idea of zooming down towards a point in a fractal set (one can google for examples!) can be turned into real mathematics with the help of a flow. To make things precise, let Ω denote the collection of all closed subsets of $\mathbb{R}^d$, and define the **magnification flow** on Ω by

$$g_t : A \mapsto e^t A$$

Thus, we dilate A about the central point 0 by the exponential factor e^t. We give Ω the **geometric topology**, which can be described as follows: let $\widehat{\mathbb{R}}^d = \mathbb{R}^d \cup \{\infty\}$ be the one-point compactification of $\mathbb{R}^d$, let $\widehat{\Omega}$ be the collection of closed subsets of $\widehat{\mathbb{R}}^d$. We take the Hausdorff metric on $\widehat{\Omega}$, and define the geometric topology on Ω to be the the relative topology on $\Omega \subseteq \widehat{\Omega}$.

Exercise 11.1. Show that a sequence of sets A_n converges to A in the geometric topology iff given $\varepsilon, K > 0$ there exists N such that for all $n > N$, $d_K(A_n, A) < \varepsilon$ where d_K is the Hausdorff metric for closed subsets of the ball of radius K about 0.

Given $F \in \Omega$, we define $S_x(F)$, the **scenery at** $x \in F$ to be the $\omega-$limit set of the translated set $(F - x)$ with respect to the magnification flow, and we define the **scenery of** F to be $S_F = \cup_{x \in F} S_x(F)$.

The **scenery flow** of F is (S_F, g_t). Thus, it is the magnification flow acting on the asymptotically small scenes of F.

Now for a smooth object such as a differentiable manifold M embedded in $\mathbb{R}^d$, the scenery flow is rather boring: the scenery at $x \in M$ is $S_x(M) = T_x(M)$, the tangent space at x, which gives a fixed point for the flow; the scenery space S_F is the tangent bundle.

But for a complicated object like a fractal set, the scenery keeps changing, and the flow can be interesting. And since we have the tangent space analogy, we can think

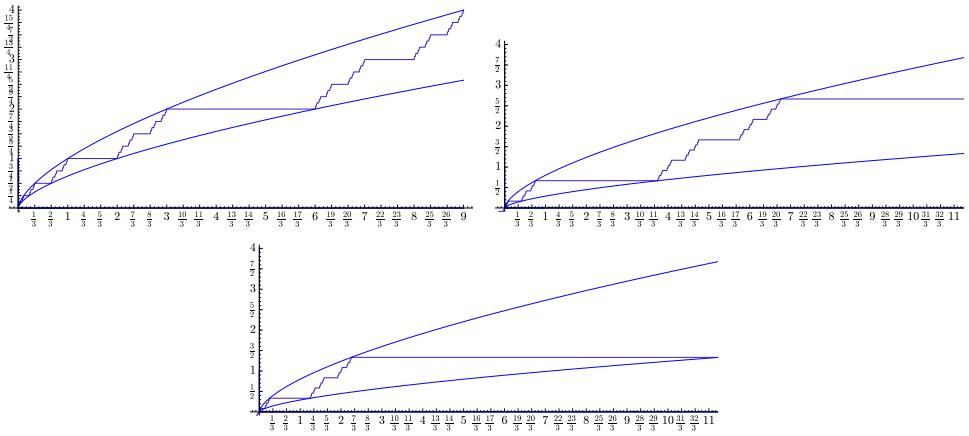


FIGURE 23. Zooming toward the point $x = 0$ of the Cantor function by the scaling scenery flow of exponent $d = \log 2/\log 3$; upper envelope t^d, lower envelope $(t/2)^d$. The first picture shows the Cantor function for the extended Cantor set; this is a scaling flow orbit of period $\log 3$, at time 0. Then zooming toward the right-hand side of the periodic point $x = 1/4$; these represent snapshots from a periodic orbit of period $2 \log 3$ at times 0 and $\log 3$. The upper and lower envelopes have changed!

of the space of sceneries as a sort of tangent space for the fractal set, which comes equipped with a natural flow action.

Now if one looks at small scales of the Cantor set, on first thought zooming down towards it is rather boring, since it is "the same everywhere", at every location and all scales. But there is a hidden assumption: the eye of our imangination naturally shifts its origin to the beginning of each subinterval. If instead we choose a point, and perform the zoom with a microscope fixed at that point, what we see will keep changing. It may in fact change in a periodic way, corresponding to a periodic orbit for the scenery flow, but the general behavior is much wilder: a nonperiodic orbit, representing positive entropy.

Indeed one can prove for the Cantor set C:

Scenery flow entropy equals the Hausdorff dimension of the limit set.

This same formula holds for some other especially nice fractal sets: hyperbolic $C^{1+\alpha}$ Cantor sets, hyperbolic Julia sets, and the limit sets of geometrically finite Kleinian groups. See [BF92], [BF96], [BF97], [BFU02], [Fis04]. It is not always true: for fractal sets associated with probability theory, the tendency is for scenery flow entropy to be infinite (this holds for the zero sets of Brownian motion and more general stable processes, see...)

To understand Fig. 24, the point $x = 1/4$ has ternary expansion .020202... The limiting scenery has large-scale structure given by ...20202.; thinking of this like the tree that determines the adic transformation for the Integer Cantor Set, we see, after the centrally located copy of C (centered at $1/4$) to its right a gap of 3, followed by

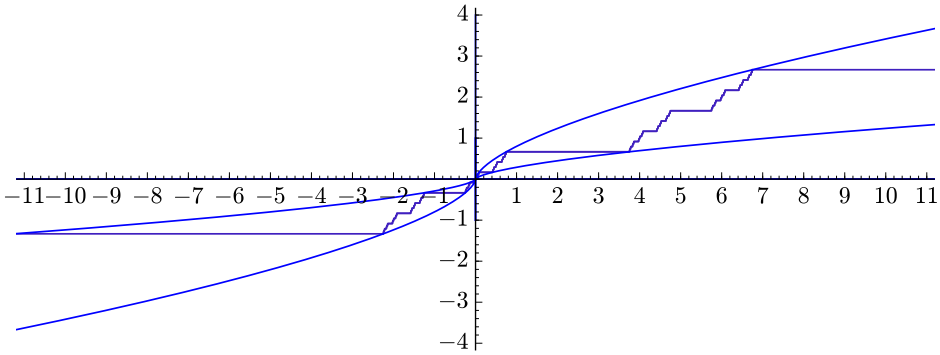


FIGURE 24. Zooming toward the point $x = 1/4$ via the scaling flow: this shows the two-sided (integrated) scenery. The right-handed scenery of $x = 3/4$ has been reflected in the origin to give the left-handed scenery for $x = 1/4$!

another copy of C, then a gap of length $3^3 = 27$; on the left we see a gap of length $3^2 = 9$, followed by a copy of C, followed by a gap of length 3^4, and so on.

Here we sketch the proof for the middle-third Cantor set.

DO ??

Exercise 11.2. Show the upper and lower envelopes for a periodic point are $(at)^d$ for some $a > 0$. Show that any such curve is invariant for the scaling flow. Understand the above comment about the Integer Cantor Set and large-scale structure. Show that while there is a universal upper envelope, there is no universal lower envelope. (Hint: consider periodic points!)

Just as the Cantor set can be more easily visualized by drawing the graph of the Canotr function, we can picture the scenery flow as acting on a space of continuous paths.

Thus, defining $\mathcal{C} = \mathcal{C}(\mathbb{R}, \mathbb{R})$ to be the space of continuous functions $f : \mathbb{R} \to \mathbb{R}$, equipped with the topology of uniform convergence on compact subsets, and given $\alpha > 0$, we define a linear map $\tau_t : \mathcal{C} \to \mathcal{C}$ by

$$\tau_s : f(t) \mapsto \frac{f(e^s t)}{e^{s\alpha}}.$$

This is the **scaling flow** of exponent α; here we take $\alpha = d = \log 2 / \log 3$. Regarding other scaling flows see SS12.1.

Proposition 11.1. *The scenery flow on the Cantor set is natually isomorphic to the scaling flow on Cantor functions.*

Proof. ... □

12. Limit theorems of probability theory

In §3.5 above, we explained how given a measure-preserving transformation T of a probability space $(X, \mathcal{A}, \mu)$ and a measurable real-valued function (an *observable*)

f on X, the sequence of functions $f_n = f \circ T^n$ defines a discrete-time stationary stochastic process. Kolmogorov's theorem gives a converse: any stationary process defines a measure-preserving transformation, the shift map on the space of paths, see §33.9.

Despite this equivalence, the points of view are very different. In ergodic theory and dynamical systems we want to understand the dynamics of the map and how that interacts with the structure of the space. Part of this study involves finding natural and interesting invariant measures, then choosing appropriate observables which determine a variety of interesting stochastic processes. In probability theory, the main emphasis is on a given stochastic process which exhibits a certain behavior, such as independence, the Markov or martingale property, with certain corellation or mixing properties. Thus what one observes is the main focus, while the underlying probability space itself has no particular importance.

A further difference is that in ergodic theory, we imagine choosing an initial point x randomly, i.e. with respect to the invariant measure μ, and then following its orbit while measuring the observable along the orbit, giving a *sample path* $f_n(x) = f(T^n x)$ for the stationary process $(f_n)_{n\geq 0}$. In probability theory, by contrast, one thinks of the stochastic process as really a *process*, that is, one sees the values of, say, a coin toss X_i up to time n, and the next value is unknown until one again flips the coin.

Further, we think of the $X_1, X_2, \ldots$ as really "random variables", like the unspecified quantities in elementary algebra but with a given distribution of values. Each distribution is by definition a probability measure P_{X_i} on the reals; if these are the same for all i, the process is termed *identically distributed*. The correlation properties of the process tell us how the next value depends on what came before. In the most random case, as for the tosses of a coin or die or the spin of a roulette wheel, the result at time n is independent of all that came before and all that will come after. This gives an *i.i.d.* (independent and identically distributed) process. Then we are essentially repeating the same experiment over and over again, starting each time from scratch which gives the independence.

Given this strong but natural assumption of i.i.d. plus some information about the distribution P_X, one then proves in probability theory a series of *limit theorems* regarding the long-term behavior.

These are, for example, the Strong Law of Large Numbers (there is also a *weak law* as we explain below), the Central Limit Theorem, and the Law of the Iterated Logarithm.

The first of these, the Strong Law, has been extended far beyond the original i.i.d. setting to Birkhoff's ergodic theorem. For the CLT and LIL, however, one in general (though there are special exceptions- see e.g. **??**) needs something close enough to independence. What conditions on the dynamics of the map and observable can guarantee this is an active theme in research as we introduce in these notes.

Beginning with the probability prespective, let now $(\Omega, \mathcal{A}, P)$ be a probability space, and $X : \Omega \to \mathbb{R}$ a random variable. As explained above in §3, the *expected value*, *mean* or *first moment* of X is just the integral of this measurable function: $\mu = \mathbb{E}(X) = \int_\Omega X(\omega) \mathrm{d}P$. As depicted in Fig, 3, this is indeed the mean value of the measurable function X, and is also the center of mass of its distribution P_X. For $a > 0$, $\mathbb{E}(X^a)$

defines the a^{th} **moment** of X. The most important of these are integer powers, $a = n = 2, 3, \ldots$, defining the *second, third,* and so on, moments.

The second moment of $|X - \mu|$ has a special name, the *variance* of X, written σ^2.

There is a clear connection with the spaces L^p. Recall that given a measure space, for example our probability space $(\Omega, \mathcal{A}, P)$, then for $0 < p < \infty$, $X \in L^p$ iff $\int |X|^p < \infty$, the norm defined by $||X||_p = (\int |X|^p)^{\frac{1}{p}}$ makes L^p a *Banach space*, i.e. a complete normed topological vector space. (So the p^{th} moment is $(||X||_p)^p$). Let us recall that only for $p = 2$ is this an inner product space, hence, since complete, a *Hilbert space*. For real values, the inner product is defined by $\langle X, Y \rangle = \int XY \mathrm{d}P$ and $||X|_2 = \langle X, X \rangle^{\frac{1}{2}}$. (For complex values, as in quantum mechanics, we have a complex vector space of functions $X : \Omega \to \mathbb{C}$ with a Hermitian inner product, see 35.11, [Rud73].) On a finite measure space, finite p^{th} moment implies finite q^{th} moment for any $q < p$, so in particular having any finite p^{th} moment for $p \geq 1$ implies finite expectation. Thus X has finite variance iff it is in the Hilbert space L^2.

Definition 12.1. Random variables X, Y defined on the same underlying space (Ω, P) are **independent** iff for all $A, \mathcal{B} \in \mathcal{A}$, then the sets $X^{-1}(A), Y^{-1}(B)$ are independent. In probability language, the events $[X \in A]$ and $[Y \in B]$ are independent; here by definition $[X \in A] = \{\omega \in \Omega : X(\omega) \in A\} = X^{-1}(A)$.

Lemma 12.1.
(i) If X, Y are independent, then
(*a*)

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y);$$

(*b*) $var(X + Y) = var(X) + var(Y)$.
(ii) If $(X_i)_{i=1}^{\infty}$ are i.i.d., then for $S_n = \Sigma_1^n X_i$, $var(S_n) = n \cdot var(X_1)$.

Proof. Part (*a*) is part of Exercise 3.3: For $X = \chi_A$, $Y = \chi_B$ we have $\chi_A \chi_B = \chi_{A \cap B}$ and the result follows. We extend to simple functions by linearity and then to limits by the Monotone Convergence Theorem, and have proved, as claimed:

$$\int XY \mathrm{d}P = \int X \mathrm{d}P \int Y \mathrm{d}P.$$

For (*b*),
$\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y = \mu + \nu$ so

$$\int_\Omega ((X + Y) - \mathbb{E}(X + Y))^2 \mathrm{d}P = \int_\Omega ((X - \mu)) + (Y - \nu))^2 \mathrm{d}P \tag{7}$$

$$= \int_\Omega (X - \mu)^2 \mathrm{d}P + \int_\Omega (Y - \nu))^2 \mathrm{d}P + \int_\Omega 2(X - \mu)(Y - \nu) \mathrm{d}P = \mathrm{var}(X) + \mathrm{var}(Y), \tag{8}$$

$$\tag{9}$$

by the previous lemma, since $(X - \mu)$ and $(Y - \nu)$ are also independent.

Part (*ii*) follows by induction. □

We mention that in L^2, since X, Y are *orthogonal* iff $\langle X, Y\rangle = 0$, then in particular if X, Y have mean zero, if they are independent then they are orthogonal. (Exercise: find an example of two functions which are orthogonal but *not* independent!)

The next statement (the Strong Law of Large Numbers) is in fact a corollary of the Birkhoff Ergodic Theorem. Nevertheless we present the proof here, for several reasons: not only is this a very important special case, but the proof is quite a bit easier and so is a good place to start to "believe" the more general statement. Also, it provides a good introduction to some methods of proof which occur frequently in probability theory.

We prove the Strong Law in stages, first assuming finite fourth moment, then finite second moment, and finally finte first moment. This simple, direct proof we learned from Marina Talet; it follows unpublished lecture notes of Grimmett.

Theorem 12.2. *(Strong Law of Large Numbers: finite fourth moment) Let $(X_i)_{i=1}^{\infty}$ be an i.i.d. sequence of random variables. We write $S_n = \Sigma_1^n X_i$. If $\mathbb{E}(X_1^4) = \mu < \infty$, then*

$$\frac{S_n}{n} \to \mu \text{ a.s.}$$

Proof.

Lemma 12.3. *We have:*

$$\mathbb{E}\big(S_n - n\mu\big)^4 < Kn^2. \tag{10}$$

Proof. Let $Z_k = X_k - \mu$ and $T_n = Z_1 + \cdots + Z_n = S_n - n\mu$. Then for all $n \geq 1$,

$$\mathbb{E}(T_n^4) = n\mathbb{E}(Z_1^4) + 3n(n-1)\mathbb{E}(Z_1^2 Z_2^2) \leq n(c_1 + c_2 n) \leq n \cdot Kn = Kn^2 \tag{11}$$

for an appropriate constant K, since when we multiply out

$$(Z_1 + \ldots Z_n)(Z_1 + \ldots Z_n)(Z_1 + \ldots Z_n)(Z_1 + \ldots Z_n)$$

the terms come in three types: those where all are the same, where there are two pairs, and where at least two are different, e.g. $Z_1^2 Z_3 Z_4$; this last case has mean zero. (We are using independence, part $i(a)$ of Lemma 12.1). To count the pairs, note that there are $\binom{n}{2} = n(n-1)/2$ ways to choose two letters i, j without order and then 6 ways to rearrange these, e.g. $1122, 1221 \ldots$, giving $3n(n-1)$. □

Now by the Lemma,

$$\mathbb{E}\left(\frac{S_n}{n} - \mu\right)^4 \leq \frac{c}{n^2}$$

Thus

$$\mathbb{E}\sum_1^{\infty}\left(\frac{S_n}{n} - \mu\right)^4 \leq \sum_1^{\infty}\mathbb{E}\left(\frac{S_n}{n} - \mu\right)^4 \leq \infty$$

Now certainly for $0 \leq f$, $\mathbb{E}(f) < \infty$ implies $f < \infty$ almost surely. Thus

$$\sum_1^{\infty}\left(\frac{S_n}{n} - \mu\right)^4 < \infty$$

a.s., and so the terms of this series converge to 0, giving for a.e. ω,

$$\left(\frac{S_n}{n} - \mu\right)^4 \to 0 \implies \frac{S_n}{n} \to \mu.$$

Next we develop two simple but key tools of probability theory: the Borel-Cantelli Lemma and Chebyshev's inequality. Making use of these we give a second proof of the Strong Law for finite fourth moment, which will lead us, combined with interpolation arguments, to the proofs for finite second and first moments.

Here we recall *De Morgan's laws*. These state for logic that when P, Q are propositions then $\sim (P \vee Q) \iff (\sim P \wedge \sim Q)$ and $\sim (P \wedge Q) \iff (\sim P \vee \sim Q)$. For sets $A, B \subseteq X$ the corresponding laws state that $(A \cap B)^c = A^c \cup B^c$ and $(A \cup B)^c = A^c \cap B^c$. This extends to intersections and unions with an arbitrary index set:

$$(\cap_{i \in I} A_i)^c = \cup_{i \in I} A_i^c \text{ and } (\cup_{i \in I} A_i)^c = \cap_{i \in I} A_i^c. \tag{12}$$

Exercise 12.1. Prove the De Morgan logical laws by using truth tables; see [Sig66]. Prove the two-set De Morgan laws from the corresponding logical laws. State a logical law corresponding to the set law with arbitrary index set, and prove both the logical and set versions.

The proof of the (easier) first part of the next item will remind the reader of the first proof we gave of the Poincaré Recurrence Theorem (Theorem 5.1), as both involve the lim sup of a sequence of sets. See Exercise 5.2!

Lemma 12.4. *(Borel-Cantelli) Let $(A_i)_{i=1}^{\infty}$ be measurable subsets of Ω.*
(i) If $\Sigma_{i=1}^{\infty} P(A_i) < \infty$, then $P[A_i$ occurs infinitely often (i.o.) $] = 0$.
(ii) If A_i are independent events, and $\Sigma_{i=1}^{\infty} P(A_i) = \infty$, then $P[A_i$ occurs i.o. $] = 1$.

Proof. The event "A_i occurs infinitely often" is, more formally and precisely, the following subset of Ω: $\{\omega \in \Omega : \omega \in A_i \text{ for infinitely many } i\}$, which equals

$$\limsup A_i = \cap_{N \geq 1} \cup_{i \geq N} A_i.$$

Writing $\widehat{A}_N = \cup_{i \geq N} A_i$, we note that these are nested decreasing, with measure bounded above by $\sum_{i \geq N} P(A_i) < \infty$. This is the tail of a convergent series hence decreases to 0 as $N \to \infty$.

To prove (ii), we shall show that for all N, $P(\widehat{A}_N) = 1$. Now for all $x \in \mathbb{R}$, $1 - x \leq e^{-x}$ (draw the tangent line to the graph of $f(x) = e^{-x}$ at 0); setting $a_i = P(A_i)$, then by De Morgan's law, together with independence,

$$1 - P(\widehat{A}_N) = P(\widehat{A}_N^c) = P((\cup_{i \geq N} A_i)^c) = P((\cap_{i \geq N} A_i^c) = \Pi_{i \geq N} P(A_i^c)$$

and this is the limit as $k \to \infty$ of

$$\Pi_{i=N}^{N+k}(1 - a_i) \leq \Pi_{i=N}^{N+k} \exp(-a_i) = \exp\left(-\Sigma_{i=N}^{N+k} a_i\right)$$

which is zero as for each N, $\Sigma_{i=N}^{\infty} a_i = \infty$. □

Lemma 12.5. *(Chebyshev inequality)*
(i)For $r > 0$, then

$$P[|X| > a] \leq \mathbb{E}(|X^r|)/a^r.$$

(ii)Let f be an increasing function on $\mathbb{R}^+$. Then

$$P[|X| > a] \leq \mathbb{E}(f \circ |X|)/f(a).$$

Proof. Part (i) is a corollary of (ii), taking $f(x) = x^r$ and applying it to the random variable $|X|$. The most common cases in applications will be $r = 1, 2, \ldots$.

For the proof of (ii), we note first that for $g \geq 0$, then

$$cP[g > c] \leq \int_\Omega g(x) \mathrm{d}P \tag{13}$$

(since $0 \leq c \cdot \chi_{[g>c]} \leq g$ (draw the graph!). Assume for simplicity that $X \geq 0$. So for f increasing, then $g = f \circ X \geq 0$ and setting $c = f(a)$, we have

$$f(a) \cdot P[f \circ X > f(a)] \leq \int_\Omega f \circ X \mathrm{d}P$$

but $f \circ X > f(a) \iff X > a$, whence (using the fact that the values of f are ≥ 0, so (13) applies),

$$P[X > a] \leq \mathbb{E}(f \circ X)/f(a).$$

Replacing a general X by $|X|$, we are done. □

Second proof of the SLLN, assuming $\mathbb{E}(X_1^4) < \infty$: Here, following Lamperti [Lam66] pp 26-27, we make use of the two previous lemmas. Without loss of generality, assume $\mu = 0$.

By Chebyshev's inequality,

$$P[|S_n| > \varepsilon n] \leq \frac{\mathbb{E}(S_n^4)}{(\varepsilon n)^4} \tag{14}$$

We have shown above in (10) that

$$\mathbb{E}\big(S_n\big)^4 < Kn^2.$$

Thus

$$P[|S_n| > \varepsilon n] \leq \frac{c}{n^2}. \tag{15}$$

Then by the first part of the Borel-Cantelli Lemma, since this series converges,

$$P[|S_n| > \varepsilon n \text{ infinitely often}] = 0 \tag{16}$$

and indeed $S_n/n \to 0$ almost surely.

Assuming $\mathbb{E}(X_1^2) < \infty$: The idea here is to show convergence along the subsequence n^2, by a similar argument to that just given. To conclude the proof, we then interpolate between these values.

First we show how this interpolation argument will work. We now assume that $X_1 \geq 0$. This will be enough, since we can find X_1^+, X_1^- nonnegative such that $X_1 = X_1^+ - X_1^-$: simply define X_1^+ to be the max of X_1 and 0, similarly for X_1^-; then by linearity we will be done.

Thus we assume that for $X_i \geq 0$, we have proved that

$$\frac{S_{n^2}}{n^2} \to \mu \tag{17}$$

almost surely, and want to conlude that

$$\frac{S_m}{m} \to \mu.$$

Now since $X_i \geq 0$, S_n is nondecreasing in n. Hence we have for any m with $n^2 \leq m \leq (n+1)^2$ that

$$S_{n^2} \leq S_m \leq S_{(n+1)^2}$$

and so

$$\frac{n^2}{(n+1)^2}\frac{S_{n^2}}{n^2} = \frac{S_{n^2}}{(n+1)^2} \leq \frac{S_m}{m} \leq \frac{S_{(n+1)^2}}{n^2} = \frac{S_{(n+1)^2}}{(n+1)^2}\frac{(n+1)^2}{(n)^2}.$$

Since both the right and left-hand sides converge to $\mu = \mathbb{E}(X_1)$, we are done.

Next we prove (17), now with a different hypothesis: without loss of generality, we assume that $\mathbb{E}(X_i) = \mu = 0$. Taking $r = 2$ in part (i) of Chebyshev's Lemma 12.5, and using part (ii) of Lemma 12.1 :

$$P[|S_{n^2}| > n^2\varepsilon] \leq \frac{\text{var}|S_{n^2}|}{n^4\varepsilon^2} = \frac{n^2\text{var}|X_1|}{n^4\varepsilon^2} = \frac{\sigma^2}{n^2\varepsilon^2} \tag{18}$$

which is summable. Therefore, by Borel-Cantelli,

$$P[|S_{n^2}| > \varepsilon n^2 \text{ infinitely often}] = 0. \tag{19}$$

Removing now the assumption that $\mu = 0$ we have:

$$P[|S_{n^2} - n^2\mu| > \varepsilon n^2 \text{ infinitely often}] = 0 \tag{20}$$

and so $S_{n^2}/n^2 \to \mu$ as $n \to \infty$.

Assuming $\mathbb{E}(|X_1|) < \infty$:

The idea here is to prove convergence along an exponential subsequence of times, growing like α^n for $\alpha > 1$. We then interpolate, as for n^2, but now we follow this by taking the limit as α decreases to 1.

This method of proof is originally due to Etemadi [Ete81]. We borrow parts from both [GS92] and [Bor95], filling in details. Here we give special thanks to conversations with M. Talet.

First we make the assumption that $X_i \geq 0$. We define a new, truncated sequence of random variables by

$$Y_i = X_i\chi_{[X_i \leq i]} = \begin{cases} X_i & \text{if } X_i(\omega) \leq i \\ 0 & \text{otherwise} \end{cases}$$

and note that $(Y_i)_{i=1}^{\infty}$ is an independent *not* identically distributed sequence (since the bound i changes with time), for which:
(1) $\mathbb{E}(Y_i) \to \mathbb{E}X_1$ as $i \to \infty$;
(2)

$$\Sigma_{n=1}^{\infty}P[X_n \neq Y_n] = \Sigma_{n=1}^{\infty}P[X_n \geq n] \leq \mathbb{E}(X_1) < \infty.$$

The first statement follows from the Monotone Convergence Theorem: for $f_n \geq 0$ and increasing to f, then $\int f_n \to \int f$. The second can be seen from the tower in Fig. 20 we used for the proof of Kac' Theorem: since the X_i are identically distributed, this is $\Sigma_{n=1}^{\infty}P[X_1 \geq n]$ and $\leq \mathbb{E}(X_1)$ the first level of the tower has mass $P[X_1 \geq 1]$ the

second $P[X_1 \geq 2]$ and so on, and the sum of these equals the integral on the left, which is $\mathbb{E}(X_1)$.

Now consider $A_n = [X_n \neq Y_n]$; we have shown that $\Sigma_{n=1}^{\infty} P(A_n) < \infty$, so by the first part of the Borel-Cantelli Lemma, $P[X_n \neq Y_n$ infinitely often$] = 0$. Therefore almost surely

$$\frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i) \to 0. \tag{21}$$

So it will be enough to show that

$$\frac{1}{N}\sum_{i=1}^{N} Y_i \to \mu. \tag{22}$$

Now we fix $\alpha > 1$ and define $k_n = \lfloor \alpha^n \rfloor$, the integer part of α^n.

Definition 12.2. We make use of the folllowing (standard) notation: given $f, g : \mathbb{R}^+ \to R$, $f \approx g \iff \lim_{t\to\infty} f(t)/g(t) = 1$, and similarly for sequences. This is called *asymptotic equivalence*.

For example, we claim that

$$k_n \approx \alpha^n. \tag{23}$$

We have:

$$\alpha^n - 1 \leq k_n \leq \alpha^n$$

Therefore

$$\frac{\alpha^n - 1}{\alpha^n} \leq \frac{k_n}{\alpha^n} \leq \frac{\alpha^n}{\alpha^n} = 1$$

so the limit is 1.

It follows that

$$\frac{k_{n+1}}{k_n} \approx \alpha \tag{24}$$

which we shall need below.

We write $S'_n = \sum_{i=1}^{n} Y_i$. Then by Chebyshev's inequality, independence of the Y_i, and the fact that $\mathrm{var}(Y_i) \leq \mathbb{E}(Y_i^2)$, we have

$$\sum_{n=1}^{\infty} P[\frac{1}{k_n}|S'_{k_n} - \mathbb{E}S'_{k_n}| > \varepsilon] \leq \frac{1}{\varepsilon^2}\sum_{n=1}^{\infty}\frac{\mathrm{var}S'_{k_n}}{k_n^2} = \frac{1}{\varepsilon^2}\sum_{n=1}^{\infty}\frac{1}{k_n^2}\sum_{i=1}^{k_n}\mathrm{var}Y_i \tag{25}$$

$$\leq \frac{1}{\varepsilon^2}\sum_{n=1}^{\infty}\frac{1}{k_n^2}\sum_{i=1}^{k_n}\mathbb{E}(Y_i^2) \tag{26}$$

$$\leq C\sum_{i=1}^{\infty}\frac{\mathbb{E}(Y_i^2)}{i^2} \tag{27}$$

We explain this last step: the idea is to change the order of summation, just like changing the order of integration of an iterated integral. We set $f(t) = \alpha^t$. Since $\alpha^n - 1 \leq k_n \leq \alpha^n$, $1 \leq i \leq k_n$ iff $1 \leq i \leq \alpha^n = f(n)$. Now for $s = f(t)$,

$f^{-1}(s) = t = \log s/\log\alpha$. Thus "$1 \le n \le \infty$ and for each fixed n, $1 \le i \le k_n$" is equivalent to "$1 \le i \le \infty$ and for each fixed i, $f^{-1}(i) \le n$", that is, $n \ge \log i/\log\alpha$.

Since from (23) $k_n \approx \alpha^n$, $1/k_n^2 \approx 1/\alpha^{2n}$ so we have, for $r = \log i/\log\alpha$,

$$\frac{1}{\varepsilon^2}\sum_{n=1}^{\infty}\frac{1}{k_n^2}\sum_{i=1}^{k_n}\mathbb{E}(Y_i^2) = \frac{1}{\varepsilon^2}\sum_{i=1}^{\infty}\mathbb{E}(Y_i^2)\sum_{n\ge r}\frac{1}{k_n^2} \approx \frac{1}{\varepsilon^2}\sum_{i=1}^{\infty}\mathbb{E}(Y_i^2)\sum_{n\ge r}\frac{1}{\alpha^{2n}} \tag{28}$$

Now for a positive decreasing function $g : \mathbb{R} \to \mathbb{R}$, then for any $r \in \mathbb{R}$,

$$\sum_{i\ge r} g(i) \le \int_{r-1}^{\infty} g(x)\mathrm{d}x.$$

Here $g(x) = 1/\alpha^{2x} = e^{-x\log\alpha^2}$, so for $C_1 = 1/2\log\alpha$, $\int_s^{\infty} g(x)\mathrm{d}x = C_1\alpha^{-2s}$. Note that

$$\alpha^{-2r} = \frac{1}{i^2}.$$

So we have, for $C = C_1\alpha^2$,

$$\sum_{n\ge\log i/\log\alpha}\frac{1}{\alpha^{2n}} = \sum_{i\ge r} g(i) \le \int_{r-1}^{\infty} g(x)\mathrm{d}x = C_1\alpha^{-2(r-1)} = C_1\alpha^2\alpha^{-2r} = \frac{C}{i^2}.$$

Next, writing P_X for the distribution of X, then

$$\mathbb{E}(Y_i^2) = \int_0^i x^2\mathrm{d}P_X = \sum_{k=0}^{i-1}\int_k^{k+1} x^2\mathrm{d}P_X.$$

Thus our sum is

$$(28) \le C\sum_{i=1}^{\infty}\frac{1}{i^2}\sum_{k=0}^{i-1}\int_k^{k+1} x^2\mathrm{d}P_X. \tag{29}$$

Now for $a_k \ge 0$

$$\sum_{i=1}^{\infty}\frac{1}{i^2}\sum_{k=0}^{i-1} a_k = 1\cdot a_0 + \frac{1}{4}(a_0+a_1) + \frac{1}{9}(a_0+a_1+a_3) + \dots \tag{30}$$

$$= a_0\sum_{i\ge 1}\frac{1}{i^2} + \dots + a_m\sum_{i\ge m+1}\frac{1}{i^2} + \dots \tag{31}$$

$$\le \sum_{k\ge 1}\frac{a_k}{k} \tag{32}$$

because

$$\sum_{i\ge m+1}\frac{1}{i^2} \le \int_m^{\infty}\frac{1}{x^2}\mathrm{d}x = \frac{1}{m}.$$

Thus, using the facts that for $x \in [k, k+1]$, $1/(k+1) \leq 1/x$ and that $(k+1)/k \leq 2$, (29) is

$$\leq C \sum_{k\geq 1} \frac{1}{k} \int_k^{k+1} x^2 \mathrm{d}P_X \leq 2C \sum_{k\geq 1} \frac{1}{k+1} \int_k^{k+1} x^2 \mathrm{d}P_X \tag{33}$$

$$\leq 2C \sum_{k\geq 1} \int_k^{k+1} \frac{1}{x} x^2 \mathrm{d}P_X \leq 2C \sum_{k\geq 0} \int_k^{k+1} x \mathrm{d}P_X = 2C\mathbb{E}(X_1) < \infty. \tag{34}$$

We have shown that

$$\sum_{n=1}^{\infty} P\left[\left|\frac{S'_{k_n}}{k_n} - \frac{\mathbb{E}(S'_{k_n})}{k_n}\right| > \varepsilon\right] \leq \infty. \tag{35}$$

Now since $\mathbb{E}(Y_i) \to \mathbb{E}X_1$,

$$\frac{\mathbb{E}S'_{k_n}}{k_n} \to \frac{\mathbb{E}S_{k_n}}{k_n} = \mathbb{E}(X_1) = \mu.$$

Hence also

$$\sum_{n=1}^{\infty} P\left[\left|\frac{S'_{k_n}}{k_n} - \mu\right| > \varepsilon\right] \leq \infty. \tag{36}$$

Therefore by the Borel-Cantelli Lemma, almost surely

$$\frac{S'_{k_n}}{k_n} \to \mu.$$

Next we interpolate. For $k_n \leq m \leq k_{n+1}$, we have

$$\alpha^n - 1 \leq k_n \leq m \leq k_{n+1} \leq \alpha^{n+1}$$

thus

$$\frac{S'_{k_n}}{k_n}\frac{k_n}{k_{n+1}} \leq \frac{S'_m}{m} \leq \frac{S'_{k_{n+1}}}{k_{n+1}}\frac{k_{n+1}}{k_n} \tag{37}$$

By (24) we know that $k_{n+1}/k_n \approx \alpha$.
Thus we shown that for all $\alpha > 1$, almost surely

$$\alpha^{-1}\mu \leq \liminf \frac{S'_m}{m} \leq \limsup \frac{S'_m}{m} \leq \alpha\mu \tag{38}$$

whence almost surely, by letting $\alpha \to 1$,

$$\frac{S'_m}{m} \to \mu.$$

and now by (22), we are done!

□

12.1. **Random walk and the CLT.** A different view of these limit theorems comes from considering the stochastic process $(S_n)_{n\geq 0}$ defined from the partial sums. Thus, let X_i be an i.i.d. sequence of random variables taking values in $\mathbb{R}$ with probabilities given by a probability distribution ρ. We define S_n for $n \geq 0$ by $S_0 = 0$, $S_n = \sum_{k=0}^{n-1} X_k$; then S_n is a **random walk** with **independent increments** (the increments of a process S_n being $X_n \equiv S_{n+1} - S_n$), or an **i.i.d. random walk** for short.

For the most basic example, the **simple random walk**, ρ is the distribution on $\{-1, 1\} \subseteq \mathbb{R}$ giving each of these points equal probability $1/2$.

From the ergodic theory point of view, we can begin with the Bernoulli shift $\sum^+ = \Pi_0^\infty\{0, 1\}$ with the left shift map σ, and define $f(\underline{x}) = 1$ if $x_0 = 1$,$= -1$ if $x_0 = 0$; then $S_n = \sum_{k=0}^{n-1} f(\sigma^k(\underline{x}))$. Conversely, for any i.i.d. random walk S_n, the increment process $(X_0, X_1, \dots)$ is a point in the shift space $\Pi^+ = \Pi_0^\infty\mathbb{R}$ with independent product measure $\otimes_0^\infty \rho$.

The location of the random walk path S_n at time n has a distribution we can understand by drawing Pascal's triangle, which counts the number of paths from the initial point to another vertex. Note that to arrive at a vertex we have to pass through a vertex above it, so the number is the sum of these if there are two. Considering the numbers occuring along level n of Pascal's triangle, starting at level 0 we have a single 1, then $1, 1$, next $1, 2, 1$ and for level 3 we have $1, 3, 3, 1$ and so on; note the symmetry and that the total along level n is 2^n. The formula for k heads in n tosses is for the number of ways to choose k items from a set of n elements, without order; this is "n choose k" or

$$\binom{n}{k} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{k!}$$

If we assign transition probabilities $1/2$ to each downward edge, see Fig. 25, and multiply these along each path, then the total becomes 1 along each level, a probability measure $\boldsymbol{\pi}_n$ on $\{0, 1, \dots, n\}$, where $\boldsymbol{\pi}_n(k)$ is the number of coin-tosses of length n such that there are k heads.

The path space for the random walk (connected by polygonal interpolation, to give the **polygonal random walk**) can be visualized as a Pascal's triangle turned on its side, Fig. 26.

We translate the distribution $\boldsymbol{\pi}_n$ to the left by $n/2$; it now has mean 0. A graph for $n = 49$ is shown in Fig. 12.1. This can be shown to, in the limit (we normalize the scale so the variance is always 1) to converge to the *Gaussian* or *standard normal* distribution Φ, where "standard" specifies that this probability distribution has mean 0 and variance $\sigma^2 = 1$. The type of convergence is that which is natural for measures: on each open interval J, $\boldsymbol{\pi}_n(J) \to \Phi(J)$. This is called in probability theory *weak convergence* or *convergence in distribution*; from the point of view of analysis it is $weak - *$ convergence; see... below. The formula for Φ is:

$$\Phi(J) = \int_J \varphi(x)\mathrm{d}x$$

where

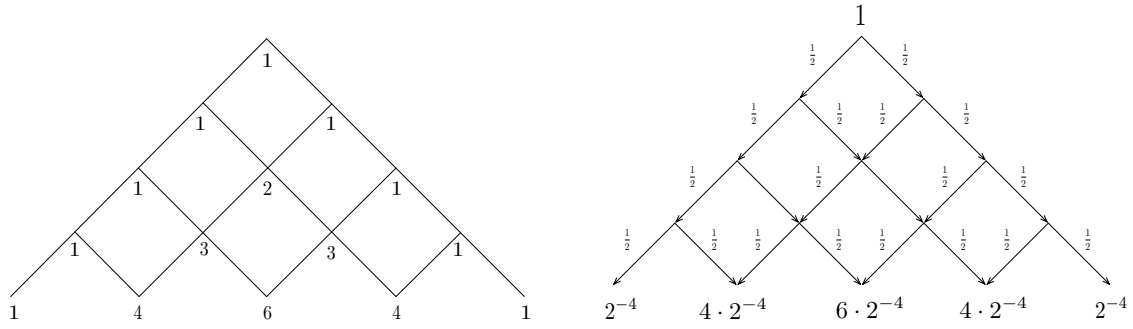


FIGURE 25. Pascal's triangle, with and without weights

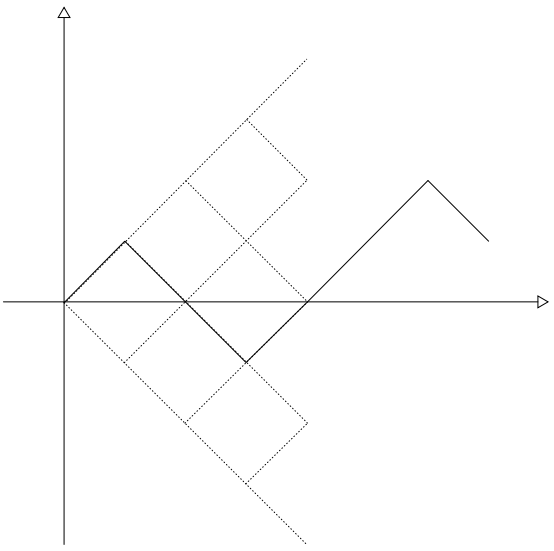


FIGURE 26. A random walk on the integers

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2};$$

see Fig. 12.1.

The reason for these factors is, as one recalls from Calculus (a beautiful argument using a double integral together with polar coordinates) that $\int_{-\infty}^{\infty} e^{-x^2} \mathrm{d}x = \sqrt{\pi}$; division by 2 gives a probability measure with variance 1.

We take a different perspective on this example in §15.4, as a countable state Markov shift.

We have already seen via the Strong Law of Large Numbers that the mean value of the partial sums converges to the mean value of X_i, that is, $S_n/n \to 0$ for a.e. sequence of the Bernoulli tosses of our fair coin.

The further classical limit theorems of probability theory emerge from considering a different rescaling, by $\sqrt{n}$.

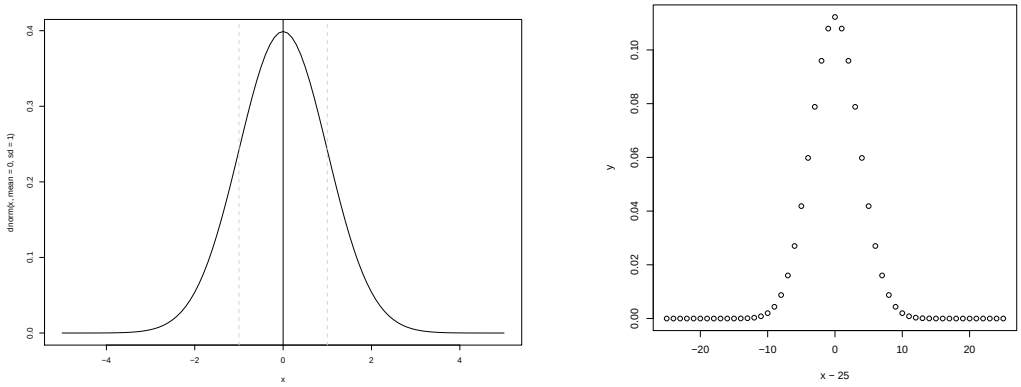


FIGURE 27. Gaussian distribution with mean 0 and variance $\sigma^2 = 1$, approximated by a binomial distribution of parameter $n+1$ for $n = 49$ steps of the simple random walk. The height of the dot indicates the value of the point mass.

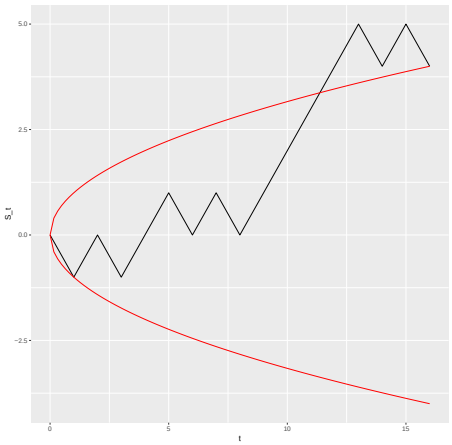


FIGURE 28. Bernoulli random walk, rescaled: $n = 2^4$ steps.

We define for $a > 0$ and $\alpha > 0$ the *scaling transformation* $\Delta_a : \mathcal{C} \to \mathcal{C}$ by

$$(\Delta_a(f))(t) = \frac{f(at)}{a^\alpha}.$$

In this case, we take $\alpha = 1/2$; Wiener measure ν on $\mathcal{C}$ is preserved by Δ_a, and the parabola $h(x) = \pm x^{\frac{1}{2}}$ is a fixed point.

In keeping the dimensions of the figures constant, the graphics program we used (rStudio) has essentially automatically rescaled the graphs by Δ_a for $a = 2^n$. The parabola is $\{(x, y) : x = y^2\}$; all parabolas $y = \pm c\sqrt{x}$ for $c > 0$ are fixed points for the scaling flow of exponent $1/2$, and this one is of special imprtance as it indicates $\pm$ the location of the variance, $\sqrt{t}$, for the distribution at time t, and gives the scaling used for the CLT.

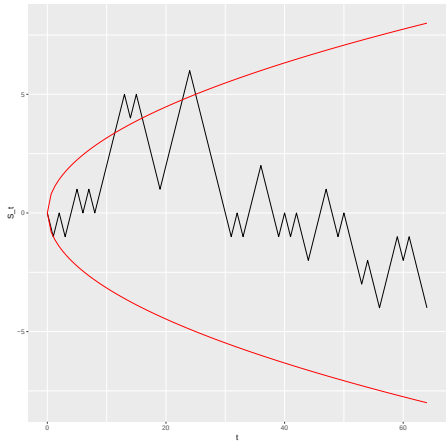


FIGURE 29. Bernoulli random walk: $n = 2^6$ steps.

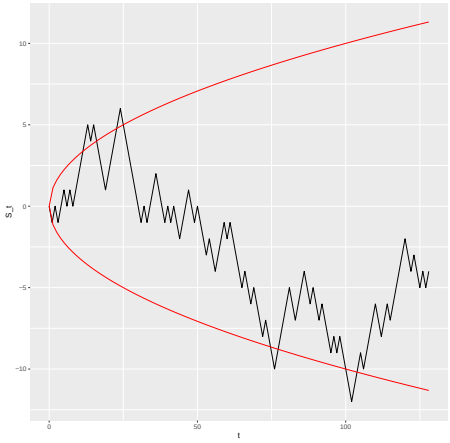


FIGURE 30. Bernoulli random walk: $n = 2^6$ steps.

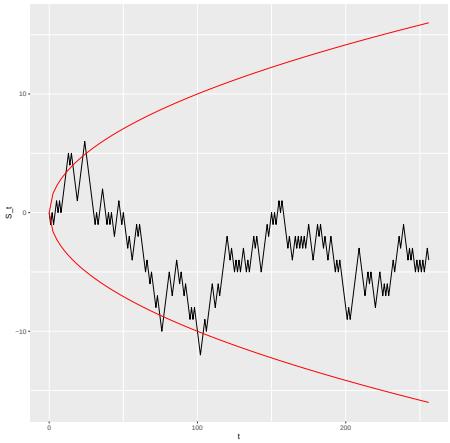


FIGURE 31. Bernoulli random walk: $n = 2^7$ steps. The parabola $\sqrt{t}$ is a scaling flow fixed point, and indicates a constant times the variance.

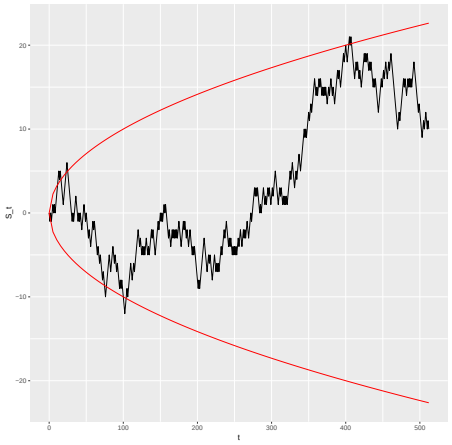


FIGURE 32. Bernoulli random walk: $n = 2^8$ steps. These graphs were all produced with the same data (i.e. the same random number generator seed) so you can follow how the graph gets rescaled.

What these figures show is the convergence under scaling of the random walk paths to paths for a continuous-time stochastic process with nowhere differentiable, but continuous, paths: *Brownian Motion*. Indeed, for the last figure, we show a random walk with i.i.d. normal increments. Such an increment sequence X_k is (by the Markov property of Brownian motion) given by $B(k+1)-B(k)$ for a Brownian path B. More precisely, the map $\Phi : B \mapsto (X_k)_{k\geq 0}$ is measure-preserving from Wiener measure ν on $\mathcal{C}$ to $\Pi^+ \equiv \Pi_0^\infty$ with infinite product measure $\mu = \otimes_0^\infty \rho$, where ρ is the normal distribution on $\mathbb{R}$. This gives a good picture of Brownian motion, since a.e. such random walk path is embedded in an actual Brownian path.

The challenge is to think about *in what sense* one has convergence to this rather crazy process. A dynamical approach is described below, via the *scaling flow* τ_s on *path space* $\mathcal{C}$, defined by $\tau_s = \Delta_{\exp(s)}$.

But for now we describe a (remarkable) first step.

Definition 12.3. Let (X, d) be a metric space, and μ a measure on X. Given a bounded uniformly continuous function $f : X \to \mathbb{R}$, we write $\mu(f) \equiv \langle \mu, f\rangle \equiv \int_X f \mathrm{d}\mu$. Note that this defines a bilinear map $\langle \cdot, \cdot \rangle : \mathcal{M} \times \mathcal{UCB} \to \mathbb{R}$ where $\mathcal{M}$ is the vector space of signed Borel measures on X; such a map (generalizing the idea of inner product) is known as a *pairing* of the two spaces.

Recall that given a probability space $(\Omega, \mathcal{A}, P)$ and $f : \Omega \to X$, the *distribution* of f is $P_f = P \circ f^{-1}$, the push-forward of P by f.

(i)Let $(\mu_n)_{n\geq 0}$ be measures on X. We say that $\mu_n \to \mu$ *weak** ("weak-star") or, in the probability terminology, *weakly*, iff for each bounded uniformly continuous function $f : X \to \mathbb{R}$,

$$\mu_n(f) \to \mu(f)$$

as $n \to \infty$.

(i) Given and a sequence of random variables f_i with values in the metric space (X, d), we say that $f_i \to f$ *in distribution* iff the distributions converge weakly, i.e. iff $P_{f_i} \to P_f$ weakly.

Theorem 12.6. *(Central Limit Theorem) Let $(X_i)_{i\geq 0}$ be i.i.d. random variables with mean zero and variance one, and let $(S_n)_{n\geq 0}$ denote the random walk $S_0 = 0, S_n = \sum_{i=0}^{n-1} X_i$. Then*

$$\frac{S_n}{\sqrt{n}} \to e^{-x^2/2}dx \text{ in distribution, as } n \to \infty.$$

That is, for all $a \in \mathbb{R}$,

$$P\left[\frac{S_n}{\sqrt{n}} < a\right] \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2}dx$$

as $n \to \infty$.

For the proof of the CLT we need some facts about the Fourier transform, some basics of which we dig into next.

12.2. **Fourier series and transforms.** The Fourier transform will be defined for a variety of functions: elements of L^1, L^2 of the reals. Further, this initial definition can be extended to finite measures and beyond that to tempered Schwartz distributions. We mention that this makes sense much more widely, to $\mathbb{R}^n$, and other locally compact abelian groups, this is the subject of Abstract Harmonic Analysis, and even to compact nonabelian groups, which takes one into Representation Theory. See e.g.

The intuitive initial idea is that the Fourier transform of a real-or complex-valued function on $\mathbb{R}$ should map us from position to frequency space, or in Quantum Mechanics, from position to momentum space.

This is supposed to imitate the Fourier series of a periodic function. Thus, for example, the spectral analyis of a periodic wave with one pure frequency, such as $\sin(t\cdot)$ or the complex wave $e^{it\cdot}$ should be a pure frequency, that is, point mass δ_t at frequency t. One should also have an inversion formula, as for Fourier series, that takes us back. And, as for that case, this inverse map $\check{\mathcal{F}}$ should essentially just be the Fourier transform itself.

However things are not quite this simple, which is what makes the subject so interesting!

We begin with $L^2(\mathbb{R}) = L^2(\mathbb{R}, \mathbb{C}; m)$ where for notational simplicity we take m to be Lebesgue measure normalized as follows: $\mathrm{d}m = \frac{1}{\sqrt{2\pi}}\mathrm{d}x$.

As above Theorem 6.1, we make $L^2(\mathbb{R})$ into a Hilbert space by defining the inner product $\langle f, g\rangle = \int_{\mathbb{R}} f\overline{g}\mathrm{d}m$.

Now by part (iii) of Theorem 6.1, the dual space of $L^\infty(\mathbb{R})$ is $ba(\mathbb{R})$, the bounded additive signed measures. If $\lambda \in ba$, and $g \in L^\infty$, we write this pairing as $\langle \lambda, g\rangle = \int_{\mathbb{R}} \overline{g}\mathrm{d}\lambda$. Given $f \in L^1$, we know that L^1 embeds in ba, by integration. That is, we can write this pairing as $\langle f, g\rangle = \int_{\mathbb{R}} f\overline{g}\mathrm{d}m$.

For $t \in \mathbb{R}$ we write $e_t : \mathbb{R} \to \mathbb{R}$ for the function

$$e_t(x) = \frac{1}{\sqrt{2\pi}} e^{itx}.$$

Noting that from Euler's formula $e^{i\theta} = cos\theta + i\sin\theta$, we have $\overline{e}_t = e_{-t}$.

Given $f \in L^2(\mathbb{R}, \mathbb{C})$, we define its *Fourier transform* $\widehat{f} = \mathcal{F}(f)$ by

$$\widehat{f}(t) = \langle f, e_t \rangle = \int_{\mathbb{R}} f\overline{e}_t \mathrm{d}m = \int_{\mathbb{R}} f(x)e_{-t}(x)\mathrm{d}m(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x)e^{-itx}\mathrm{d}x. \quad (39)$$

For the first formula we have used the above observation regarding $f \in L^1 \subseteq ba$, and that $e_t \in L^\infty$.

We define the *inverse Fourier transform* $\check{f} = \check{\mathcal{F}}(f)$ by:

$$\check{f}(t) = \langle f, \overline{e}_t \rangle = \int_{\mathbb{R}} f e_t \mathrm{d}m = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x)e^{itx}\mathrm{d}x. \quad (40)$$

Note that $\check{f}(t) = \widehat{f}(-t)$.

Lemma 12.7. *On the Hilbert space $L^2(\mathbb{R})$, $\mathcal{F}$ is an isometry, with inverse map $\check{\mathcal{F}}$.*

As we shall see, the Fourier transform can be extended by linearity and continuity to a variety of other spaces, of functions, measures and Schwartz distributions. Note that the definition of $\widehat{f}$ and $\check{f}$ also makes sense for L^1, and beyond that to finite measures, since e_t is a bounded function so the integrals clearly exist there. However, the inverse transform may not be defined on $\widehat{f}$, as we see by example.

Clearly, the basic idea of the Fourier transform is to generalize the spectral (i.e. frequency) analysis of Fourier series to the real line.

To get an intuitive feeling for this map, we begin with some examples:

–For δ_0, point mass at 0, the transform of $\widehat{\delta_0}$ is the constant function 1

–More generally, $\widehat{\delta_t} = e_{-t} = e^{-it}$, and so by linearity, the transform of $\sum_{-n}^{n} a_k\delta_k$ is ...

sin/cos series...

Note however that for these examples, $\check{\mathcal{F}}$ is not defined on these transforms, as they arr nonintegrable functions.

We note further that:

–the Gaussian function $e^{-x^2/2}$ is a (unique, up to multiples by a nonzero constant) fixed point for the Fourier transform. This is normalized for the measure $\mathrm{d}m$, as it has mean zero, integral one, and variance one.

–the transform of the Gaussian function $e^{-x^2/2}$ scaled by σ is scaled by $1/\sigma$.

Note that this relates to $\mathcal{F}(\delta_0) = 1$: the sequence $\varphi_n = \ldots$ converges to δ_0, and....converges to 1.

Convolution.... φ_n is an *approximate identity* in that....

These are all related to the *uncertainty principle*: that for $f \in L^2$,

$$||f||_2||\widehat{f}||_2 \geq 1.$$

Thus, on the space $\S = \S(\mathbb{R})$ of *tempered distributions*, $\mathcal{F}$ is an isometry, with inverse map $\check{\mathcal{F}}$.

Given a measure μ on $\mathbf{r}$ we define $\mathcal{F}(\mu) = \widehat{\mu} = g$ where

$$g(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-itx}\mathrm{d}\mu \quad (41)$$

Here is an example from probability theory:

Definition 12.4. Given a $\mathbb{R}$- or $\mathbb{C}$- valued random variable X defined on a probability space $(\Omega, \mathcal{A}, P)$ then the *characteristic function* of X is the inverse Fourier transform of the distribution P_X of X.

Thus ...???

Remark 12.1. Much of the usefulness of the Fourier transform in analysis is due to its basic properties, see Theorem 7.2 of [Rud73]; in particular, that
-convolution is turned into multiplication.

In probability theory (see Chapter 15 of [**?**], the normalized *inverse* Fourier transform is known as the *characteristic function* of a probability distribution on $\mathbb{R}$. One shows that sums of i.i.d. random variables yield convolutions of their distributions and so again one can use multiplication in extimates. We shall next see some important applications of this idea.

It is not an accident that the Gaussian distribution is a fixed point for the Fourier transform. In physics, the Fourier transform takes you from position to momentum space, and in wave analysis from position space to frequency space.

We mention that the Laplace transform has similar properties, since in faact the Laplace transform on the imaginary axis is exactly the characteristic function, hence also is of use in probability, where it is known as the *moment generating function*. See Postscript 1 in Chapter 15 of [**?**].

12.3. **Proof of the CLT.** We shall need the following basic fact from analysis:

Lemma 12.8.

$$e^x = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n \tag{42}$$

and hence also

$$\lim_{n\to\infty} \left(1 + \frac{x}{n} + o\left(\frac{x}{n}\right)\right)^n = e^x. \tag{43}$$

Proof. The first statement is a basic fact from analysis, one of the common equivalent definitions of the exponential function (the others being as a solution of a differential equation, or as the Taylor expansion). It can be proved from the Taylor series by using the Binomial Theorem to turn the product into a series. We skip the details (see e.g. Wikipedia, *exponential function*.)

Recall that that $f(x) = o(x)$ means for all $\varepsilon > 0$, $\exists x_0$ such that for all $x \geq x_0$ we have $|f(x)| < \varepsilon x$. We write this as

$$f(x) = \pm\varepsilon x.$$

Thus $f(x) = o(x)$ is equivalent to: "given $\varepsilon > 0$, eventually $f(x) = \pm\varepsilon x$." Now (43) means given a function $f(x)$ such that $f(x) = o(x)$ then

$$\lim_{n\to\infty} \left(1 + \frac{x}{n} + f(x/n)\right)^n = e^x$$

but,

$$\left(1 + \frac{x}{n} + \pm\varepsilon\frac{x}{n}\right)^n = \left(1 + \frac{x}{n}(1 \pm \varepsilon)\right)^n \to e^{(1\pm\varepsilon)x}$$

for every ε, proving the statement.

□

Lemma 12.9. *Suppose that* $\varphi : \mathbb{R} \to \mathbb{R}$ *satisfies*

$$\varphi(x) = 1 - \frac{1}{2}x^2 + o(x^2).$$

Then for

$$\varphi_n(x) = \left(\varphi\left(\frac{x}{n^{\frac{1}{2}}}\right)\right)^n = \left(1 - \frac{x^2}{2n} + o\left(\frac{x^2}{n}\right)\right)^n,$$

we have

$$\varphi_n \to e^{-x^2/2}$$

as $n \to \infty$.

Proof. This now follows from (43). □

The idea for proving the CLT is as follows. The Fourier transform will convert the partial sum to a product, where we apply the above argument. The Gaussian is a fixed point for the Fourier transform. Taking the Fourier transform again, proves the CLT.

Proof. (of CLT) We take X_n i.i.d. with finite second moment, moreover for simplicity of the constants, with mean zero and variance one. Thus $S_n = X_1 + \cdots + X_n$ has variance n.

Let P_X be the distribution of $X = X_k$ (for any k).

Let us write $\mathcal{F}$ for the Fourier transform operator.

Let $\varphi_X = \mathcal{F}(P_X)$ denote the Fourier transform of P_X. Then because X has first and second moments $0, 1$ we have

$$\varphi_X(x) = 1 - \frac{1}{2}x^2 + o(x^2).$$

The Fourier transform of the distribution of S_n is

$$\mathcal{F}(P_n) \equiv \mathcal{F}(P_{S_n}) = \varphi_n(x) = \left(\varphi\left(\frac{x}{n^{\frac{1}{2}}}\right)\right)^n.$$

Hence by the Lemma, $\lim_{n\to\infty} \mathcal{F}(P_n) = e^{-x^2/2}$. Also, $\mathcal{F}(\lim_{n\to\infty} P_n) = \lim_{n\to\infty} \mathcal{F}(P_n)$ by linearity and continuity of the Fourier transform.

Thus

$\lim_{n\to\infty} P_n = \mathcal{F}(\mathcal{F} \lim_{n\to\infty} P_n) = \mathcal{F}(e^{-x^2/2}) = e^{-x^2/2}\mathrm{d}x$, finishing the proof. □

The CLT describes convergence of the distribution of the random walk at time n to the *standard normal* distribution (that with mean $\mu = 0$ and variance $\sigma^2 = 1$). This applies to the simple random walk, where S_n has the binomial distribution of point masses shown in Fig. 12.1. But it applies to *any* i.i.d. $(X_i)_{i\geq 0}$ with finite positive variance, including continuous distribution or a mixture of continuous and discrete.

When one changes the formula to include any μ and any $\sigma^2 > 0$, the CLT states that

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}}$$

converges to the standard normal.

There are also precise asymptotic upper and lower envelopes for the random walk path, curves slightly larger than the above parabola which follows the variance of S_n:

Theorem 12.10. *(Law of the Iterated Logarithm) For $(X_i)_{i\geq 0}$ as above, then almost surely,*

$$P\left[\limsup \frac{|S_n|}{\sqrt{n \log\log n}} = \sqrt{2}\right] = 1.$$

We note that X_1 having finite p^{th} moment implies finite q^{th} moment for any $q < p$, so having any finite p^{th} moment for $p \geq 1$ implies finite expectation (hence the strong law) and finite p^{th} moment for $p \geq 2$ implies the distributional convergence of the CLT, and the upper and lower bounds of the Law of the Iterated Logarithm (which of course implies the strong law for this case).

But what can one say for $0 < p < 2$, still with the i.i.d. assumption? This is a fascinating story answered by Paul Lévy: one then gets the α-stable distributions for $\alpha \in (0, 1)$. There now is an additional skewness parameter $\xi \in [-1, 1]$, with $\xi = 0$ being the *symmetric* stable law (*law* means distribution!) and $\xi = \pm 1$ being the *completely asymmetric* case.

Lévy completely characterized those distributions which converge to the (α, ξ)-stable laws not only for the standard normalization $S_n/n^{1/\alpha}$ but for any normalization $S_n/a(n)$. This is called the *basin of attraction* of the (α, ξ)-stable law. The characterization is in terms of the Fourier transform of the distribution. This can be thought of as a *stable central limit theorem*; see [Lam66] and also see the introduction to **??**.

Most conventionally in applications to ergodic theory and dynamical systems the only limit law that comes up other than the Ergodic Theorem (i.e. the Strong Law) is the CLT, but that is not quite true: all of the laws for $\alpha > 0$, completely asymmetric with $\xi = 1$, occur naturally in the context of expanding maps with indifferent fixed points. (For $\alpha \in (0, 1)$ one gets the Mittag-Leffler distributions, which correspond to the inverses of the stable processes). See Fig.??

Furthermore, Poisson distributions and Pareto distributions can naturally occur. See ??

–Markov process

–Limit theorems

–recurrence; infinite measure

12.4. **Brownian motion and the scaling flow.** We consider the space $\mathcal{C}^+ = \mathcal{C}(\mathbb{R}^+, \mathbb{R})$, of continuous functions from $\mathbb{R}^+$ to $\mathbb{R}$, with the topology $\mathcal{T}$ of uniform convergence on compact subsets of $\mathbb{R}^+$. This topology makes $\mathcal{C}^+$ a Polish space (Def. 5.2), which is good from the point of view of measure theory. We call this *one-sided path space*. Similarly we define *two-sided path space* to be $\mathcal{C} = \mathcal{C}(\mathbb{R}, \mathbb{R})$.

We define a flow on on path space $\mathcal{C}$, respectively $\mathcal{C}^+$, by $(\tau_s B)(t) = B(e^s t)/e^{s/2}$. This is the *scaling flow* τ_s of exponent $1/2$.

Theorem 12.11. *There exists a unique probability measure ν on $\mathcal{C}^+$ satisfying:*
(i) $B(0) = 0$, for a.e. $B \in \mathcal{C}$;
(ii) the increments *$B(t+s) - B(s)$ for $t > 0$ are independent of $B(r)$ for all $r < s$.*
(iii) the distribution of $B(t+s) - B(s)$ is $\mathcal{N}(0,t)$.

Parts $(i) - (iii)$ can be summarized as: B has stationary, in fact i.i.d. Gaussian increments, with mean zero and variance $\mathrm{var}(B(t)) = t$.

Thus the distribution of $B(t)$ is $\mathcal{N}(0,t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} dx$.

This is called the *Wiener process* or *Brownian Motion*. The measure ν^+ can be extended to $\mathcal{C}(\mathbb{R}, \mathbb{R})$, i.e. to paths $B(t)$ with $t \in \mathbb{R}$; we call this *two-sided path space*. We do this by taking an independent copy of the one-sided process, reflecting the paths and joining them at time 0.

We define the *increment flow* h_s on $\mathcal{C}$ by

$$h_s : B(t) \mapsto B(t+s) - B(s).$$

Theorem 12.12.
(i)The scaling and increment flows for Brownian motion are ergodic, moreover are each Bernoulli flows of infinite entropy.
(ii)Let S_n be a random walk on $\mathbb{R}$ with i.i.d. increments X_i of mean 0 and variance 1. Write $S(t)$ for the corresponding polygonal path in $\mathcal{C}^+$, and μ for the measure on path space corresponding to $S(t)$. Then μ-a.e. path $S(t)$ is a generic point for the scaling flow $(\mathcal{C}^+, \nu, \tau_t)$.
iii the two flows satisfy the commutation relation

$$h_b \tau_a = \tau_a h_{e^{-s}b}.$$

For proofs of $(i), (ii)$, see [Fis87], [FT12]. Part (iii) is immediately verified. This says that the following diagram commutes; a consequence is that the increment flow has the interesting property of being isomorphic to a speeded-up copy of itself! Compare the discussion of the geodesic and horocycle flows below in Fig. 66.

$$\begin{array}{ccc} \mathcal{C} & \xrightarrow{h_{e^{-a}\cdot b}} & \mathcal{C} \\ {\scriptstyle \tau_a}\uparrow & & \uparrow{\scriptstyle \tau_a} \\ \mathcal{C} & \xrightarrow[h_b]{} & \mathcal{C} \end{array}$$

12.5. **Fundamental solution of the heat equation.** There is a close relationship between the Heat Equation and Brownian motion: the *fundamental solution* to the heat equation is just the distribution of $B(t)$! For one dimension (this is true much more generally, for example on manifolds with a Riemannian metric), this fundamental solution is therefore $\Phi : \mathbb{R} \times \mathbb{R}^+ \to$
R^+ where

$\Phi(x,t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} dx$. The interpretation is that one begins at time 0 with point mass concentrated at 0; it evolves as a normal distribution with variance t at time t.

See Fig. 12.4.

Click HERE for rotatable 3-d image!

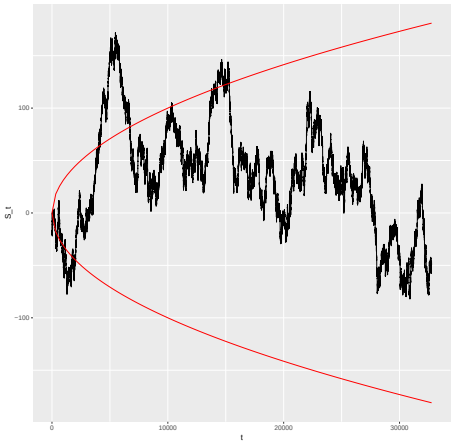


FIGURE 33. Typical path of an i.i.d. random walk with standard normal increments, here with $n = 2^{15}$ steps, now close to a Brownian path..

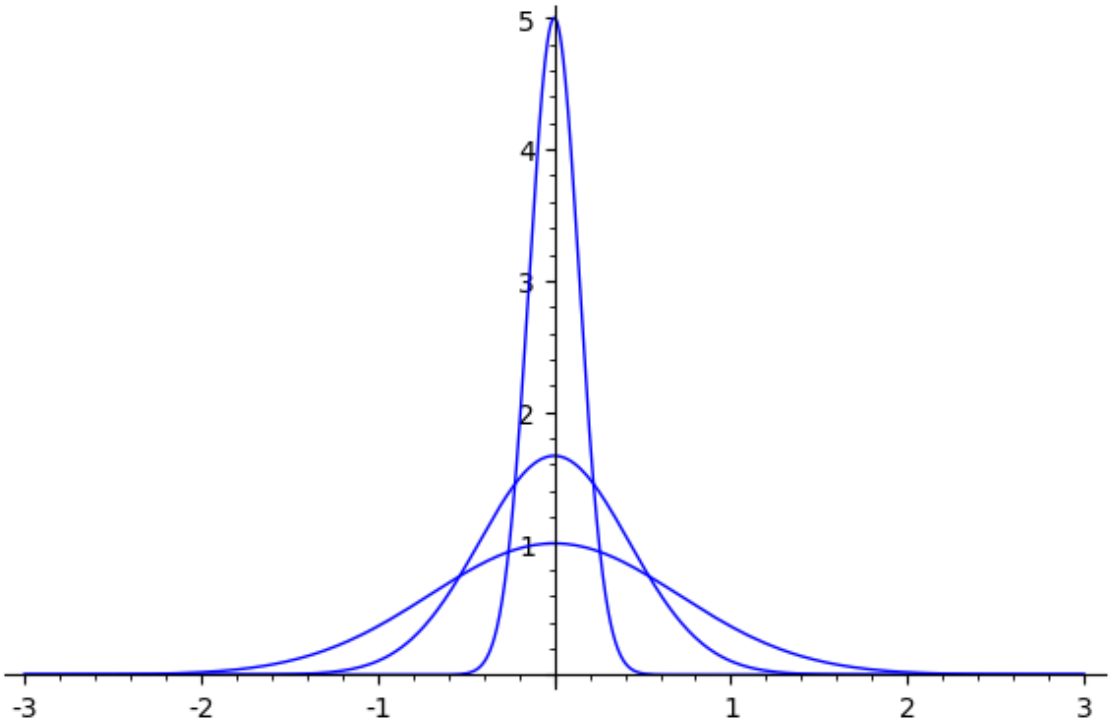


FIGURE 34. Distribution of $B(t)$: evolution of fundamental solution

12.6. **The shift flow on the Ornstein-Uhlenbeck velocity process.**

12.7. **The shift flow on White Noise.**

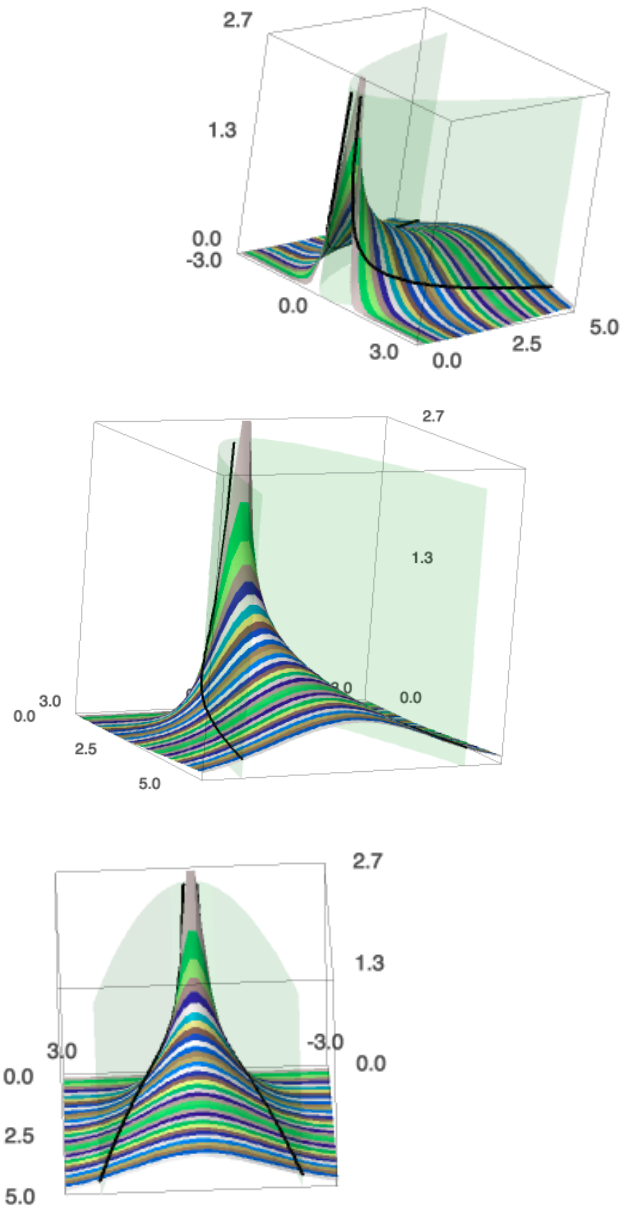


FIGURE 35. Fundamental solution of heat equation

13. Proofs of ergodicity

14. Weak mixing, eigenfunctions and rotations

Here we follow mostly [Fur81]; property (*iii*) is so central that Furstenberg takes it as his *definition* of weak mixing!

Proposition 14.1. *These are equivalent, for a measure-preserving transformation T of a probability space $(X, \mathcal{A}, \mu)$:*

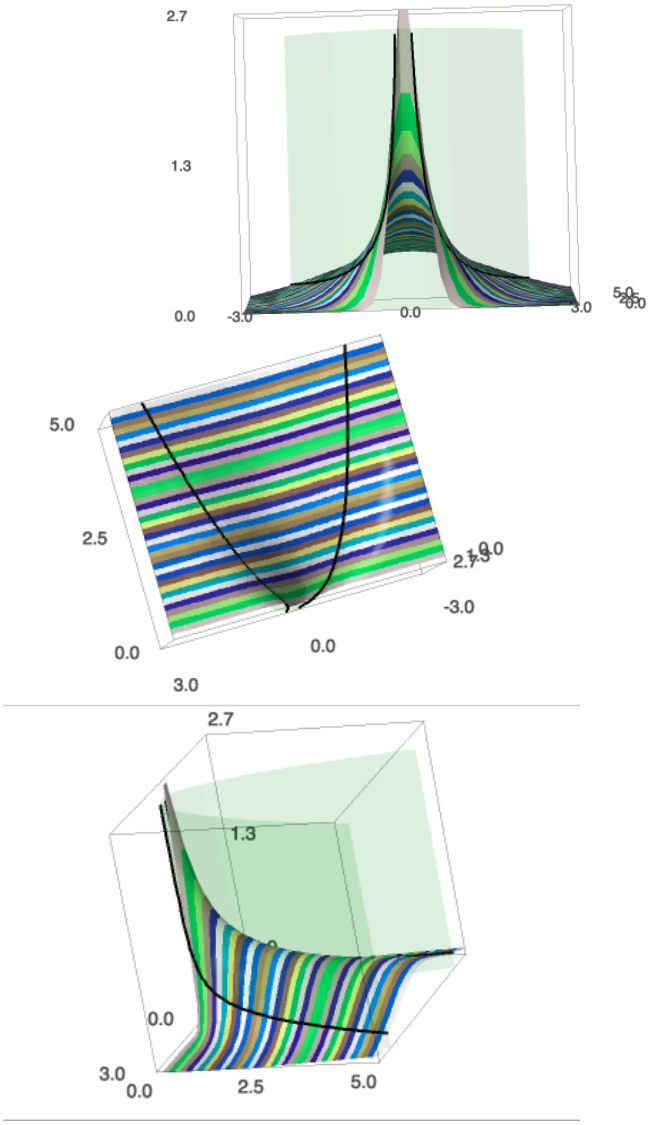


FIGURE 36. The black curve projects to a parabola on the $t-x$ plane: it is where a vertical paraboloid $t = c \cdot x^2$ (indicating constant σ) meets the surface.

(i) T is weak mixing
(ii) $T \times T$ is weak mixing
(iii) $T \times T$ is ergodic
(iv) for any ergodic measure-preserving transformation of a probability space $(Y, \mathcal{B}, \nu, S)$, $T \times S$ is ergodic.

Exercise 14.1. Given T an irrational circle rotation R_θ, show that $T \times T$ (which is a map on the torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$) is not ergodic.

For the proof of the proposition we shall need:

Lemma 14.2. *For a measure-preserving transformation T of a probability space $(X, \mathcal{A}, \mu)$, then T is mixing if and only if $T \times T$ is mixing.*

Proof. If $T \times T$ is mixing then T is: we just consider the sets $A \times X$, $B \times X$. For the converse,

$$\mu \times \mu(T \times T)^{-k}(A \times C) \cap (B \times D) = \mu \times \mu(T^{-k}A \cap B) \times (T^{-k}C \cap D) \tag{44}$$

$$= \mu(T^{-k}A \cap B)\mu(T^{-k}C \cap D) \to \mu A \mu B \mu C \mu D = \mu \times \mu(A \times C)\mu \times \mu(B \times D). \tag{45}$$

But it is enough to have verified mixing for rectangles, as they generate the product $\sigma-$algebra. □

Proof. of Proposition: Applying the lemma to convergence along a subsequence of density one, by 8.6 we have a proof of $(i) \iff (ii)$.

For $(ii) \implies (iii)$ we apply Proposition 8.1 to the transformation $T \times T$.

We next show that $(iii) \implies (i)$. By Lemma 8.5 it will be enough to show that

$$\frac{1}{n}\sum_{k=0}^{n-1}(\mu(A \cap T^{-k}B) - \mu A \mu B)^2 \to 0.$$

We have:

$$\frac{1}{n}\sum_{k=0}^{n-1}(\mu(A \cap T^{-k}B) - \mu A \mu B)^2 \tag{46}$$

$$= \frac{1}{n}\sum_{k=0}^{n-1}(\mu(A \cap T^{-k}B))^2 + (\mu A \mu B)^2 - 2\mu(A \cap T^{-k}B)\mu A \mu B. \tag{47}$$

Now

$$\frac{1}{n}\sum_{k=0}^{n-1}(\mu(A \cap T^{-k}B))^2 = \frac{1}{n}\sum_{k=0}^{n-1}\mu \times \mu(A \times A) \cap (T \times T)^{-k}(B \times B) \to (\mu A)^2(\mu B)^2 \tag{48}$$

by the ergodicity of $T \times T$. And

$$\frac{1}{n}\sum_{k=0}^{n-1}\mu(A \cap T^{-k}B) \to \mu A \mu B$$

since T is ergodic. Therefore

$$\frac{1}{n}\sum_{k=0}^{n-1}\big(\mu(A \cap T^{-k}B) - \mu A \mu B\big)^2 \to 2(\mu A \mu B)^2 - 2(\mu A \mu B)^2 = 0$$

finishing the proof.

Next is $(iv) \implies (iii)$: assume that T is such that for any S ergodic, the product $T \times S$ is ergodic. Let Y be a singleton with point mass and S the identity map on Y; this is ergodic. So $T \times S$ is ergodic by (iv), and this is isomorphic to T, and we have that T itself is ergodic. Then again by (iv), $T \times T$ is ergodic, giving (iii).

Finally we show $(i) \implies (iv)$. Write $\widehat{X} = X \times Y$, $\widehat{A} = \mathcal{A} \times \mathcal{B}$, $\widehat{\mu} = \mu \times \nu$ and $\widehat{T} = T \times S$. From Prop. 8.3, $\widehat{T}$ is ergodic iff for any $A, B \in \widehat{\mathcal{A}}$, $f_1 = \chi_A, f_2 = \chi_B$ we have

$$\frac{1}{N}\sum_{n=0}^{N-1}\int_{\widehat{X}}(f_1)\ (f_2 \circ \widehat{T}^n)\mathrm{d}\widehat{\mu} \to \int_{\widehat{X}} f_1 \mathrm{d}\widehat{\mu}\int_{\widehat{X}} f_2 \mathrm{d}\widehat{\mu} \text{ as } N \to \infty. \tag{49}$$

We prove this for $L^2(\widehat{X}) = L^2(X \times Y)$, where it will be enough to verify for a dense set: the collection of all finite sums of functions of the form $f(x, y) = g(x)h(y)$ for $g \in L^2(X), h \in L^2(Y)$.

For this we verify (49) for $f_1(x, y) = g_1(x)h_1(y)$, $f_2(x, y) = g_2(x)h_2(y)$, extending to the rest of the dense set by linearity (of the integral and the sum). Now by Fubini's theorem

$$\int_{\widehat{X}} f_1 \mathrm{d}\widehat{\mu}\int_{\widehat{X}} f_2 \mathrm{d}\widehat{\mu} = \int_X g_1 \mathrm{d}\mu \int_X g_2 \mathrm{d}\mu \int_X h_1 \mathrm{d}\mu \int_X h_2 \mathrm{d}\mu. \tag{50}$$

We wish to show that

$$\frac{1}{N}\sum_{n=0}^{N-1}\int_{\widehat{X}} f_1 \cdot\ f_2 \circ \widehat{T}^n \mathrm{d}\widehat{\mu} = \frac{1}{N}\sum_{n=0}^{N-1}\left(\int_X g_1 \cdot\ g_2 \circ T^n \mathrm{d}\mu\ \int_Y h_1 \cdot\ h_2 \circ S^n \mathrm{d}\nu\right) \tag{51}$$

converges to (50) as $N \to \infty$.

...

OBS: can use erg of T also

...

We carry this out first for g_1 constant, then for $\int g_1 = 0$; the general case follows by linearity.

For g_1 constant, $\int_X g_1 \cdot\ g_2 \circ T^n \mathrm{d}\mu = \int_X g_1 \mathrm{d}\mu \int_X g_2 \circ T^n \mathrm{d}\mu = \int_X g_1 \mathrm{d}\mu \int_X g_2 \mathrm{d}\mu$ (since $\int_X g_2 \circ T^n \mathrm{d}\mu = \int_X g_2 \mathrm{d}(\mu \circ T^{-n}) = \int_X g_2 \mathrm{d}\mu$) so these constants can be pulled out of the sum and (51) reduces to showing:

$$\frac{1}{N}\sum_{n=0}^{N-1}\int_Y h_1 \cdot\ h_2 \circ S^n \mathrm{d}\nu \to \int_Y h_1 \mathrm{d}\nu \int_Y h_2 \mathrm{d}\nu$$

which is true by the ergodicity of S.

We move on to the case $\int g_1 = 0$, where we use a different trick to separate the factors under the sum of (51). We recall Hölder's inequality:

$$||\phi\psi||_1 \leq ||\phi||_2 |||\psi||_2 \tag{52}$$

which we use in these two forms, one for L^2 and one for $\mathbb{R}^d$:

$$\left(\int_X \phi\psi \mathrm{d}\mu\right)^2 \leq \left(\int_X |\phi\psi| \mathrm{d}\mu\right)^2 \leq (||\phi||_2)^2 (||\psi||_2)^2 \tag{53}$$

and:

$$\left(\frac{1}{N}\sum_{k=0}^{N-1} a_k b_k\right)^2 \leq \left(\frac{1}{N}\sum_{k=0}^{N-1} |a_k b_k|\right)^2 \leq \frac{1}{N}\sum_{k=0}^{N-1} |a_k|^2 \frac{1}{N}\sum_{k=0}^{N-1} |b_k|^2. \tag{54}$$

Defining $a_k = \int_X g_1 \cdot\ g_2 \circ T^n$ and $b_k = \int_Y h_1 \cdot\ h_2 \circ S^n \mathrm{d}\nu$, we are to show that

$$\frac{1}{N}\sum_{k=0}^{N-1} a_k b_k \to (50)$$

which equals 0 in this case. Now from (54),

$$\left(\frac{1}{N}\sum_{k=0}^{N-1} a_k b_k\right)^2 \le \frac{1}{N}\sum_{k=0}^{N-1} |a_k|^2 \frac{1}{N}\sum_{k=0}^{N-1} |b_k|^2. \tag{55}$$

We shall show the first of these $\to 0$ while the second remains bounded. From Lemma 8.5, $\frac{1}{n}\sum_{k=0}^{n-1}|c_k| \to 0 \iff \frac{1}{n}\sum_{k=0}^{n-1}|c_k|^2 \to 0$; taking

$$c_k = \int_X g_1 \cdot\ g_2 \circ T^n \mathrm{d}\mu - \int_X g_1 \mathrm{d}\mu \cdot \int_X g_2 \mathrm{d}\mu$$

then having

$$\frac{1}{N}\sum_{n=0}^{N-1}\left(\int_X g_1 \cdot\ g_2 \circ T^n \mathrm{d}\mu - \int_X g_1 \mathrm{d}\mu \cdot \int_X g_2 \mathrm{d}\mu\right)^2 \to 0$$

for $g_1, g_2 \in L^2$ is equivalent to weak mixing of T. We are assuming T is weak mixing, so this is true for our particular choice of g_1 which has integral 0, whence $\frac{1}{N}\sum_{n=0}^{N-1}\left(\int_X g_1 \cdot\ g_2 \circ T^n \mathrm{d}\mu\right)^2 = \frac{1}{N}\sum_{k=0}^{N-1}|a_k|^2 \to 0$. On the other hand, by (53),

$$\frac{1}{N}\sum_{k=0}^{N-1}|b_k|^2 = \frac{1}{N}\sum_{n=0}^{N-1}\left(\int_Y h_1 \cdot\ h_2 \circ S^n \mathrm{d}\nu\right)^2 \le \frac{1}{N}\sum_{n=0}^{N-1}(||h_1||_2)^2\ (||h_2||_2)^2 = (||h_1||_2)^2\ (||h_2||_2)^2$$

which is finite. Hence (55)$\to 0$, finishing the proof. □

Two consequences are:

Corollary 14.3.
(a)If T is weak mixing then $T \times T \times \cdots \times T$ is weak mixing.
(b)If T is weak mixing then T^m is weak mixing.

Proof. The proof of (a) will use several of the elements of Prop. 14.9. Since T is weak mixing, by (iii) $T \times T$ is ergodic, so by (iv) $T \times (T \times T)$ is ergodic. By induction, $S = (T \times T \times \cdots \times T)$ (n times) is ergodic. That isn't quite enough! But so is $S \times S = (T \times T \times \cdots \times T)$ ($2n$ times); hence now by (ii), we do know that S is weak mixing.

For an alternative argument, we wish to prove (iv): that for any S ergodic, $(T \times T \times \cdots \times T) \times S$ is ergodic. But since T is weak mixing, by $(iv) T \times S$ is ergodic; again by (iv) $T \times (T \times S) = (T \times T) \times S$ is ergodic, and by induction we are done.

To prove (b), consider the discrete space $Y = \mathbb{Z}_m$ with normalized counting measure ν and with transformation S the cyclic permutation, $k \mapsto k + 1 (\mathrm{mod}\, m)$. This is ergodic, so by (iv) above, $T \times S$ on the product space $X \times \mathbb{Z}_m$ is ergodic. Now the induced map of $T \times S$ on the set $X \times \{0\}$ is isomorphic to T^m; and we know that an induced map is ergodic iff that holds for the original map. Thus T^m is ergodic. To show it is weak mixing, we repeat the proof beginning with $T \times T$, since that is weak mixing; so $(T \times T)^m \cong T^m \times T^m$ is ergodic, and by (ii) T^m is weak mixing. □

Definition 14.1. $f \in \mathcal{L}^2(X,\mu)$ (with complex values) is a **eigenfunction** for T with **eigenvalue** λ if $f \neq 0$ and $U(f) = \lambda f$ for U the Koopman operator $f \mapsto f \circ T$.

Lemma 14.4. *If f is an eigenfunction for T, then $|\lambda| = 1$. If T is ergodic then $|f|$ is constant, and if T is weak mixing, then f is constant.*

Proof. Since U is unitary, $\langle f, f\rangle = \langle Uf, Uf\rangle = \langle \lambda f, \lambda f\rangle = |\lambda|^2\langle f, f\rangle$ so since $f \neq 0$, $|\lambda| = 1$.

If T is ergodic, then $|f(Tx)| = |\lambda| \cdot |f(x)| = |f(x)|$ so $|f|$ is constant, and since f is an eigenfunction this constant is nonzero.

If T is weak mixing, then $T \times T$ is ergodic. Since $|f| \neq 0$, the exists a set X_1 of measure one such that $f(x) \neq 0$ for every $x \in X_1$; for $(x,y) \in X_1 \times X_1$, define $g(x,y) = f(x)/f(y)$. Then $g(Tx,Ty) = \lambda f(x)/\lambda f(y) = g(x,y)$ is constant (say $= c$) on a subset $\widehat{X} \subseteq X_1 \times X_1$ of measure one; this may not be symmetric, but it contains a symmetric set of measure one (take the intersection with its image under the map $(x,y) \mapsto (y,x)$) and so now, switching x and y, $c = 1$ and then $f(x) = f(y)$ so f is indeed a.s. constant as well. □

Definition 14.2. A **rotation factor** is R_θ acting on $\mathbb{R}/\mathbb{Z}$ for some $\theta \in [0,1)$ together with an invariant probability measure μ. We know that there are two cases: $\theta \notin \mathbb{Q}$ and μ is Lebesgue measure, or θ is rational and μ is normalized counting measure on a finite invariant subset of the circle.

Theorem 14.5. *These are equivalent:*
(i) T is weak mixing;
(ii) T has no nonconstant eigenfunction;
(iii) T has no rotation factor.

Proof. We just proved $(i) \implies (ii)$ (in Lemma 14.12).

We show $(ii) \iff (iii)$. First we remark that the equation for an eigenfunction, $f(Tx) = \lambda f(x)$, has this dynamical interpretation: we have the commutative diagram

$$\begin{array}{ccc} X & \xrightarrow{T} & X \\ \downarrow f & & \downarrow f \\ f(X) & \xrightarrow{\cdot\lambda} & f(X) \end{array} .$$

We can improve this semiconjugacy so the image $f(X) \subseteq \mathbb{C}$ is contained in the circle, by normalization: indeed, if T has a nonconstant eigenfunction f then we know its eigenvalue is $\lambda = e^{i\theta}$, for some $\theta \in [0, 2\pi)$. Now define $\varphi : X \to S^1 = \{z \in \mathbb{C} : |z| = 1\}$ by $\varphi(x) \equiv f(x)/|f(x)|$ and define R_θ on S^1 by $z \mapsto \lambda z$. Then $\varphi \circ T(x) = \lambda f(x)/|\lambda||f(x)| = \lambda\varphi(x)$ so we have this commutative diagram:

$$\begin{array}{ccc} X & \xrightarrow{T} & X \\ \downarrow \varphi & & \downarrow \varphi \\ S^1 & \xrightarrow{R_\theta} & S^1 \end{array}$$

Let $\nu = \mu \circ \varphi^{-1}$ be the pushed-forward measure on S^1. Then φ is a semiconjugacy from (X,μ,T) to (Y,ν,R_θ) where $Y \subseteq S^1$ is the image of X. Since this latter map is a

factor of an ergodic transformation, it too must be ergodic. There are two possibilites, that θ is rational or irrational, leading to the two types of rotation factor.

To prove the converse, given a rotation factor we have the above commutative diagram. We define $f(x) = \varphi(x)$, and note that this is a (nonconstant) eigenfunction with eigenvalue λ, since $f(Tx) = \varphi(Tx) = R_\theta(\varphi(x)) = \lambda f(x)$. □

Definition 14.3. A transformation has **pure point spectrum** iff the collection $\mathcal{E}$ of its eigenfunctions spans $L^2(X, \mu)$.

Example 16. We note that for any $\theta \in [0, 1)$, R_θ has pure point spectrum; indeed there are many more eigenfunctions for R_θ : for $n \in \mathbb{Z}$ and $g = e^{in\theta}$ the eigenvalue is λ^n. In multiplicative notation, for $g(z) = z^n$ then $g(\lambda z) = \lambda^n z^n$. And these form a basis for L_2 of the circle. Note that the spectrum is finite or dense depending on the irrationality of the angle.

Example 17. -skew products

-isometric extensions
-subsequence ergodic theorem via random ergodic theorem
-statement of Furstenberg-Zimmer

14.1. **Weak mixing for flows.**

Definition 14.4. Let τ_t be a measure-preserving flow on a probability space $(X, \mathcal{A}, \mu)$. The flow is **mixing** iff for every $A, B \in \mathcal{A}$ we have for
(i)

$$\mu(A \cap \tau_t B) \to \mu A \mu B \text{ as } t \to \pm\infty.$$

It is **weak mixing** iff
(ii)

$$\frac{1}{T}\int_{t=0}^{T} |\mu(A \cap \tau_t B) - \mu A \mu B| \to 0 \text{ as } T \to \pm\infty.$$

We also consider, as for transformations, the third condition:
(iii)

$$\frac{1}{T}\int_{t=0}^{T} \mu(A \cap \tau_t B) \to \mu A \mu B \text{ as } T \to \pm\infty.$$

As for transformations,

Proposition 14.6. *For flows on a probability space, mixing implies weak mixing implies ergodic.*

Lemma 14.7. *For a bounded $f : \mathbb{R} \to \mathbb{C}$ these are equivalent:*
(a)

$$\frac{1}{T}\int_{t=0}^{T} |f(x)|\,dx \to 0$$

(b)

$$\frac{1}{T}\int_{t=0}^{T} |f|^2(x)\,dx \to 0$$

(c)

$$f(x) \to 0 \text{ in density },$$

i.e. along a set of times of Cesáro density one in the positive reals.

Proof. In fact, the arguments given above for sequences in the proof of Lemma 8.5 generalize immediately to the integrals. □

Proposition 14.8. *These are equivalent, for a measure-preserving flow τ_t of a probability space $(X, \mathcal{A}, \mu)$:*
(i) τ_t is weak mixing;
(ii) Given $A, B \in \mathcal{A}$, there is a subset $K \subseteq \mathbb{R}^+$ of Cesáro density one such that τ_t is mixing along K; that is,

$$\lim_{t \in K; t \to \infty} \mu(A \cap \tau_t B) \to \mu A \mu B.$$

Proposition 14.9. *These are equivalent, for a measure-preserving flow τ_t of a probability space $(X, \mathcal{A}, \mu)$:*
(i) τ_t is weak mixing
(ii) $\tau_t \times \tau_t$ is weak mixing
(iii) $\tau_t \times \tau_t$ is ergodic
(iv) for any ergodic measure-preserving flow η_t of a probability space $(Y, \mathcal{B}, \nu)$, $\tau_t \times \eta_t$ is ergodic.

Example 18. The rotation flow on the circle $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, $\tau_t : x \mapsto x + t (\text{mod} 1)$ is ergodic, but the flow $\tau_t \times \tau_t$ on the torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ is not ergodic.

Lemma 14.10. *For a measure-preserving flow τ_t of a probability space $(X, \mathcal{A}, \mu)$, then τ_t is mixing if and only if $\tau_t \times \tau_t$ is mixing.*

Two consequences are:

Corollary 14.11.
(a)If τ_t is weak mixing then $\tau_t \times \tau_t \times \cdots \times \tau_t$ is weak mixing.
(b)If τ_t is weak mixing then τ_s is a weak mixing transformation for each $s \in \mathbb{R}$.

Definition 14.5. $f \in \mathcal{L}^2(X, \mu)$, we define the action of the Koopman one-parameter group $U_t : f \mapsto f \circ \tau_t$. We say f is a **eigenfunction** for τ_t with **eigenvalue** $\lambda \in \mathbb{C}$ iff $f \neq 0$ and for every t, $U_t(f) = \lambda^t f$.

Lemma 14.12. *If f is an eigenfunction for τ_t, then $|\lambda| = 1$. If τ_t is ergodic then $|f|$ is constant, and if τ_t is weak mixing, then f is constant.*

Definition 14.6. A **rotation factor** is $R_{\theta t} = R_\theta^t$ acting on $\mathbb{R}/\mathbb{Z}$ for some nonzero $\theta \in \mathbb{R}$ together with the unique invariant measure μ, equal to Lebesgue measure. Thus $R_{\theta t}(x) = x + \theta t (\text{mod} 1)$.

This is just a linear time-change of the speed-one rotation flow R_t, where $\theta = 1$.

Theorem 14.13. *These are equivalent:*
(i) τ_t is weak mixing;
(ii) τ_t has no nonconstant eigenfunction;
(iii) τ_t has no rotation factor.

15. Markov shifts

15.1. **Markov measures on the full shift.** We recall some of the notation and definitions from §4.3. We begin with a finite set $\mathcal{A}$ called the **alphabet**, consisting of d **symbols** or **letters**; these can also be called **digits** in the case where $\mathcal{A} = \{1, 2, \dots, d\}$ or $\mathcal{A} = \{0, 1, 2, \dots, d-1\}$.

We write $\Sigma = \Pi_{-\infty}^{\infty}\mathcal{A}$; we define the **left shift map** $\sigma : \Sigma \to \Sigma$: for $x = (\dots x_{-1}.x_0x_1\dots) \in \Sigma$, then $(\sigma(x))_i = x_{i+1}$. To keep track of locations, we have placed a "decimal point" immediately to the left of the 0^{th} coordinate; thus $\sigma : (\dots x_{-1}.x_0x_1\dots) \mapsto (\dots x_{-1}x_0.x_1\dots)$. The **one-sided shift space** is $\Sigma^+ = \Pi_0^{\infty}\mathcal{A}$, acted on by the left shift map, (also denoted σ, with the same definition for $i \geq 0$, so now $\sigma : (.x_0x_1\dots) \mapsto (.x_1\dots)$. Giving the set $\mathcal{A}$ the discrete topology and Σ the product topology, then Σ is a compact topological space which is metrizable; a convenient metric is $d(x, y) = 2^{-m}$ where $m = \inf\{i : x_i \neq y_i\}$.

Exercise 15.1. *The spaces Σ^+ and Σ are homeomorphic to the middle-thirds Cantor set, the shift on Σ is a homeomorphism and on Σ^+ it is a $d-$to-1 continuous map. (See Exercise 4.7.)*

We define a **thin cylinder set** to be a subset of Σ of the form, for $k \leq m \in \mathbb{Z}$, $[x_k \dots x_m] = \{w \in \Sigma : w_k = x_k, \dots, w_m = x_m\}$ this is a clopen (closed and open) set. The collection of these is denoted $\mathbb{C}_k^m$. The decimal point again helps us keep track of the 0^{th} coordinate; taking $\mathcal{A} = \{1, 2, \dots, d\}$, then $[.2] \in \mathbb{C}_0^0$ and $[01.0] \in \mathbb{C}_{-2}^0$. A **general cylinder set** is a finite union of thin cylinders; we write $*$ for "no restriction on the symbols" so e.g. for an alphabet with 3 symbols, some general cylinders which are unions of sets in $\mathbb{C}_0^4$ are $[**.*2*]$ or $[.*12*0]$.

The cylinder sets are clopen sets which generate the topology and hence the Borel $\sigma-$algebra $\mathcal{B}$ for Σ_A. For the one-sided shift space for Σ_A^+, the thin cylinders $[.x_0 \dots x_m]$ generate the $\sigma-$algebra $\mathcal{B}^+$.

Exercise 15.1. Let $\mathcal{B}_0^+$ denote the algebra generated by the collection of thin cylinder sets. Show that this algebra contains all general cylinder sets, and consists of all sets which are finite unions of thin cylinders.

We write Δ for the unit simplex in $\mathbb{R}^d$: $\Delta \equiv \{\boldsymbol{\pi} = (\pi_1, \dots, \pi_d) : \pi_i \geq 0, \sum_{i=1}^d \pi_i = 1\}$. Thus Δ is the convex set spanned by the standard basis column vectors. An element $\boldsymbol{\pi}$ of Δ is a **probability vector**; a choice of $\boldsymbol{\pi}$ serves to define a probability distribution on the alphabet, equivalently on the collection $\mathbb{C}_0^0$ of 0-cylinder sets. We next examine what is needed to extend this to all of $\mathcal{B}^+$.

One says M is **row-stochastic** iff each of its rows is a probability vector. Writing $\underline{1}$ for the column vector with entries identically 1, then M is row-stochastic iff $M\underline{1} = \underline{1}$. Other common names for this are a **probability** or **stochastic** matrix.

We shall need the following, the proof of which is delayed until Lemma 16.6:

Lemma 15.1. *A $(d \times d)$ matrix M is row-stochastic if and only if it is nonnegative and $\Delta^t M \subseteq \Delta^t$ M.*

But first we relax these hypotheses, so as to see exactly what is needed. Thus, given a $(d \times d)$ $\mathbb{C}-$ or $\mathbb{R}-$ valued matrix M and a vector $\boldsymbol{\pi}$, we define a function μ

on the collection of all thin cylinder sets in Σ_A^+ as follows:

$$\mu([x_0 \dots x_m]) = \pi_{x_0} M_{x_0 x_1} \cdots M_{x_{m-1} x_m} \tag{56}$$

We have:

Lemma 15.2. *M satisfies $M\underline{1} = \underline{1}$ iff μ as defined in (56) extends by additivity to general cylinder sets of Σ_A^+, in which case it is finitely additive on the algebra generated by the thin cylinder sets. μ is $\sigma-$invariant if and only if $\boldsymbol{\pi}^t M = \boldsymbol{\pi}^t$. In this case, μ extends to an invariant function on the algebra for the two-sided shift space.*

If M and $\boldsymbol{\pi}$ are nonnegative, then μ is a finitely additive measure on Σ_A^+, and extends in a unique way to be countably additive on the Borel $\sigma-$algebra of Σ_A^+. This is a probability measure iff $\boldsymbol{\pi} \in \Delta$, and as above extends to Σ_A by invariance iff $\boldsymbol{\pi}^t M = \boldsymbol{\pi}^t$.

Proof. As in Exercise 15.1, the algebra $\mathcal{B}_0^+$ generated by the thin cylinders is the collection of all finite unions of thin cylinders. So we extend μ to such a set A by additivity. Thus for example consider a cylinder set which terminates in a given symbol, taking e.g. $\mathcal{A} = \{0, 1\}$, with $A = [.11 * 0]$, we add the contributions from the thin cylinders of length 4 which make it up, which is well-defined as that decomposition is unique.

Note however that a thin cylinder set is itself a union of longer thin cylinders; thus e.g. $[.11] = [.11*] = [.111] \cup [.110]$. But here, $\mu([.11]) = \mu([.11*]) = \mu([.111] \cup [.110]) = \mu([.111]) + \mu([.110]))$, making use of the fact that $M\underline{1} = \underline{1}$.

If $A = A_1 \cup A_2$, disjoint, with each a union of thin cylinders, then we let n be the longest length of any of these cylinders, and decompose each of the cylinders for A_1, A_2 into cylinders of length n. Then by the previous observation the measure of each A_i is the sum of the measures of these longer cylinders, and moreover, $\mu(A) = \mu(A_1) + \mu(A_2)$. Thus μ is defined additively on $\mathcal{B}_0^+$.

We observe that this part remains valid for real or complex entries.

For the case of nonnegative entries, M is stochastic, and μ is a finitely additive measure on the algebra $\mathcal{B}_0^+$. The extension to a countably additive measure on the Borel sigma-algebra $\mathcal{B}^+$ is guaranteed by Alexandroff's extension theorem, Theorem 33.15 below, making use of the fact that the shift space is compact.

To prove shift-invariance it is sufficient to check this on thin cylinders; we need to show that e.g. $\mu([.11])$ is equal to $\mu(\sigma^{-1}([.11])) = \mu([.111] \cup [.011])$ which by additivity we know is equal to $\mu([.111]) + \mu([.011])$; this now follows from the hypothesis $\boldsymbol{\pi}^t M = \boldsymbol{\pi}^t$. Thus μ is invariant on the one-sided space. We extend this measure to the two-sided space first on cylinders (invariantly), defining e.g. $\mu([011.]) \equiv \mu([.011])$; this extends to the full sigma-algebra $\mathcal{B}$ as before. □

From now on, a probability matrix will generally be denoted by P.

Given P together with a nonnegative row vector $\boldsymbol{\pi}^t$, we write $\mu(P, \boldsymbol{\pi})$ for the invariant measure just defined. The quadruple $(\Sigma_A^+, \mathcal{B}, \mu, \sigma)$ is termed a **Markov chain**. In the special case where $\boldsymbol{\pi}^t$ is invariant and hence so is the measure μ, this quadruple is known as a **Markov shift**, and as noted above, the measure then extends invariantly to the bilateral shift space Σ_A.

These are both special cases of the more general concept in probability theory of a **Markov process**. For this probabilistic interpretation the symbols of the alphabet of Σ^+ are thought of as the *states* of the system.

To explain this, given a measure space $(X, \mathcal{A}, \mu)$, we recall the definition of conditional measure: for $A \subseteq X$ with $0 < \mu(A) < \infty$, we define $\mu_A(E) = \mu(A \cap E)/\mu(A)$. This is a probability measure; one also writes $\mu(E|A)$ (the probability of E given A) for $\mu_A(E)$.

For $k \leq m$, we define $\mathcal{B}_k^m$ to be the $\sigma-$algebra generated by the cylinder sets $\mathbb{C}_k^m$. We write $\mathcal{B}_k^\infty$ for the $\sigma-$algebra generated by $\cup_{m \geq k} \mathbb{C}_k^m$ and similarly for $\mathcal{B}_{-\infty}^m$ and $\mathcal{B}_{-\infty}^\infty$, which equals $\mathcal{B}$.

We say μ on Σ is a **Markov measure** (or **satisfies the Markov property**) iff relative to any given time k, the future depends at most on the present; precisely, for $A \in \mathcal{B}_{-\infty}^{k-1}, B \in \mathcal{B}_k^k$ and $C \in \mathcal{B}_{k+1}^\infty$, then

$$\mu_{A \cap B}(C) = \mu_B(C).$$

We note that a Markov chain satisfies this property, as e.g. for $k = 0$, taking $A = [l.], B = [.i]$ and $C = [. * j]$ then $\mu_{A \cap B}(C) = \mu(A \cap B \cap C)/\mu(A \cap B) = \mu([l.jk])/\mu([l.j]) = \pi_l P_{lj} P_{jk}/\pi_l P_{lj} = P_{jk} = \pi_j P_{jk}/\pi_j = \mu_B(C)$, the same calculation working for other cylinder sets. Hence, the matrix entry P_{ij} gives the probability of making a transition from state i to state j, and for this reason, the matrix P is called the **transition matrix** of the Markov chain.

Remark 15.1. A Markov chain is a Markov measure with stationary (i.e. unchanging in time) transition probabilities. The concept of Markov measure is more general than this; one can consider stochastic processes for which the state space is infinite, or even continuous (such as a manifold or Lie group) and time can also be continuous; moreover the transition probabilities (now a **Markov operator** rather than a matrix) for such a **Markov process** may depend on time.

For example, assuming we have discrete time and discrete states, then a nonstationary sequence of $(d \times d)$ probability matrices $(P_k)_{k \geq 0}$, together with an initial nonnegative vector $\boldsymbol{\pi}^t$, determines a Markov measure on Π_0^∞ known as a **nonhomogeneous** Markov chain.

Returning to the current setting, we note that the definition of the Markov property is inherently time-asymmetric; however, this lack of symmetry is an illusion, as one has the equivalent expression:

$$\mu_B(A \cap C) = \mu_B(A)\mu_B(C); \tag{57}$$

This says that *the past and future are independent relative to the present*, and is valid for any "present" time k.

In this case where μ is invariant, one can write the corresponding transition matrix from future state j to present state i: assuming that $\boldsymbol{\pi} > 0$, this is it is $\widetilde{P}$ where

$$\widetilde{P}_{ij} = \mu([.ij])/\mu([. * j]) = (\pi_i/\pi_j)P_{ij};$$

here we have used the fact that by invariance, $\mu([. * j]) = \mu([.j])$. Writing Π for the diagonal matrix with entries $\Pi_{ii} = \pi_i$, in matrix form this is

$$\widetilde{P} = \Pi P \Pi^{-1}. \tag{58}$$

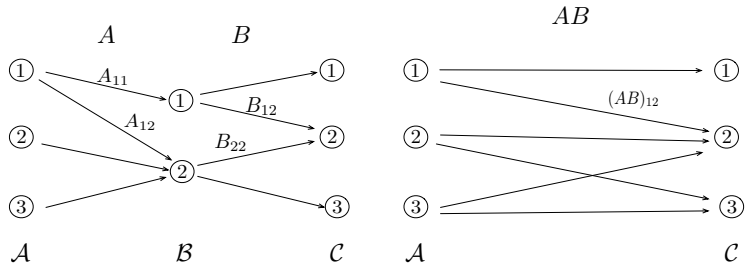


FIGURE 37. Matrix multiplication as a "sum over histories": $(AB)_{12} = A_{11}B_{12} + A_{12}B_{22}$

Note that the matrix $\widetilde{P}$ is **column-stochastic**, i.e. column sums are 1, since (equivalently) $\underline{1}^t\widetilde{P} = \underline{1}^t$. Note also that P and $\widetilde{P}$ are similar matrices; we return to this point below in equations (65) and (66).

The next proposition illustrates the probabilistic significance of the powers of the matrix P. The key to understanding this is an interesting representation of a matrix, illustrated in Fig. 37. Given a $(l \times m)$ matrix A, we define alphabets $\mathcal{A} = \{1, \ldots, l\}$ and $\mathcal{B} = \{1, \ldots m\}$ (considered as disjoint sets, with $\mathcal{A}$ followed by $\mathcal{B}$), and draw a graph whose vertices are the elements of $\mathcal{A} \cup \mathcal{B}$, and draw an edge from symbol $i \in \mathcal{A}$ to $j \in \mathcal{B}$ iff $A_{ij} \neq 0$. We label this edge with the corresponding entry. Now we interpret matrix multiplication as follows. Given a second, $(m \times n)$ matrix B, then we add the vertices for $\mathcal{C} = \{1, \ldots, d\}$, with the corresponding edges from $\mathcal{B}$ to $\mathcal{C}$ labelled by the entries of B. Now the ij^{th} entry of the product AB is the inner product of the i^{th} row of A with the j^{th} column of B; that is,

$$(AB)_{ij} = \sum_{k=1}^{m} A_{ik}B_{kj},$$

and in the graph this is exactly the sum over all edge paths connecting i to j (and passing through some k) of the products $A_{ik}B_{kj}$ of the entries along this path.

The same works for an arbitrary product $M_0M_1 \ldots M_n$ of matrices; the graph is a finite version of a **Bratteli diagram**, see §??.

Remark 15.2. This interpretation of a matrix product as the sum over all possible paths yields an easy proof of the associative law for matrix multiplication. Thus for matrices A, B, C which are $(m \times n), (n \times o)$ and $(o \times p)$ respectively, we have $(AB)_{ik} = \sum_{j=1}^{n} a_{ij}b_{jk}$ and $(BC)_{kl} = \sum_{k=1}^{0} b_{jk}c_{kl}$ whence

$$((AB)C)_{il} = \sum_{k=1}^{0}(\sum_{j=1}^{n} a_{ij}b_{jk})c_{kl} = \sum_{k=1}^{0}\sum_{j=1}^{n} a_{ij}b_{jk}c_{kl} = \sum_{j=1}^{n}\sum_{k=1}^{0} a_{ij}b_{jk}c_{kl} = \sum_{j=1}^{n} a_{ij}(\sum_{k=1}^{0} b_{jk}c_{kl})$$

where the middle sums are the sum over possible paths of length three. The point is that this path composition has erased all of the previous association information.

(This sum over paths is perhaps reminiscent of the "sums of histories" of Feynman diagrams in quantum mechanics!)

Proposition 15.3. *Given a (row) stochastic matrix P, probability row vector $\boldsymbol{\pi}^t$ and Markov measure μ on Σ^+, the matrix entry of the m^{th} power P^m_{ij} for $m \geq 1$ gives the transition probability from state i to state j after a gap of time m, and hence $\boldsymbol{\pi}^t_m = \boldsymbol{\pi}^t P^m$ gives the distribution of states at time m, for initial distribution $\boldsymbol{\pi}^t$.*

Proof. The probability of being in state j at time $(m+1)$ given that we are in state i at time m is $\mu([.**\cdots*ij])/\mu([.**\cdots*i])$, and from the definition of μ this is P_{ij}. For $m > 1$, the matrix product automatically sums over all the possible paths in the shift space (see Fig. 37), completing the proof. □

We are most interested in the case where μ is a $\sigma-$invariant probability measure. Recall from Definition 41.1 that (Σ, μ, σ) is mixing iff for all $A, B \in B$, $\mu(\sigma^{-m}A \cap B) \to \mu A \mu B$ as $m \to \infty$. Given a probability row vector $\boldsymbol{\pi}^t$, let us write $Q_{\boldsymbol{\pi}}$ for the $(d \times d)$ matrix each of whose rows is $\boldsymbol{\pi}^t$.

Proposition 15.4. *Given a Markov shift $(\Sigma, \mathcal{B}, \mu, \sigma)$ defined by a $(d \times d)$ probability matrix P and invariant probability row vector $\boldsymbol{\pi}^t$, then*
(i)

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} P^k = Q$$

exists; this satisfies $Q^2 = Q$, $PQ = QP = Q$;
(ii) the Markov shift is ergodic iff the limit is $Q = Q_{\boldsymbol{\pi}}$.
(iii) the Markov shift is mixing iff $\lim_{k\to\infty} P^k = Q_{\boldsymbol{\pi}}$.

Proof. Recalling the proof of Lemma 8.2, by the Birkhoff ergodic theorem, for any $A, B \in \mathcal{B}$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{0}^{N-1} \chi_B(T^n(x)) \to \overline{\chi}_B, \quad \text{almost-surely, whence}$$

$$\chi_A\left(\frac{1}{N} \sum_{0}^{N-1} \chi_B \circ T^n\right) \to \chi_A \overline{\chi}_B \quad \text{a.s., and so, integrating,}$$

$$\frac{1}{N} \sum_{k=0}^{N-1} \mu(A \cap T^{-k}B) \to \int_X \chi_A \overline{\chi}_B \mathrm{d}\mu \text{ as } N \to \infty.$$

Choosing now $a, b \in \mathcal{A}$, we set $A = [.a]$ and $B = [.b]$, and define a matrix Q by $Q_{ab} = \frac{1}{\pi_a} \int_X \chi_A \overline{\chi}_B \mathrm{d}\mu$. Let us note that for the special case when the map is ergodic, $\int_X \chi_A \overline{\chi}_B \mathrm{d}\mu = \mu A \mu B$ whence $Q_{ab} = \pi_b$ and hence each row is $\boldsymbol{\pi}^t$.

Now for $x_0 = a, x_k = b$ we have $\mu[x_0 \dots x_k] = \pi_a P_{ax_1} P_{x_1x_2} \cdots P_{x_{k-1}b}$.

Thus $\mu[a**\cdots*b] = \sum \pi_a P_{ax_1} P_{x_1x_2} \cdots P_{x_{k-1}b}$ where the sum is taken over all thin cylinders of length k beginning with a and ending with b. From Proposition 15.3, $\sum P_{ax_1} P_{x_1x_2} \cdots P_{x_{k-1}b} = P^k_{ab}$ while $[a**\cdots*b] = A \cap \sigma^{-k}B$ whence $\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} P^k_{ab} = \frac{1}{\pi_a} \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \mu[a**\cdots*b] = \frac{1}{\pi_a} \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \mu(A \cap T^{-k}B) = \frac{1}{\pi_a} \int_X \chi_A \overline{\chi}_B \mathrm{d}\mu = Q_{ab}$.

Hence the limit is Q with rows $\boldsymbol{\pi}^t$ in the ergodic case.

Conversely, we are given that for each a, b, $\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} P^k = Q_{\boldsymbol{\pi}}$. This is equivalent to

$$\frac{1}{N} \sum_{k=0}^{N-1} \mu(A \cap T^{-k} B) \to \mu A \mu B$$

for the special sets $A = [.a]$, $B = [.b]$. We shall prove this remains true for general thin cylinders; from there it will extend by additivity to general cylinders and so to arbitrary Borel sets.

Given then $A = [.x_0 x_1 \dots x_n = a]$ and $B = [.b = y_0 x_1 \dots y_m]$, we have that for any $k \geq 1$,

$$\begin{aligned} \mu(A \cap T^{-k-n} B) &= \mu[.x_0 \dots x_{n-1} a * * \cdots * * b y_1 \dots y_n] \\ &= \pi_a P_{a x_1} \cdots P_{x_{n-1} x_n} \cdot P^k_{ab} \cdot P_{b y_1} \cdots P_{y_{m-1} y_m} \end{aligned} \tag{59}$$

whence

$$\begin{aligned} &\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \mu(A \cap T^{-k-n} B) \\ &= \pi_a P_{a x_1} \cdots P_{x_{n-1} x_n} \cdot \big(\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} P^k_{ab} \big) \cdot P_{b y_1} \cdots P_{y_{m-1} y_m} \\ &= \pi_a P_{a x_1} \cdots P_{x_{n-1} x_n} \cdot (\pi_b) \cdot P_{b y_1} \cdots P_{y_{m-1} y_m} = \mu A \mu B. \end{aligned}$$

To prove (iii), if the shift is mixing, then for $A = [.a]$ and $B = [.b]$,

$$\mu(A \cap T^{-k} B) \to \mu A \mu B$$

so $P^k_{ab} = \frac{1}{\pi_a} \mu[a * * \cdots * b] = \frac{1}{\pi_a} \mu(A \cap T^{-k} B) \to \pi_b$ whence $P^k \to Q_{\boldsymbol{\pi}}$. Conversely, checking for thin cylinders, by (59)

$$\begin{aligned} &\lim_{N\to\infty} \mu(A \cap T^{-k-n} B) = \pi_a P_{a x_1} \cdots P_{x_{n-1} x_n} \cdot \big(\lim_{N\to\infty} P^k_{ab} \big) \cdot P_{b y_1} \cdots P_{y_{m-1} y_m} \\ &= \pi_a P_{a x_1} \cdots P_{x_{n-1} x_n} \cdot (\pi_b) \cdot P_{b y_1} \cdots P_{y_{m-1} y_m} = \mu A \mu B. \end{aligned}$$

□

Definition 15.1. A $(d \times d)$ matrix M with nonnegative real entries is M is **primitive** iff for some $m > 0$, M^m is strictly positive, i.e. has entries all nonzero.

The basic fact about primitive matrices is the Perron-Frobenius Theorem, proved below in §16: that there exist (up to normalization) unique nonnegative left, and right, eigenvectors.

Proposition 15.5. *The Markov shift is ergodic iff the transition matrix P is irreducible. It is mixing iff P is primitive.*

Proof. Since P is row-stochastic, by Lemma 15.1 it preserves Δ^t_∞. If P is primitive, then the unique left and right eigenvectors must, up to positive multiples, be $\boldsymbol{\pi}^t$ and $\underline{1}$. As in the proof of Theorem 16.1, the images $\Delta^t_k = \Delta^t \cdot P^k$ nest down to $\Delta^t_\infty = \{\boldsymbol{\pi}^t\}$. Thus for every $\mathbf{v} \in \Delta$, $\mathbf{v^t} P^m \to \boldsymbol{\pi}^t$. We choose e.g. for the (3×3) case, $\mathbf{v^t}$ to be the row vector $\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$, noting that $\begin{bmatrix} 1 & 0 & 0 \end{bmatrix} P^m$ gives the first row of P^m. Hence P^m converges to the matrix $Q_{\boldsymbol{\pi}}$. Then the previous Proposition implies mixing for the Markov shift.

Conversely, if the Markov shift is mixing then by (*iii*) of Proposition 15.4 $\lim_{k\to\infty} P^k = Q_{\boldsymbol{\pi}}$, whence it has a unique fixed point $\boldsymbol{\pi}^t$ in Δ^t_∞ hence a unique (up to multiples) nonnegative left eigenvector. By the Perron-Frobenius Theorem for the nonprimitive case §18.5, if the matrix P is not primitive then there exist other eigenvectors, contradicting this fact.

DO: Proof for ergodic case ! ?? □

Summarizing, we have:

Corollary 15.6. *These are equivalent, for a Markov chain on a finite state space with transition matrix P and invariant row vector $\boldsymbol{\pi}^t$, determining the measure μ on path space Π, with left shift map σ:*
(a) P is primitive;
(b) the transformation (Π, μ, σ) is mixing;
(d)P^n converges (to $\to Q_{\boldsymbol{\pi}}$).

And these are equivalent:
(e) P is irreducible;
(f) the transformation (Π, μ, σ) is ergodic;
(g)$\frac{1}{N}\sum_{k=0}^{N-1} P^k$ converges.

15.2. **Markov measures for subshifts of finite type.** Given a $(d \times d)$ matrix A with entries 0 and 1,p we define $\Sigma_A \subseteq \Sigma$ to be the set of all $x \in \Sigma$ such that $A_{x_i x_{i+1}} = 1$ for all $i \in \mathbb{Z}$. These are the **allowed strings**. We call Σ_A a (bilateral, or two-sided) **subshift of finite type** or *sft* for short. An alternative name is **topological Markov shift**. The *one-sided sft* is the corresponding subset $\Sigma^+_A \subseteq \Sigma^+$.

The matrix A defines a finite graph, whose vertices are the symbols and whose edges indicate the allowed transitions. That is, $A_{ij} = 1$ iff there is a (directed) edge from state (vertex) i to state j. See Fig. ???

Exercise 15.2. *If $d > 1$ and A is primitive, then Σ^+_A and Σ_A are homeomorphic to the Cantor set.*

Given an allowed string x and $k, m \in \mathbb{Z}$ with $k \le m$, we write $[x_k \dots x_m] = \{w \in \Sigma_A : w_k = x_k, \dots, w_m = x_m$; this is a **thin cylinder set**, and the collection of these is denoted $\mathbb{C}^m_k$. The decimal point again helps us keep track of the 0^{th} coordinate; thus $[01.0] \in \mathbb{C}^0_{-2}$. A **general cylinder set** is a finite union of thin cylinders; we write $*$ for "no restriction on the symbols" so e.g. for an alphabet with 3 symbols, some general cylinders which are unions of sets in $\mathbb{C}^4_0$ are $[**.*2*]$ or $[.*12*0]$.

The cylinder sets are clopen sets which generate the topology and hence the Borel $\sigma-$algebras $\mathcal{B}$ for Σ_A and $\mathcal{B}^+$ for Σ^+_A.

A matrix M which satisfies $(A_{ij} = 0) \implies (M_{ij} = 0)$ will be termed **compatible** with A. Given a compatible probability matrix P and a nonnegative vector $\boldsymbol{\pi}$, we define the Markov measure on Σ^+_A and Σ_A as above.

Given now a $0-1$ matrix A and a probability matrix P which is compatible with A, we decorate the graph of the *sft* Σ_A, labelling each edge with the transition probabilityP_{ij}. This is a **probability graph**, with the outgoing edges from a symbol indicating the chance of taking that path, while the unlabeled graph for the subshift

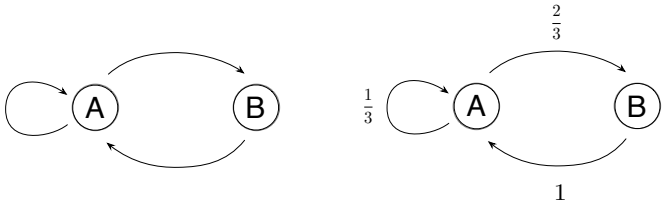


FIGURE 38. Possibility graph; probability graph.

of finite type is a graph of *possibilities*; see Fig. 38, where the matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ has been replaced by the compatible probability matrix $\begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ 1 & 0 \end{bmatrix}$.

15.3. **Markov partitions and the geometric Markov property.** In Fig. 4 we see how independence can be interpreted geometrically as product measure. From (57), the Markov property can be expressed as: *the past and future are independent relative to the present*, more precisely where A, B, C are elements of the past, present and future sigma-algebras respectively, then

$$\mu_B(A \cap C) = \mu_B(A)\mu_B(C) \tag{60}$$

Note that the independence in Fig. 4 could just as well be illustrated using parallelograms. Thus the relative independence of the Markov property is shown in the left-hand side of Fig. 39. In the middle and final picture the probabilities do not satisfy the Markov property; this is easiest to see in the final picture, where assuming A, B, and C are the 1- cylinders $[a.]$, $[.b]$ and $[.*c]$, $[a.b]$ and $[.bc]$ are allowed but $[a.bc]$ is *not* allowed. Hence the transitions are not given by a subshift of finite type.

In Definition 4.8 we defined a *topological Markov partition* to be a partition which codes a map as a subshift of finite type. So the first figure depicts a toplogical Markov partition, while the third is certainly not.

Now suppose that for a hyperbolic automorphism M of $\mathbb{T}$ we choose segments l_u, l_s of E^u, E^s respectively, such that:
(1) these segments include the origin;
(2) the endpoints of segment l_u belong to l_s and vice-versa.

We claim that if extended far enough, these segments form the boundaries of rectangles which satisfy the geometric Markov property.

The reason is remakable in its simplicity. Since $M^{-1}(l_u) \subseteq l_u$, a rectangle like A in Fig. 39 cannot occur, since its upper boundary is part of $M^{-1}(l_u)$ but *not* a subset of l_u, and similarly for the rectangle C. The interior of the rectangle B does not contain any of $l_u \cup l_s$. Thus the upper boundary of A which is part of $M^{-1}(l_u)$ and hence l_u must either be part of the boundary of B or outside of B. Therefore A must "pass completely through", giving independence of A and C relative to B. This same argument also guarantees the topological Markov property for the partition.

Thus properties $(1), (2)$ give another version of the geometric Markov property. We mention that while the Figure illustrates the case of a two-torus automorphism, something very similar works in all dimensions.

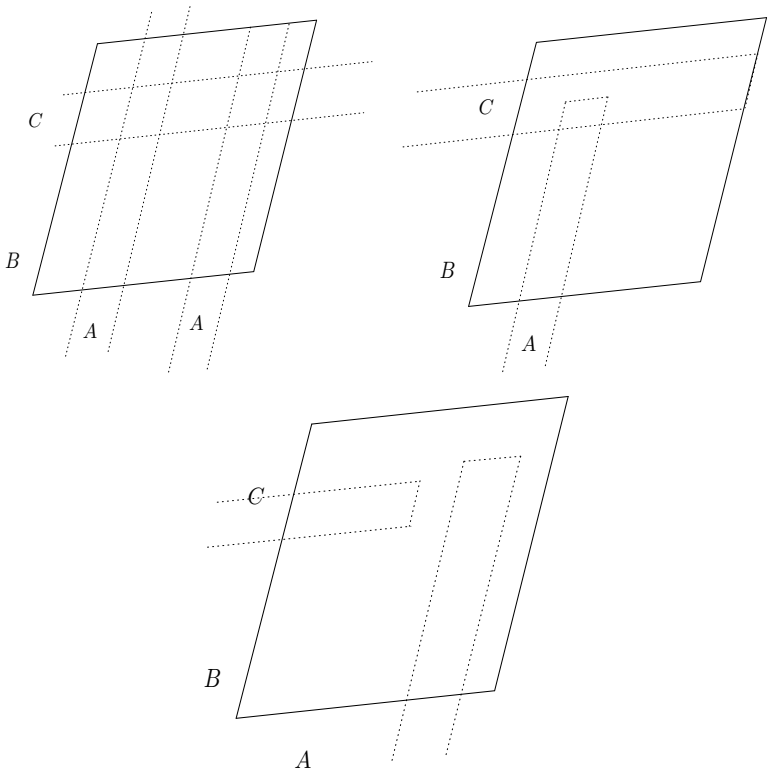


FIGURE 39. The Markov property: future and past partition elements A, C pass completely through present element B, giving relative independence; two examples of nonMarkov intersections.

This beautiful connection between the geometry and the probabilty of the Markov property is due to Adler and Weiss [AW70]; the extension of these ideas to higher dimensions were first made by Bowen and Sinai.

...Lebesgue and Parry measures

15.4. **Countable state Markov shifts: key examples.** Although in these notes finite state Markov shifts are of primary importance, the countable (or even uncountable) state case comes up naturally even in the study of these simpler objects.

We treat here some basics, returning later for a deeper study. Let X_i be an i.i.d. sequence of random variables taking values in $\mathbb{R}$ with probabilities given by a probability distribution ρ. We define S_n for $n \geq 0$ by $S_0 = 0$, $S_n = \sum_{k=0}^{n-1} X_k$; then S_n is a **random walk** with **independent increments** (the increments of a process S_n being $X_n \equiv S_{n+1} - S_n$), or an **i.i.d. random walk** for short.

The simplest example (not surprisingly known as the **simple random walk**) is where ρ is a distribution on $\mathbb{Z} \subseteq \mathbb{R}$ giving probabilities $1/2$ to 1 and to -1.

Note that we have already encountered i.i.d. partial sums S_n in the limit theorems of probability theory, §12.

From the ergodic theory point of view, we can begin with the Bernoulli shift $\Sigma^+ = \Pi_0^\infty\{0,1\}$ with the left shift map σ, and define $f(\underline{x}) = 1$ if $x_0 = 1$,$= -1$ if $x_0 = 0$; then $S_n = \sum_{k=0}^{n-1} f(\sigma^k(\underline{x}))$. Conversely, for any i.i.d. random walk S_n, the increment process $(X_0, X_1, \dots)$ is a point in the shift space $\Pi^+ = \Pi_0^\infty\mathbb{R}$ with independent product measure $\otimes_0^\infty\rho$.

Of course X_n is a Markov shift. But the partial sum process S_n is also a Markov chain: now the transition matrix P with index set $\mathbb{Z}$ and $P_{ij} = 1/2$ iff $|i-j| = 1$, 0 otherwise, and the initial probability vector is $\boldsymbol{\pi}$ with $\pi_i = 1$ for $i = 0$.

Note that P cannot be easily drawn in traditional matrix form, as the index set is biinfinite; however it fits a more general definition of matrix perfectly well; see below ??.

The process S_n has the Markov probability measure denoted μ defined from the stochastic matrix P and initial probability vector $\boldsymbol{\pi}$ as in (56). Of course this is not shift invariant, as $\boldsymbol{\pi}$ is not an invariant vector. Indeed, setting $\boldsymbol{\pi}_0^t = \boldsymbol{\pi}^t$, $\boldsymbol{\pi}_n^t = \boldsymbol{\pi}^t P^n$, then $\boldsymbol{\pi}_n^t$ gives the distribution of the random walk S_n at time n.

This has a **binomial distribution**, which can be understood by drawing Pascal's triangle, which counts the number of paths from the initial point to another vertex, and then assigning transition probabilities $1/2$ to each edge, see Fig. 25, giving at time n the probability distribution $\boldsymbol{\pi}_n$. Indeed, the path space for the random walk (connected by polygonal interpolation, to give the **polygonal random walk**) can be visualized as a Pascal's triangle turned on its side, Fig. 26.

One might search for an initial distribution which is invariant, so as to come up with a *stationary* Markov process for the random walk.

There is a natural choice: let ρ be counting measure on $\mathbb{Z}$, so now $\boldsymbol{\pi}$ satisfies $\pi_i = 1$ for all $i \in \mathbb{Z}$. Note that indeed $\boldsymbol{\pi^t} P = \boldsymbol{\pi}^t$; however the resulting shift-invariant measure ν is now infinite.

Writing as above μ for the random walk starting at 0 at time 0, then the relationship between the two is that μ is the conditional measure for ν, relative to the set $A = \{(S_n) : S_0 = 0\}$. That is, $\mu = \nu_A$. In probability language, we say that this is the *random walk conditioned to start at* 0.

There is a natural way to approximate this by finite-state Markov shifts: let $\mathcal{A}_d = \mathbb{Z}_d = \mathbb{Z}/d\mathbb{Z}$, the integers mod d; this is our state space, with $(d \times d)$ transition matrix $P_{ij} = 1/2$ iff $|i-j| = 1 (\text{mod } d)$, 0 otherwise. Then taking $\boldsymbol{\pi_d^t}$ to be $[1, 1, \dots 1]$, this is invariant, and the resulting measures ν_d on paths in $\Pi_\mathbb{Z}$ converge to ν. ?? (edge effect) (reflected better?)

example: renewal shift; recoding doubling map; induced of renewal is ctable state Bernoulli; fte and inf measure; prove basic thms- invariance; bilateral iff unilateral; extension; all Markov shifts are r walks; fte state approx...existence of invariant vector w/o PF ??? recoding doubling map.

16. The Perron-Frobenius Theorem

We present a simple and entirely geometric proof of the Perron-Frobenius Theorem. The proof may be original in this form, although it borrows ideas from several sources; in particular, parts of the proof are like that in [KH95], and another part can be seen as a finite-dimensional version of an argument of Walters [Wal75]. Our

inspiration originally came from a construction of a certain measure, in the ergodic and information theory of subshifts of finite type, given by Shannon and later independently by Parry; further references will be given and this history will be explained in §2.

We thank Jarek Kwapisz and Pierre Arnoux for helpful suggestions regarding the proof.

There are many proofs in the literature; one of the nicest is without doubt the projective metric proof of Birkhoff and Samelson ([Bir57], [Bir67], [Sam56]), which is also geometrical, see references and comments below. Our proof has the advantage that it can be presented more readily in an undergraduate or graduate course, as less background preparation is necessary.

16.1. **Proof of the theorem.** To understand the action of a linear transformation on a vector space, it is clearly a good idea to try to identify any invariant subspaces, as a way of simplifying the description of the map. Since the zero-dimensional subspace $\{\mathbf{0}\}$ is always invariant, we should start by considering subspaces with dimension 1.

Definition 16.1. Given a complex $(d\times d)$ matrix M, a vector $\mathbf{v} \in \mathbb{C}^d$ is an *eigenvector* for M iff $\mathbf{v} \neq \mathbf{0}$ and there exists $\lambda \in \mathbb{C}$ (possibly zero) such that $M\mathbf{v} = \lambda\mathbf{v}$. (Note that $\mathbf{v} = \mathbf{0}$ should not be allowed here as it gives the 0-dimensional space mentioned above, and furthermore since then *any* λ would work!) For a real matrix M, we consider it as a complex matrix acting on $\mathbb{C}^d$, so eigenvectors and eigenvalues are allowed to be complex.

See Lemma 35.60 for the geometric meaning of complex eigenvalues and eigenvectors for real matices.

The theorem shows that any primitive nonnegative $(d \times d)$ matrix defines what is called a partially hyperbolic map of $\mathbb{R}^d$. Here is the definition:

Definition 16.2. Let M be a differentiable manifold. A diffeomorphism f of M is *partially hyperbolic* iff there exists an invariant splitting of the tangent bundle $TM = E^s \oplus E^u$ such that there exist $\mu < \lambda \in \mathbb{R}$ and $c > 0$ such that for all $n \geq 1$,

$$||Df^n(\mathbf{v})|| \leq \mu^n ||\mathbf{v}||$$

for all $\mathbf{v} \in E^s(p)$, for all $p \in M$, and

$$||Df^n(\mathbf{v})|| \geq \lambda^n ||\mathbf{v}||$$

for all $\mathbf{v} \in E^u(p)$. It is *hyperbolic* if we can take $\mu < 1 < \lambda$ here. We recall that f is an *Anosov* diffeomorphism iff it is hyperbolic (i.e. on all of M) for M a compact manifold. The idea behind the study of partial huperbolicity is to see which properties originally proved for Anosov diffeomorphisms may go through in the presence of weaker conditions. Note in particular that for partial hyperbolicity we may well have $\mu < \lambda < 1$ or $! < \mu < \lambda < 1$ or a triple splitting $TM = E^s \oplus E^c \oplus E^u$ where $\mu < 1 < \lambda$ and there is no asymptotic expansion or contraction at all in E^c; this is an important special case where E^c is called the *central* direction.

We note that Birkhoff's proof (see Theorem 24.5) offers explicit extimates on both $|\mu|$ and λ. This becomes especially critical when studying a *sequence* of maps, i.e. *non-stationary* dynamics.

Of course for $M = \mathbb{R}^d$, just multiplying f by a constant c shifts the spectrum $\{\mu, \lambda\}$ to $\{c\mu, c\lambda\}$. The real content of (partial) hyperbolicity is when there is a finite invariant measure, forcing recurrence, or when M is a compact manifold. There are many fascinating examples to go with the theory, see e.g. .Rodriguez-Hertz, Hasselblatt.....

Theorem 16.1. *Let M be a $(d \times d)$ matrix with nonnegative real entries which is primitive.*
(i) Then there exists (up to multiplication by a constant) a unique strictly positive right eigenvector; there is a unique strictly positive left eigenvector as well. The eigenvalues are equal and positive.
(ii) Any other (possibly complex) eigenvalue has modulus strictly less than this eigenvalue λ.

Proof. We note that since M is nonnegative, for any positive eigenvector the corresponding eigenvalue is necessarily positive.

First we go through the proof of (i), then fill in the details in several lemmas. Part (ii) will follow from (i), by an argument which we borrow from [KH95].

To prove (i), we begin by showing that M has at least one positive right eigenvector.

The **positive cone** is $(\mathbb{R}^d)^+ = \{\mathbf{v} : \ v_i \geq 0\}$. We write Δ for the unit simplex in $\mathbb{R}^d$, i.e. Δ is the convex set spanned by the standard basis column vectors. We set $||\mathbf{w}|| = \sum_{i=1}^d |w_i|$; note that with this choice of norm the map $\mathbf{w} \mapsto \mathbf{w}/||\mathbf{w}||$ projects the positive cone, minus its vertex $\mathbf{0}$, onto Δ. We define $f_M : \Delta \to \Delta$ by

$$f_M(\mathbf{v}) = \frac{M\mathbf{v}}{||M\mathbf{v}||}.$$

Note that f_M has a fixed point in Δ if and only if M has an eigenvector in the positive cone.

Thus, we wish to show that f_M has at least one fixed point. One approach is to apply the Brower fixed point theorem, since f_M is a continuous map of a topological $(n-1)$-ball. This would provide a nonnegative eigenvector. We shall instead give an elementary argument, as we will need the same method later on. Writing $\Delta_0 = \Delta$, $\Delta_1 = f_M(\Delta)$, $\ldots, \Delta_k = (f_M)^k(\Delta)$, and $\Delta_\infty = \cap_{n=0}^\infty \Delta_k$, we have (Lemma 16.3) that Δ_∞ is compact and convex. Since some power of M is positive, Δ_∞ is contained in the interior of Δ (so if there is an eigenvector in the positive cone, in fact it is strictly positive). We show Δ_∞ has at most d extreme points (Lemma 16.3). The map f_M permutes this finite set, hence some power m fixes all of these points. So (equivalently) there exists a positive right eigenvector for M^m. This part of the proof is like that in [KH95].

Next for M itself we show existence, and at the same time uniqueness, first under a special additional assumption: that M is *column-stochastic*, i.e. it preserves Δ in its action on column vectors.

This implies that the map f_M on the simplex Δ_∞ is the restriction of M and so is affine. From the previous step we know that Δ_∞ has at most d extreme points; we claim that in fact it has a single point.

If there are ≥ 2 extreme points for Δ_∞, consider the line segment in Δ containing these points. Since f_M^m fixes these two points, and is affine, it fixes the entire segment; hence it fixes the point where the line extending the segment encounters the boundary

of Δ. But any fixed point in Δ for f_M^m is in Δ_∞ which we know is strictly inside of Δ, giving a contradiction.

Next we show how to reduce to the column-stochastic case. We begin with the general matrix M; we know from above that for some power m, there exists a column vector $\mathbf{w}$ and $\lambda > 0$ such that

$$M^m \mathbf{w} = \lambda \mathbf{w}.$$

Therefore, in its action on row vectors $\mathbf{v}^t$ ($*^t$ indicates transpose), the matrix $(1/\lambda)M^m$ preserves the hyperplane

$$H_{\mathbf{w}}^t \equiv \{\mathbf{v}^t : \ \mathbf{v} \cdot \mathbf{w} = 1\}$$

(Lemmas 16.5, 16.6). A change of basis produces a positive matrix P whose action on row vectors maps the simplex Δ^t into itself (see §16.3). Now P is row-stochastic, and some power is positive, so applying the above proof (on rows rather than columns), P has a unique positive left eigenvector $\boldsymbol{\pi}$.

The conjugacy of M with P shows that M^m also has a unique positive left eigenvector. Since the left and right positive eigenvectors for P both have the same eigenvalue ($= 1$), this fact passes by similarity over to M. A simple argument then implies existence and uniqueness for M itself (Lemma 16.4). By duality (i.e. exchanging the role of left and right), M has a unique positive right eigenvector. This finishes the proof of part (i). □

Remark 16.1. We mention that the fact that the maximum eigenvalues λ for M and M^t are equal is general: from Lemma 35.52 below, for any real rectangular matrix A, $||A|| = ||A^t||$ where this is the operator norm.

16.2. **Some details.**

Lemma 16.2. *For M as in the Theorem, the image by f_M of any convex set is convex, and moreover for any finite set of points, the image of their convex hull is the convex hull of the image of the points.*

Proof. For two points, the statement is that the image of the segment $[\mathbf{v}, \mathbf{w}]$ with endpoints $\mathbf{v}, \mathbf{w}$ is the segment (possibly a point) $[f_M(\mathbf{v}), f_M(\mathbf{w})]$. Indeed, since M is linear, the image of a line segment in the positive cone $(\mathbb{R}^d)^+ = \{\mathbf{v} : \ v_i \geq 0\}$ is a line segment, and when normalized to Δ this gives either a line segment or a point, with those extreme points. It follows from this statement that the image by f_M of a convex set is convex.

Now consider the case of three points $\mathbf{v}, \mathbf{w}, \mathbf{z} \in \Delta$; given a point $\mathbf{x} = a\mathbf{v} + b\mathbf{w} + c\mathbf{z}$ where $a + b + c = 1$ and $a, b, c \geq 0$, there is a point $\tilde{\mathbf{x}}$ on the segment $[\mathbf{v}, \mathbf{w}]$ such that $\mathbf{x}$ lies on the segment $[\tilde{\mathbf{x}}, \mathbf{z}]$. Indeed, take $\tilde{\mathbf{x}} = a/(a+b)\mathbf{v} + b/(a+b)\mathbf{w}$. Since by the above each of these segments is mapped by f_M to a segment with the corresponding extreme points, the result follows; the general induction step is similar. □

Lemma 16.3. *For M as in the Theorem, $\Delta_k = f_M^k(\Delta)$ is convex, compact and nonempty and the number of its extreme points is at most d. The same holds for Δ_∞. Also, $f_M(\Delta_\infty) = \Delta_\infty$. The map f_M acts on the set of extreme points as a permutation.*

Proof. Since f_M is continuous, each Δ_k is a compact set. From the previous lemma, we have that Δ_k is a convex set, and that the extreme points $\mathrm{Ext}(\Delta_k)$ satisfy $f_M(\mathrm{Ext}(\Delta_k)) \supseteq \mathrm{Ext}(\Delta_{k+1})$. Thus $\#\mathrm{Ext}(\Delta_k) \leq d$ for all k. Since Δ_k is compact, convex and nonempty, so is the intersection $\Delta_\infty = \cap_{n=0}^\infty \Delta_k$.

We next show that $f_M(\Delta_\infty) = \Delta_\infty$. We make the following more general
CLAIM: Let $f : X \to X$ be a continuous map on a topological space X, and let K be a compact subset of X. If $f(K) \subseteq K$, then for $K_n = f^n(K)$ and $K_\infty = \cap_{n=0}^\infty K_n$, we have that $K = K_0 \supseteq K_1 \supseteq K_2 \ldots$ and that $f(K_\infty) = K_\infty$.

But this is exactly Lemma 4.3 above!

Now we return to Δ_∞. We need to show $\mathrm{Ext}(\Delta_\infty)$ is finite. For $x \in \Delta_\infty$, for each l, there are real numbers $\lambda_i^{(l)}$ and an integer $j = j(l)$ such that

$$x = \sum_{i=1}^{j} \lambda_i^{(l)} \mathbf{e_i}^{(l)}$$

where $\{\mathbf{e_1}^{(l)}, \ldots, \mathbf{e_j}^{(l)}\} = \mathrm{Ext}(\Delta_l)$. Let us write $m = \min_{l \geq 0}\{j(l) = \#\mathrm{Ext}(\Delta_l)\}$. Thus there exists J such that for every $l \geq J$, $\#\mathrm{Ext}(\Delta_l) = m$. We claim that $\#\mathrm{Ext}(\Delta_\infty) \leq m$.

Now by compactness of Δ_l and Δ_∞, there exists for each i a subsequence of $(\mathbf{e_i}^{(l)})_{l=J}^\infty$ which converges to some point $\mathbf{e_i}^{(\infty)} \in \Delta_\infty$. We claim that $\mathrm{Ext}(\Delta_\infty) \subseteq \{\mathbf{e_i}^{(\infty)}\}_{i=1}^m$. (Here the order on each set $\mathrm{Ext}(\Delta_l)$ is of no importance, once it has been fixed.) Indeed, any point $x \in \Delta_\infty$ can be written as a convex combination

$$x = \sum_{i=1}^{m} \lambda_i^{(l)} \mathbf{e_i}^{(l)}$$

for each $l \geq J$; by compactness of $[0, 1]$, for each i there exists a subsequence of $\lambda_i^{(l)}$ converging to λ_i^∞ such that we have $x = \sum_{i=1}^m \lambda_i^\infty \mathbf{e_i}^{(\infty)}$; hence $\{\mathbf{e_i}^{(\infty)}\}_{i=1}^m \supseteq \mathrm{Ext}(\Delta_\infty)$, as stated.

Finally we show f_M permutes $\mathrm{Ext}(\Delta_\infty)$. First we show that given $\mathbf{b} \in \mathrm{Ext}(\Delta_\infty)$, there exists some $\mathbf{a} \in \mathrm{Ext}(\Delta_\infty)$ which maps onto it. Indeed, since the map is onto, there exists some preimage $\mathbf{c} \in \Delta_\infty$; if $\mathbf{c}$ is not extreme, it is a nontrivial convex combination of the extreme points; but by the previous lemma, its image is a (generally different, since f_M may not be linear)) convex combination of the images of these points; yet $\mathbf{b}$ is extreme; so this combination must be trivial, and at least one of these extreme points must map onto $\mathbf{b}$.

Then, since $\mathrm{Ext}(\Delta_\infty)$ is finite, it follows that f_M acts as a permutation. □

Lemma 16.4. *Let $f : X \to X$ be a function on a set such that for some $m > 1$, f^m has a unique fixed point. Then the same is true for f.*

Proof. Let x be the unique fixed point for f^m. Then x is a periodic point for f, of (least) period k which divides m. We want to show that $k = 1$. But this is true, since the orbit of x provides k distinct fixed points for f^m. Lastly, uniqueness for f follows from uniqueness for f^m. □

As noted in the proof of the theorem, the previous lemma lets us work in what follows with M rather than M^m.

Lemma 16.5. *Let M be a $(d \times d)$ real matrix. Then $M\mathbf{w} = \lambda\mathbf{w} \iff (1/\lambda)M$ maps the hyperplane*

$$H^t_{\mathbf{w}} = \{\mathbf{v}^t : \ \mathbf{v} \cdot \mathbf{w} = 1\}$$

into itself (by multiplication on the right of row vectors).

Proof. This can be deduced from the fact that taking the orthogonal complement is idempotent in finite-dimensional vector spaces, but we include a simple proof for completeness. By associativity of matrix multiplication, $(\mathbf{v}^t(1/\lambda)M)\mathbf{w} = \mathbf{v}^t(1/\lambda)(M\mathbf{w}) = 1$. This shows ($\Longrightarrow$). The hypothesis of the converse states:

$$\text{For fixed } \mathbf{w}, \text{ defining } \mathbf{z} = (1/\lambda)M\mathbf{w}, \text{ then } \mathbf{v} \cdot \mathbf{w} = 1 \implies \mathbf{v} \cdot \mathbf{z} = 1. \tag{61}$$

We shall show:
If (61) holds then $\mathbf{w} = \mathbf{z}$.

Without loss of generality, assume $||\mathbf{w}|| = 1$. Now by (61), since $\mathbf{w} \cdot \mathbf{w} = 1$ then $\mathbf{w} \cdot \mathbf{z} = 1$. Since $\mathbf{z} \cdot \mathbf{w} = 1$, then using (61) again, $\mathbf{z} \cdot \mathbf{z} = 1$ so also $||\mathbf{z}|| = 1$. Now we can write $\mathbf{z} = (\mathbf{z} \cdot \mathbf{w})\mathbf{w} + \widehat{\mathbf{w}}$ where $\mathbf{w} \cdot \widehat{\mathbf{w}} = 0$ (since, defining $\widehat{\mathbf{w}}$ by the first equation, the second follows). We claim $\widehat{\mathbf{w}} = 0$, which will finish the proof. But $\mathbf{w} \cdot \widehat{\mathbf{w}} = 0$ implies that $(\mathbf{w} + \widehat{\mathbf{w}}) \cdot \mathbf{w} = 1$, so by (61) $(\mathbf{w} + \widehat{\mathbf{w}}) \cdot \mathbf{z} = 1$, but this is $(\mathbf{w} + \widehat{\mathbf{w}}) \cdot \big((\mathbf{z} \cdot \mathbf{w})\mathbf{w} + \widehat{\mathbf{w}}\big) = \mathbf{z} \cdot \mathbf{w} + \widehat{\mathbf{w}} \cdot \widehat{\mathbf{w}}$, so $\widehat{\mathbf{w}} \cdot \widehat{\mathbf{w}} = 0$ as desired. □

Write $\underline{1}$ for the column vector all of whose entries are 1, and $\underline{1}^t$ for the corresponding row vector. Making use of that notation, we have this useful linear algebra characterization of being row- and column- stochastic:

Lemma 16.6. *For a nonnegative $(d \times d)$ matrix M, M is row-stochastic $\iff$ $M\underline{1} = \underline{1} \iff \Delta^t M \subseteq \Delta^t$; M is column-stochastic $\iff \underline{1}^t M = \underline{1}^t \iff M\Delta \subseteq \Delta$.*

Proof. This is immediate from the definition and Lemma 16.5. □

16.3. **The change of basis.** For of notation let us now assume we are in the (3×3) case. Given M and $\mathbf{w}$ such that $M\mathbf{w} = \lambda\mathbf{w}$, with

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix},$$

then defining

$$W = \begin{bmatrix} w_1 & 0 & 0 \\ 0 & w_2 & 0 \\ 0 & 0 & w_3 \end{bmatrix}$$

we set

$$P = \frac{1}{\lambda} W^{-1} M W.$$

Note that $P\underline{1} = \underline{1}$, so the matrix W has given a change of basis which transforms $(1/\lambda)M$ to P which is row-stochastic; this completes the proof as explained above.

Remark 16.2. The change-of-basis from $(1/\lambda)M$ to P can be visualized as follows: the standard basis for P of row vectors, which spans the simplex Δ^t, is taken to the row vectors $\begin{bmatrix} w_1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & w_2 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & w_3 \end{bmatrix}$, which span a hyperplane which is orthogonal to $\mathbf{w}$, and which is preserved by the matrix $(1/\lambda)M$. Figure **??** shows the row-action of M on the simplex formed by the intersection of this hyperplane with the positive cone.

16.4. **The maximum eigenvalue.** Now we show that the eigenvalue λ is maximal in modulus.

Now we give the proof of (ii) of Theorem 16.1, for which we follow [KH95].

Lemma 16.7. *Let M be as in the Theorem, with λ the eigenvalue for the positive eigenvector $\mathbf{v}$. Suppose μ is another eigenvalue. Then $|\mu| < \lambda$.*

Proof. First suppose that μ is real. Then there is a column vector $\mathbf{w} \in \mathbb{R}^d$ such that $M\mathbf{w} = \mu\mathbf{w}$. Now consider the plane spanned by $\mathbf{v}$ and $\mathbf{w}$. If $\mu = \lambda$, then the action of M simply dilates this plane by the constant λ, so the ray where $(\mathbb{R}^d)^+$ meets this plane is taken to itself. However we know the image of Δ lies inside $\text{int}\Delta$, giving a contradiction. If $\mu > \lambda$, then that ray is mapped outside of the cone. If $-\mu \geq \lambda$, the same argument works for the map M^2.

Next suppose μ is complex. By Lemma 35.60, there is a plane where the matrix acts with respect to a (possibly non-orthogonal, as noted above) basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ by $\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \rho \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. We consider the action of M on the $3-$dimensional space generated by $\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}$. If θ is rational, then some power (say k) of this matrix dilates the $(\mathbf{u}_1, \mathbf{u}_2)$-plane by ρ^k. Considering say the $(\mathbf{u}_1, \mathbf{v})$ plane, we are now in the situation considered above for a real eigenvalue. If θ is irrational, some power of the rotation returns us arbitrarily close to the identity, and the same argument works again. □

16.5. **Acknowledgements and history.**

The proof we give here of the Perron-Frobenius Theorem arose in the following way: we realized that the definition of Shannon-Parry measure as given in [AW70] could be rewritten in matrix form, and that this amounted to a change-of-basis; in trying to produce a simple proof of the Perron-Frobenius Theorem for use in a class on dynamical systems given at Stony Brook in Fall 1993, we gave first the proof for the stochastic case, and noticed that the same equation could reduce the general case to this one. Later we realized that this approach is essentially a translation of Walters' proof of the Ruelle-Perron-Frobenius Theorem [Wal75] to the Markov case. There an abstract fixed point theorem is used to give the first step; here because of the finite dimensions, we are able to give an elementary argument. It turns out that the first part of this was anticipated in the book by Katok and Hasselblatt [KH95].

We wish to thank Jarek Kwapicsz, then (in 1993) a student in the dynamics class at Stony Brook, who made a nice observation which simplified the part of the proof about affine maps, and Pierre Arnoux, who read and commented on the whole section and contributed the proof of Lemma 16.5.

16.6. **Parry measure.** We return to considering a $0-1$ $(d \times d)$ matrix A and its subshift of finite type Σ_A^+.

If we draw the graph for the matrix A, then each edge corresponds to an entry 1 of the matrix, that is, $A_{ij} = 1$ iff there is a (directed) edge from state (vertex) i to state j. Given a row-stochastic matrix P which is compatible with A, we can now label each edge by the transition probabilityP_{ij}. One can think of the graph with these labels as the probability graph, while the unlabeled graph for the subshift of finite type is a graph of *possibilities*. There are in general many possible probability matrices to choose from!

We denote by $\mathcal{M}_A^{\text{Mark}}$ the set of all the shift-invariant measures $(P, \boldsymbol{\pi}^t)$. These are the invariant (one-step) Markov measures on Σ_A^+. We also write $\mathcal{M}_A$ for the set of all invariant probability measures on Σ_A^+.

Out of all the possible choices of such measures, there is one probability matrix P and hence one Markov measure in $\mathcal{M}_A^{\text{Mark}}$ of special importance, for two reasons mentioned above: first it turns out to be the measure of maximal entropy (for which the measure theoretic entropy equals the topological entropy), and second, geometrically it corresponds to Lebesgue measure on the torus. The formula for this measure was discovered by Shannon [SW63] and later independently by Parry [Par64] (we follow convention in naming this after Parry); it can be defined as follows.

Suppose our $0-1$ matrix A is primitive. Now from the Perron-Frobenius Theorem, we know such a matrix has, up to constant multiples, a unique strictly positive right eigenvector $\mathbf{w}$ and unique positive left eigenvector $\mathbf{v^t}$, both with eigenvalue $\lambda > 0$. We normalize these so that $\sum v_i w_i = 1$.

We define the diagonal matrix W from $\mathbf{w}$ as before. We set $P = \frac{1}{\lambda} W^{-1} A W$. We know that P has a unique left eigenvector $\boldsymbol{\pi}^t$, normalized to be in Δ^t; one checks that in fact $\pi_i = w_i v_i$. This is a left eigenvector with eigenvalue 1, so is invariant. Hence it defines an invariant measure, $\mu = \mu(P, \boldsymbol{\pi}^t)$; this is **Shannon-Parry measure**.

Theorem 16.8. (equidistribution property) *The measure is equidistributed on cylinders in the sense that for all sets $[x_0 \dots x_m] \in \mathbb{C}_0^m$, setting $c_1 = \min\{v_i w_j\}$ and $c_2 = \max\{v_i w_j\}$, then for all $m \geq 0$,*

$$c_1 \lambda^{-m} \leq \mu([x_0 \dots x_m]) \leq c_2 \lambda^{-m}. \tag{62}$$

Proof. From $P = \frac{1}{\lambda} W^{-1} A W$ we have that (equivalently)

$$P_{ij} = \frac{1}{\lambda} \frac{w_j}{w_i} A_{ij}.$$

Hence for an allowed string x,

$$\mu([x_0 \dots x_m]) = \pi_{x_0} P_{x_0 x_1} \cdots P_{x_{m-1} x_m} = \lambda^{-m} (v_{x_0} w_{x_0}) \frac{w_{x_1}}{w_{x_0}} \dots \frac{w_{x_m}}{w_{x_{m-1}}} = \lambda^{-m} v_{x_0} w_{x_m} \tag{63}$$

as everything else cancels. □

As an application we would like to count the number of cylinder sets of length m, $\#\mathbb{C}_0^m$. We give three methods, the first based on the equidistribution of Parry measure, the second which relies on the graph interpretation of matrix multiplication

given above in Fig. 37, and the third making use of the opertor norm. First we note that in the special case of matrices with entries $0, 1$ we have the following:

Proposition 16.9. *Given a $(d \times d)$ nonnegative integer matrix A, the number of cylinder sets $\#\mathbb{C}_0^m$ is the number of paths in the finite (stationary) Bratteli diagram from time 0 to time m, and equals $\sum_{i,j} A_{ij}^m = ||A^m|| = ||A^m 1||$ where the first norm is the L_1-norm of the matrix and the second of the vector.*

Remark 16.3. A basic fact is that in a finite dimensional vector space all norms are equivalent (see Lemma 35.41) In the statement above we used the L_1 norm $||M|| = \sum_{i,j} |M_{ij}|$. But for linear operators there is a second norm: $||M||_{op} = \max_{||\mathbf{v}||=1} ||M\mathbf{v}||$ which is very useful because of the following submultiplicative property:

$$||AB||_{op} \leq ||A||_{op} ||B||_{op}$$

(Exercise: prove this, and interpret it geometrically).

The next statement tells us that the topological entropy of the subshift of finite type Σ_A for A primitive with entries $0, 1$ is $h = \log \lambda$, see Def. 4.7.

Corollary 16.10. *The number of cylinder sets grows exponentially with exponent h; that is there exist $C_1, C_2 > 0$ such that for all $m \geq 0$,*

$$C_1 \ e^{hm} \leq \#\mathbb{C}_0^m \leq C_2 \ e^{hm}.$$

Proof. For the first proof, by the equidistribution property (62), setting $C_1 = 1/c_2$ and $C_2 = 1/c_1$. For the second, note that from Proposition 16.9, $\#\mathbb{C}_0^m = \sum_{i,j} A_{ij}^m = ||A^m 1||$. Now by the Perron-Frobenius Theorem the growth rate of any nonnegative vector under the action of A is given by λ^m. Thus

$$C_1 \lambda^m \leq \#\mathbb{C}_0^m \leq C_2 \lambda^m$$

as claimed.

Lastly, from the Remark, there are $\widetilde{c}_1, \widetilde{c}_2 > 0$ such that for all n, $\widetilde{c}_1 ||A^n||_{op} \leq ||A^n|| \leq \widetilde{c}_2 ||A^n||_{op}$ but since $A\mathbf{w} = \lambda \mathbf{w}$ with λ the largest eigenvalue, necessarily $||A^n||_{op} = \lambda^n$, giving the third proof.

□

As a consequence we have that

$$\lim \frac{1}{m} \log \#\mathbb{C}_0^m = \log \lambda = h$$

which corresponds to the weaker statement: for all $\epsilon > 0$ then for m sufficiently large, we have

$$\lambda^{m(1-\varepsilon)} \leq \#\mathbb{C}_0^m \leq \lambda^{m(1+\varepsilon)}. \tag{64}$$

Remark 16.4. Equation 64....Shannon-McMillan-Breiman Theorem (sometimes known as the *Ergodic Theorem of Information)*.....

Remark 16.5. We have seen above in (57) and (58) how to write the time-reversed matrix $\widetilde{P}$ for a general Markov shift; it is $\widetilde{P} = \Pi P \Pi^{-1}$; for the Shannon–Parry measure, this is

$$\widetilde{P} = \Pi P \Pi^{-1} = VWPW^{-1}V^{-1} = VAV^{-1} \tag{65}$$

which makes sense, as we have

$$\mu([x_0 \dots x_m]) = \pi_{x_0} P_{x_0 x_1} \cdots P_{x_{m-1} x_m} = \lambda^{-m} v_{x_0} w_{x_m} = \widetilde{P}_{x_0 x_1} \cdots \widetilde{P}_{x_{m-1} x_m} \pi_{x_n}. \tag{66}$$

This new formula will be useful below.

Example 19. Let's calcuate the Parry measure for the golden shift; this will lead us to the natural invariant measure for the Markov interval map of Fig. 17.....

17. Entropy

-variational principle (Misz proof)

18. Measure theory of adic transformations

18.1. **Primitive case: Unique ergodicity for stationary adic transformations.**

18.2. **The lemma of Bowen and Marcus.** As above, we have a $(d \times d)$ 0-1 matrix A, and assume A is primitive. Thus we have the Perron-Frobenius right and left eigenvectors $\mathbf{w}, \mathbf{v}^t$ with eigenvalue λ. First we give a quite different characterization of Shannon-Parry measure. The significance of this will become clear below.

On the space Σ_A^+, we define a second measure ν, by

$$\nu([x_0 \dots x_m]) = \mu([x_0 \dots x_m])/v_{x_0} = w_{x_0} P_{x_0 x_1} \cdots P_{x_{m-1} x_m} = \lambda^{-m} w_{x_m}. \tag{67}$$

Here we have used (63). This is a Markov measure with initial distribution $\mathbf{w^t}$; note in particular that $\nu[a] = w_a$. The measure ν is respectively invariant and is a probability measure if and only if $\mathbf{w^t}$ happens to be invariant and a probability vector, but that is not the general case.

We shall call ν the **Shannon-Parry eigenmeasure**; the reason for this name will only become clear in section §27.

We say a measure m on Σ_A^+ has the **Bowen-Marcus property** iff $m_t(s) = m([x_0 \dots x_t])$ for $x_t = s \in \mathcal{A}$ is well-defined; that is, the measure of a thin cylinder set depends only on the last letter.

Note that the Shannon-Parry eigenmeasure ν has this property.

Lemma 18.1. *(Bowen-Marcus) If A is a primitive $0-1$ $(d \times d)$ matrix, then for a measure m on Σ_A^+ which has the Bowen-Marcus property, m is a constant multiple of ν.*

Proof. As above, , we have Perron-Frobenius eigenvectors $A\mathbf{w} = \lambda \mathbf{w}$, $\mathbf{v}^t A = \lambda \mathbf{v}^t$. We shall show that $\exists \gamma > 0$ such that for any $a \in \mathcal{A}$,

$$m[a] = \gamma \cdot \nu[a].$$

This same proof will work for any cylinder set $[a_0 \dots a_l]$, with the same constant γ.

The key idea is to use the mixing property of the invariant measure μ. Choosing two symbols a and s, and fixing a length t, write $[* * \cdots * s]$ for $\sigma^{-t}([s])$, the cylinder set of length t ending in symbol s, and $[a * \cdots * s]$ for $[a] \cap [* * \cdots * s]$; mixing tells us that:

$$\mu([a * \cdots * s]) \to \mu[a]\mu[s]$$

as $t \to \infty$; for the equivalent but non-invariant measure ν, from (67) this becomes:

$$\nu[a * \cdots * s] \to \mu[a] \cdot \mu[s]/v_a = \nu[a] \cdot \mu[s]. \tag{68}$$

We next define a number $\gamma_{t,s}$ by

$$m([* * \cdots * s]) = \gamma_{t,s}\nu([* * \cdots * s]). \tag{69}$$

We shall prove that $\gamma = \gamma_{t,s}$ does not depend on t or s; this will be our constant, with $m = \gamma\nu$.

First note that the same factor $\gamma_{t,s}$ works for all thin cylinders $[b_0 b_1 \dots b_t = s]$. This is because by assumption both m and ν have the property that all cylinders of this length ending in s have the same measure, and adding them up gives (69).

Now we fix the choice of the symbol a. Then $m([a * \cdots * s]) = \gamma_{t,s}\nu([a * \cdots * s])$ since this is a union of thin cylinders.

Now

$$m[a] = \sum_{s \in \mathcal{A}} m([a * \cdots * s]) = \sum_{s \in \mathcal{A}} \gamma_{t,s} \cdot \nu([a * \cdots * s])$$

for each $t \geq 1$, so by (68) this equals

$$\lim_{t\to\infty} \sum_{s \in \mathcal{A}} \gamma_{t,s} \cdot \nu([a * \cdots * s]) = \nu[a] \cdot \lim_{t\to\infty} \sum_{s \in \mathcal{A}} \gamma_{t,s} \cdot \mu[s] = \gamma \cdot \nu[a]$$

where $\gamma = \lim_{t\to\infty} \sum_{s \in \mathcal{A}} \gamma_{t,s} \cdot \mu[s]$ exists.

For clarity we rephrase this last step of the argument: we know that given $\varepsilon > 0$, for any a, s, we have for t large enough,

$$\nu[a * \cdots * s] = (1 \pm \varepsilon)\nu[a] \cdot \mu[s];$$

therefore, for a different ε',

$$(1 \pm \varepsilon')m[a] = \sum_{s \in \mathcal{A}} \gamma_{t,s} \cdot \nu[a] \cdot \mu[s];$$

so

$$(1 \pm \varepsilon')m[a]/\nu[a] = \sum_{s \in \mathcal{A}} \gamma_{t,s} \cdot \mu[s],$$

and hence the limit indeed exists.

So

$$m[a] = \gamma \cdot \nu[a].$$

The constant γ does not depend on s, t or a. Indeed, starting with any other cylinder set $[a_0 \dots a_k]$ in place of $[a]$, at each stage $\gamma_{t,s}$ is the same (for $T > k$) and we end up with the same equation:

$$m([a_0 \dots a_k]) = \gamma \cdot \nu([a_0 \dots a_k]).$$

This proves that for all sets in the Borel $\sigma-$algebra, the same is true, so we are done. □

This is Lemma 2.4 in [BM77]; the proof above is based on the one given there.

18.3. **Finite coordinate changes.** We define a group of transformations on Σ_A^+ as follows. Consider two finite allowed words $(x_0 \dots s), (y_0 \dots s)$ which have the same length and end in the same letter s; we define a map $\gamma : \Sigma_A^+ \to \Sigma_A^+$ as follows: for $w \in [x_0 \dots x_{t-2}s]$, with $w = (x_0 \dots x_{t-2}sw_tw_{t+1} \dots)$, $\gamma(w) = (y_0 \dots y_{t-2}sw_tw_{t+1} \dots)$; on the rest of Σ_A^+ we define γ to be the identity. This is well-defined as $\gamma(w)$ is an allowed string. We call the group of all such maps the **group of finite coordinate changes** of Σ_A^+.

Given a group of homeomorphisms acting on a topological space, we say the action is **uniquely ergodic** if there is a unique invariant probability measure.

We have:

Proposition 18.2. *If A is a primitive $(d \times d)$ $0-1$ matrix, then the action of the group of finite coordinate changes on Σ_A^+ is uniquely ergodic.*

Proof. If m is a probability measure which is invariant for this group, then the Bowen-Marcus property is fulfilled; hence $m = \nu/\nu(\Sigma_A^+)$. □

18.4. **Stationary adic transformations.** We next see how essentially the same orbits can be generated by a single transformation, an **adic transformation** in the terminology of Vershik [Ver94], [Ver95].

Given an $(d \times d)$ $0-1$ matrix A, we define an equivalence relation on Σ_A^+ by:

$$x \sim y \iff \exists N : \forall k \geq N, x_k = y_k.$$

We note that the equivalence class $< x >$ of $x \in \Sigma_A^+$ is exactly the stable set $W^s(x)$ of x for the shift map; this is the set of all y such that $d(\sigma^m(x), \sigma^m(y))$ as $m \to \infty$ (indeed, the distance equals 0 eventually).

We put an order on this countable set as follows. First, assume we are given an order on the symbol set which depends only on the symbol which follows. That is, for each fixed $j \in \mathcal{A}$, defining $\mathcal{A}_j = \{i : A_{ij} = 1\}$, there is a function $\mathcal{O}_j : \mathcal{A}_j \to \{1, 2 \dots, \#\mathcal{A}_j\}$. We call $\mathcal{O}$ an **edge order**. We then use this to order $W_s(x)$ lexicographically, defining this inductively as follows: if N is the least i such that $x_i \neq y_i$, then writing $j = x_{N+1} = y_{N+1}$, we define $x < y$ iff $\mathcal{O}_j(x_N) < \mathcal{O}_j(y_N)$. We define the **successor** of $x \in \Sigma_A^+$, succ(x), to be the least point in $W_s(x)$ which is greater then x, if that exists. One can prove [FT23] that the number of points in Σ_A^+ without a successor or immediate predecessor is at most countable. Call this set $\mathcal{N}_\mathcal{O}$. We define a transformation $T_\mathcal{O} : \Sigma_A^+ \setminus \mathcal{N}_\mathcal{O}$ by $T_\mathcal{O}(x) = \text{succ}(x)$. This is the **adic transformation** defined by the edge order $\mathcal{O}$.

For the simplest example, let $A_{ij} = 1$ for all i, j so $\Sigma_A^+ = \Sigma^+ = \Pi_0^\infty\{1, \dots, d\}$ is the full shift, and order the symbols by their labels, $0 < 1 < 2 < \dots < (n-1)$. There is one point which has no successor, the point $(.111\dots)$; if we define the map there to be $(.0000\dots)$, the only point with no predecessor, then this extension T is continuous. Then T is the Kakutani-von Neumann **adding machine** or **d-adic odometer**.

We now extend the definition of unique ergodicity as follows. Let $(X, \mathcal{B})$ be a set with a $\sigma-$algebra, and let $T : X \setminus N \to X$ be a measurable map where N is a countable set. We shall now say that T is **uniquely ergodic** if there is a unique

invariant non-atomic probability measure. Note that in the case of a shift space Σ_A^+, nonatomic is equivalent to points having mass zero.

Theorem 18.3. (Unique ergodicity for stationary adic transformations) *Let A be an $(d \times d)$ primitive $0-1$ matrix, and let $\mathcal{O}$ be an edge order. Then $(\Sigma_A^+, T_{\mathcal{O}})$ is uniquely ergodic, and $m = \nu/\nu(\Sigma_A^+)$ is the unique $T_{\mathcal{O}}$-invariant non-atomic measure on Σ_A^+.*

Proof. Consider the set of words of length t which end in the same symbol s. By induction, this finite set is totally ordered, that is, there is a 1-to-1 order-preserving map to $\{1, \dots, w_t(s)\}$ where $w_t(s)$ is the number of such words.

We write $\Sigma_{A,\mathcal{O}}^+ \equiv \Sigma_A^+ \setminus < \mathcal{N}_{\mathcal{O}} >$ where $< \mathcal{N}_{\mathcal{O}} >$ is the countable set of points equivalent to $\mathcal{N}_{\mathcal{O}}$, i.e. $\Sigma_{A,\mathcal{O}}^+$ is the set of points whose whose $T_{\mathcal{O}}$-orbit is defined for all past and future times.

Now consider two cylinder sets $[x_0 x_1 \dots x_{t-2} s]$ and $[y_0 y_1 \dots y_{t-2} s]$ of length t. By the above observation, one of these words is least, say $(x_0 x_1 \dots x_{t-2} s)$, and there exists k such that $T^k([x_0 x_1 \dots x_{t-2} s]) = [y_0 y_1 \dots y_{t-2} s]$ in $\Sigma_{A,\mathcal{O}}^+$. If m is a probability measure on Σ_A^+ which gives mass 0 to points, then $m(< \mathcal{N}_{\mathcal{O}} >) = 0$, and if m is $T_{\mathcal{O}}-$invariant, then m satisfies the Bowen-Marcus property. Hence $m = \nu/\nu(\Sigma_A^+)$. □

Remark 18.1. In fact primitivity is necessary and sufficient for adic transformations, if we assume the measure is nonatomic. Primitivity implies minimality for stationary adics; see Theorem 41.8 below. However for minimality, even in the stationary case, this is sufficient but not necessary; see

Remark 18.2. For more on adic transformations see [Ver94], [Ver95] and the references cited therein.

18.5. **Perron-Frobenius theory for nonprimitive matrices: irreducible matrices.** Given a nonnegative $(d \times d)$ matrix M, we say a state $a \in \mathcal{A}$ (the alphabet) **communicates to** state b iff there exists an $n \geq 0$ such that $M_{ab}^n > 0$. Here $M^0 = I$, so every state communicates to itself. Note that this partitions $\mathcal{A}$ into **communicating classes**: the largest subalphabets for which every state communicates with every other. The matrix is termed **irreducible** iff there is a single communicating class, $\mathcal{A}$ itself. Equivalently, M is irreducible iff given $a, b \in \mathcal{A}$ then there exists $n \geq 0$ such that $M_{ab}^n > 0$; for the stronger condition of *primitive*, the iterates are *eventually positive*, i.e. there exists N such that for all $n \geq n$ this works simultaneously for all pairs of states.

The basic example of a matrix which is irreducible but not primitive is a permutation matrix, e.g. the matrix $M = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$ encountered in Example 45. Here M permutes the states in a **cycle** (a, b, c) of period 3, that is, $M : a \mapsto b \mapsto c \mapsto a$. Note also that $M^3 = I$.

As we shall see, essentially, up to a grouping of elements, this is all that can happen. Consider for example M with block form like the permutation matrix, with the blocks B_i and 0 each $(k \times k)$ and primitive: $M = \begin{bmatrix} 0 & B_1 & 0 \\ 0 & 0 & B_2 \\ B_3 & 0 & 0 \end{bmatrix}$, so for $A_s = B_s^3$, then for

$A_1 = B_3B_2B_1, A_2 = B_1B_3B_2, A_3 = B_2B_1B_3$ we have $M^3 = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & A_3 \end{bmatrix}$, with a block form like the identity matrix I; moreover since the diagonal blocks are primitive, there is a further power with diagonal blocks all strictly positive.

There is no need here for the blocks to be all square. Indeed, with M a $(d \times d)$ matrix, let $l_1 + l_2 + \cdots + l_k = d$ with $l_i \geq 1$ and let B_i be $(l_{i+1} \times l_i)$ for $i < k$ and $(l_1 \times l_i)$ for $i = k$, so the matrix sizes are compatible, then A_i is is square, $(l_i \times l_i)$ for each i. (A way of expressing that the matrix sizes are compatible is to state that the diagonal 0 blocks of M are square).

For an example, consider $M = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} = \left[\begin{array}{cc|c} 0 & 0 & 1 \\ 0 & 0 & 1 \\ \hline 1 & 1 & 0 \end{array}\right] = \begin{bmatrix} 0 & B_1 \\ B_2 & 0 \end{bmatrix}$ so $M^2 = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} = \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 1 & 0 \\ \hline 0 & 0 & 1 \end{array}\right]$ with $A_1 = B_1B_2$ and $A_2 = B_2B_1$.

In fact, what happens in general is exactly this, after a change of order on the alphabet:

Theorem 18.4. *Let M be a nonnegative irreducible $(d \times d)$ matrix. Then the alphabet can be partitioned into equivalence classes called* **period classes** *$\mathcal{B}_1, \ldots, \mathcal{B}_k$ such that after a change of order on $\mathcal{A}$, the block form of the matrix is a permutation matrix of these classes. Furthermore, each subblock $M_{\mathcal{B}_i\mathcal{B}_i}$ is primitive, and there exists $n \geq 0$ (the* **period** *of M) such that M^n has block form like the identity matrix: there are primitive blocks on the diagonal.*

To prove this, we can assume that M is a $0-1$ matrix. This defines a map f_A on $\mathcal{A}$, by $f(a) = b$ iff $M_{ab} = 1$. We consider the forward orbit $\mathcal{O}(a) = \{b : f_M(a) = b \text{ for some } n \geq 0\}$ Note that since $M^0 = I$, $a \in \mathcal{O}(a)$. Note that for a communicating class C and some $a \in C$, then $C \subseteq \mathcal{O}(a)$, and $a, b \in C$ have the same orbit.

Now suppose that a belongs to two cycles, of lengths p, q. We claim that if $\gcd(p, q) = 1$, i.e. p, q are relatively prime, then there exists $n > 0$ such that a communicates to all of both cycles at time n.

For this we recall that for $p, q \in \mathbb{N}^*$, then $\gcd(p, q)$, the **greatest common divisor**, has an equivalent definition as the least $k \geq 0$ such that there exist $n, m \in \mathbb{Z}$ with $np + mq = k$. Since for this to hold one of n, m has to be positive and the other negative, equivalently there are $n, m > 0$ such that $k = np - mq$ or $k = mq - np$. This means that (in the first case) $k = np (\mathrm{mod}\, q)$. That is, counting by p units along a circle of length q, we return with a minimum distance of gap k to 0, and hence with future iterates will cover all multiples of k.

Hence if $k = 1$, if we allow ourselves to walk with step lengths either p or q, after some time l we can step on any integer. Equivalently, restricting the the subalphabet of the two orbits, M^l has row a all positive. But it follows that this is true for any other b in the two cycles, and hence M^l is strictly positive.

One can moreover say the following about the eigenvalues of M.

18.6. **Perron-Frobenius theory for nonprimitive matrices: reducing to the irreducible case.**

19. An example: arithmetic codes for hyperbolic toral automorphisms

DO FIRST: example. Introduce NS dynamics. Periodic case. Diagram. Diag/ev's/Boxes. Write down arithmetic code!!!

NEXT: factorization-semigp

NEXT- periodic; general; statement;

NEXT- eqn/split/MP

LATER: skew/ randm/ T flow

Here we analyse in detail a special case, hyperbolic orientation-preserving automorphisms of the two-torus, which will serve to introduce and illuminate a number of important themes in these nores.

In §4.13, see also Example 13, we analysed one particular example, the golden map given by the action on column vectors of the matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$. In particular, we found explicitly the eigenvalues and eigenvectors, giving us the decomposition of the tangent space $\mathbb{R}^2$ of the torus $\mathbb{T} = \mathbb{R}^2/\mathbb{Z}^2$ into expanding and contracting subspaces, $\mathbb{R}^2 = E^u \oplus E^s$. For this we diagonalized the matrix by finding the roots of the characteristic polynomial. Then we used this decomposition to construct a Markov partition for the map.

The key idea of this construction, pioneered by Adler and Weiss in [AW70], is that any partition of the torus whose boundaries consist of line segments from the stable and unstable subspaces of $\mathbf{0}$ will give a Markov partition, due to their remarkable insight in giving a geometric interpretation to the Markov property of probability theory.

Moreover, if this partition is sufficiently fine, it will generate, i.e. separate points, thus providing an a.s. bijective code of the map as a subshift of finite type.

What was discovered by Manning and Adler later on, see also [AF05], was that a more careful construction can yield an especially nice coding. In particular, given a (2×2) nonnegative integer matrix M which is hyperbolic (equivalently, trace ≥ 2), then one can find a Markov partition such that the dynamics of M is coded by the edge shift defined by exactly the same matrix.

The partition is algorithmically and arithmetically defined. The construction is due to Pierre Arnoux in his thesis, [Arn94]. In fact, understanding this procedure is aided by extending our purvue to include *nonstationary dynamics*, given by a sequence of such matrices, as explained in this section, see [AF05]. A deeper look, see §??, takes us into the Teichmüller flow and Veech's way of studying interval exchange transformations, which in fact led to Arnoux' insights.

This same example will lead us in other directions. First, to a concrete case of the Perron-Frobenius theorem; next, to nonstationary dynamics, adic transformations, interval exchanges, the Teichmüeller flow, group boundaries, and the Osceledec theorem, as well as some questions in number theory.

We begin with the following observation, noted above in §4.12. The matix $M = A^2$ has a *second* nice factorization, as

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

We recall that $SL(2,\mathbb{Z})$ is the group of (2×2) matrices with integer entries and with determinant 1. We shall write $SL(2,\mathbb{N})$ for the subsemigroup whose entries are all ≥ 0.

The next lemma is well-known and we do not know the proper attribution; we learned this simple proof a long time ago, perhaps from Rauzy. See also [ES80].

Lemma 19.1. *$SL(2,\mathbb{N})$ is the free semigroup on the two generators*

$$Q = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad \textit{and } P = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Proof. We note that the identity $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is included here as $I = Q^0 = P^0$. Let $A \in SL(2,\mathbb{N})$, with $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. We claim that if $A \neq I$, then either the first column is $\geq$ the second, in the sense that $a \geq b$ and $c \geq d$, or the reverse. If both of these conditions fail then either $a > b$ and $d > c$ or the reverse. We verify this formally:

$$\begin{aligned} &\sim ([(a \geq b) \wedge (c \geq d)] \vee [(a \leq b) \wedge (c \leq d)]) \\ \iff &[(a < b) \vee (c < d)] \wedge [(a > b) \vee (c > d)] \\ \iff &[(a < b) \wedge (c > d)] \vee [(c < d) \vee (a > b)] \end{aligned}$$

However the second of these (i.e. the reverse condition, $b > a$ and $c > d$) cannot happen as this would imply that $bc > ad$ so $ad - bc < 0$, but by assumption the determinant is one.

Since therefore $a > b$ and $d > c$, we have: $a \geq b+1$ and $d \geq c+1$ so the determinant is:

$$ad - bc \geq (b+1)(c+1) - bc = bc + b + c + 1 - bc = b + c + 1.$$

Since $\det A = 1$, we have b and $c = 0$ in which case $A = I$, as claimed.

Now we show that $A \in SL(2,\mathbb{N})$ can be factored as a product of nonnegative powers of Q and P. Writing $A = A_0$, if $A_0 \neq I$ then remove the smaller column from the larger to form A_1. This amounts to writing

$$A_1 = A_0 Q^{-1} \quad \text{or } A_1 = A_0 P^{-1};$$

note that the new matrix A_1 is still in $SL(2,\mathbb{N})$. If A_1 again has one column larger than the other then we continue, producing a sequence $A_0, A_1, \ldots, A_n$. This process terminates with a matrix A_n with determinant one and which has neither column larger than the other. So as shown above, $A_n = I$. Thus, reversing the process, we have factored A as a product of nonnegative powers of Q and P.

We have proved a little more: the preceding argument shows that an element of $SL(2,\mathbb{N})$ which is not the identity can be factored either as $A = A_1 P$ or as $A = A_1 Q$, but not both. Therefore the decomposition of A in terms of Q and P is unique, and

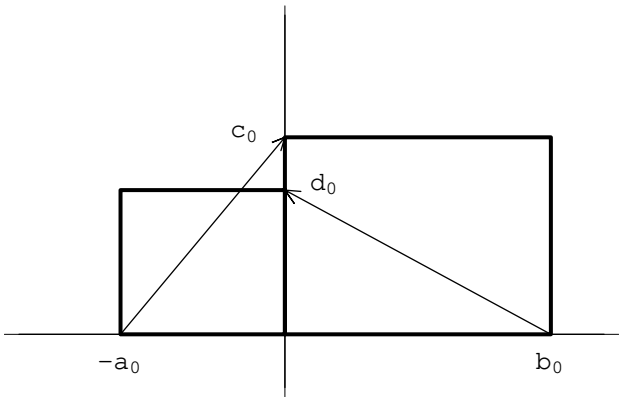


FIGURE 40. Two Boxes

this implies that there can be no nontrivial relations in the semigroup $SL(2,\mathbb{N})$; hence it is free. □

19.1. **The additive and multiplicative families.** Consider a matrix $B = \begin{bmatrix} a & c \\ -b & d \end{bmatrix}$ satisfying:

(1) $a, b, c, d \geq 0$
(2) $\det B = 1$
(3) and either (parity **0**) $0 < a < 1 \leq b$ and $d < c$, and for or (parity **1**) $0 < b < 1 \leq a$ and $c < d$.

Each B can be pictured as defining a pair of rectangles or *boxes*, $a \times d$ and $b \times c$, of total area one, since $\det B = ad + bc = 1$. Note that the longer of a, b is 1, and that for parity **0** the smaller box is on the left (smaller both in width and in height) and the larger on the right; for parity **1** this switches.

Next we describe a process for defining an infinite sequence of such pairs of boxes, associated to a biinfinite contined fraction expansion.

Theorem 19.2. *Given a sequence* $(\dots n_{-1}.n_0 n_1 \dots) \in \Pi_{-\infty}^{\infty}\mathbb{N}^*$*, plus a choice of parity* **0** *or* **1***, we define matrix sequences* B_k*,* D_k*,* A_k*, all of determinant one, by*

$$B_i = \begin{bmatrix} a_i & c_i \\ -b_i & d_i \end{bmatrix},$$

$$D_i = \begin{bmatrix} \lambda_i & 0 \\ 0 & \lambda_i^{-1} \end{bmatrix}$$

and
for parity **0**:
$a_i = [n_i n_{i+1} \dots], b_i = 1, d_i/c_i = [n_{i-1} n_{i-2} \dots]$*, and* $\lambda_i = 1/a_i$*, and* $A_i = \begin{bmatrix} 1 & 0 \\ n_i & 1 \end{bmatrix}$,
for parity **1**:
$b_i = [n_i n_{i+1} \dots], a_i = 1, c_i/d_i = [n_{i-1} n_{i-2} \dots]$*, and* $\lambda_i = 1/b_i$ *and* $A_i = \begin{bmatrix} 1 & n_i \\ 0 & 1 \end{bmatrix}$.
Then these satisfy the equation $B_{i+1} = A_i B_i D_i$.

TO DO: diagram; splitting; Anof fam; Markov coding).

A sequence $(A_i)_{i\in\mathbb{Z}}$ of (2×2) integer matrices defines a linear mapping family on the square torus, via the action on the left on column vectors. Thus, $\mathbb{T}_i = \mathbb{T} = \mathbf{r}^2/\mathbb{Z}^2$ for all i.

As we see in the next proposition, a sequence of diagonalizations defines a second mapping family, now on the tori $\mathbb{T}_{\Lambda_i} = \mathbb{R}^2/\Lambda_i$ where Λ_i is a sequence of parallelogram lattices, each generated by the columns of the corresponding matrix B_i^{-1}.

We have:

Proposition 19.3.
(a)These are equivalent for a sequence $(A_i)_{i\in\mathbb{Z}}$ in $GL(2,\mathbb{Z})$:
(i) Column vector sequences $\mathbf{v}_i^0, \mathbf{v}_i^1$ are an eigenvector sequence pair with eigenvalues λ_i^0, λ_i^1 for the mapping family on columns defined by $f_i(\mathbf{v}) = A_i\mathbf{v}$.
(ii)For the sequences of invertible real matrices B_i and D_i defined by: the columns of B_i are the vectors $\mathbf{v}_i^0, \mathbf{v}_i^1$; the D_i are diagonal matrices with entries λ_i^0, λ_i^1, then these satisfy the equation:

$$A_iB_i = B_{i+1}D_i.$$

(iii) The following diagram commutes, for the action on column vectors, with B_i invertible and D_i diagonal:

$$\begin{array}{ccccccccc}
\cdots\mathbb{R}^2 & \xrightarrow{A_0} & \mathbb{R}^2 & \xrightarrow{A_1} & \mathbb{R}^2 & \xrightarrow{A_2} & \mathbb{R}^2 & \cdots \\
\uparrow{\scriptstyle B_0} & & \uparrow{\scriptstyle B_1} & & \uparrow{\scriptstyle B_2} & & \uparrow{\scriptstyle B_3} & \\
\cdots\mathbb{R}^2 & \xrightarrow{D_0} & \mathbb{R}^2 & \xrightarrow{D_1} & \mathbb{R}^2 & \xrightarrow{D_2} & \mathbb{R}^2 & \cdots
\end{array}$$

(b) In the above situation, the following diagram commutes, where Λ_i is the lattice in $\mathbb{R}^2$ generated by the columns of B_i^{-1}, and where $\mathbb{T} = \mathbb{R}^2/\mathbb{Z}^2$ and $\mathbb{T}_{\Lambda_i} = \mathbb{R}^2/\Lambda_i$.

$$\begin{array}{ccccccccc}
\cdots\mathbb{T} & \xrightarrow{A_0} & \mathbb{T} & \xrightarrow{A_1} & \mathbb{T} & \xrightarrow{A_2} & \mathbb{T} & \cdots \\
\uparrow{\scriptstyle B_0} & & \uparrow{\scriptstyle B_1} & & \uparrow{\scriptstyle B_2} & & \uparrow{\scriptstyle B_3} & \\
\cdots\mathbb{T}_{\Lambda_0} & \xrightarrow{D_0} & \mathbb{T}_{\Lambda_1} & \xrightarrow{D_1} & \mathbb{T}_{\Lambda_2} & \xrightarrow{D_2} & \mathbb{T}_{\Lambda_3} & \cdots
\end{array}$$

The eigenvectors for the second mapping family D_i are the same as for A_i, taking for the second family the standard basis vectors as the eigenvector sequence.
(c)In the above situation, the eigenvector matrices B_i can be normalized to have determinant ± 1; we can additionally have the max of the entries of $\mathbf{v}_i^0$ be 1. The eigenvalues then satisfy $\lambda_i^s\lambda_i^u = 1$ or -1.

Proof. Assuming (ii),

$$A_iB_i = B_{i+1}D_i$$

so

$$A_i\mathbf{v}_i^0 = A_iB_i\begin{bmatrix}1\\0\end{bmatrix} = B_{i+1}D_i\begin{bmatrix}1\\0\end{bmatrix} = \lambda_i^0\mathbf{v}_{i+1}^0$$

The converse argument works, so $(i) \iff (ii)$.

(iii): From equation (ii), we have

$$B_1^{-1} A_0 B_0 = D_0,$$

with change-of-basis matrices B_i^{-1} diagonalizing the actions of the A_i on column vectors. This gives equivalently the commutative diagram.
(b):The image by $B_i^{-1}$of the column vectors $(0,1)$ and $(1,0)$ which generate the integer lattice $\mathbb{Z} \oplus \mathbb{Z}$ are the columns of B_i^{-1}, giving the image lattice Λ_i. On the other hand the standard basis vectors $(0,1)$, $(1,0)$ are an eigenvector sequence for the diagonal mapping family and are taken by B_i to the eigenvector sequences $\mathbf{v}_i^0, \mathbf{v}_i^1$. Since these are eigenvectors, $f(\Lambda_i) = \Lambda_{i+1}$ so the maps are well-defined on the tori. The diagram commutes, and by linearity the λ_i are the same for the two families.
(c): Since A_i is an invertible matrix with integer entries, $\det A_i = \pm 1$; if $\det B_i \neq \pm 1$, we replace it by αB_i such that $\alpha^2 = |\det B_i|^{-1}$. Then $\pm 1 = \det(B_{i+1}^{-1} A_1 B_i) = \det D_i^{-1}$ so $\lambda_i^s \lambda_i^u = \pm 1$. Then let β be the max of the moduli of the entries of $\mathbf{v}_i^0$; multiply B_i on the right by the diagonal matrix with entries β^{-1}, β; this new matrix still has determinant 1 and has the property desired. □

Frequently in math it turns out that by considering an apparently harder or more general actually simplifies...shed light on...

Tree of free semigroup; not just vertices but path to boundary at infinity.

We return to this subject below in §25.6 from a different perspective.

We call Q, P the **additive generators** because of their connection with the additive continued fraction.

–nonstationary subshift of finite type

–mult family/ add family, infinite/ fte measure base

We have introduced continued fractions and the Gauss map in §11.4.

There, we wrote the continued fraction of an irrational $x \in (0,1)$ as

$$x = [n_0 n_1 \dots] = \cfrac{1}{n_0 + \cfrac{1}{n_1 + \cdots}},$$

with $n_i \in \mathbb{N}^* = \{1, 2, \dots\}$, and noted that the Gauss map is isomorphic to the left shift σ on $\Pi_0^\infty \mathbb{N}^*$.

We shall need a bilateral version of the continued fraction: $x = [\dots n_{-1}.n_0 n_1 \dots] \in \Pi_{-\infty}^{\infty} \mathbb{N}^*$, with the left shift map σ.

We begin the expansion of x with n_0 rather than with the more traditional choice of n_1 to agree with the usual shift notation of ergodic theory, where 0 indicates the coordinate of present time; this is especially natural since we are considering the bilateral shift.

We call Q, P the **additive generators** because of their connection with the additive continued fraction.

–nonstationary subshift of finite type

–mult family/ add family, infinite/ fte measure base

We have introduced continued fractions and the Gauss map in §11.4.

There, we wrote the continued fraction of an irrational $x \in (0,1)$ as

$$x = [n_0 n_1 \dots] = \cfrac{1}{n_0 + \cfrac{1}{n_1 + \cdots}},$$

with $n_i \in \mathbb{N}^* = \{1, 2, \dots\}$, and noted that the Gauss map is isomorphic to the left shift σ on $\Pi_0^\infty \mathbb{N}^*$.

We shall need a bilateral version of the continued fraction: $x = [\dots n_{-1}.n_0 n_1 \dots] \in \Pi_{-\infty}^{\infty} \mathbb{N}^*$, with the left shift map σ.

We begin the expansion of x with n_0 rather than with the more traditional choice of n_1 to agree with the usual shift notation of ergodic theory, where 0 indicates the coordinate of present time; this is especially natural since we are considering the bilateral shift.

We define $\mathcal{B}$ to be the collection of matrices $B = \begin{bmatrix} a & c \\ -b & d \end{bmatrix}$ satisfying:

(1) $a, b, c, d \geq 0$
(2) $\det B = 1$
(3) $\mathcal{B}$ is a union of disjoint sets $\mathcal{B} = \mathcal{B}^{\mathbf{0}} \cup \mathcal{B}^{\mathbf{1}}$, where for $B \in \mathcal{B}^{\mathbf{0}}$, $0 < a < 1 \leq b$ and $d < c$, and for $B \in \mathcal{B}^{\mathbf{1}}$, $0 < b < 1 \leq a$ and $c < d$.

We say $B \in \mathcal{B}$ has **parity** $\epsilon = \mathbf{0}$ or $\epsilon = \mathbf{1}$ when it is in $\mathcal{B}^{\mathbf{0}}$ or $\mathcal{B}^{\mathbf{1}}$ respectively.

20. Nonstationary and random dynamics

20.1. **Skew products.**

Exercise 20.1. *Show the Morse-Thue substitution dynamical system can be modelled as a skew product over the odometer transformation. (Hint: you can do this with two points in the fiber over each point).*

20.2. **Examples and introduction.**

There are several initial motivations for a study of nonstationary dynamics:

–it can provide a sort of "completion" for a class of stationary dynamical systems;

–even for a fixed map, a limiting procedure used in some construction (e.g. for Markov partitions, for the scenery flow, for stable manifolds) may be viewed as a nonstationary system;

-"random perturbations" of a stationary system give a nonstationary one;

–a stationary non-hyperbolic system may arise as the transverse dynamics to a hyperbolic stationary one;

–nonstationary dynamics can arise in renormalization theory for the "non-fixed-point" case;

–nonstationary constructions already arise naturally in ergodic theory, e.g. in cutting-and-stacking constructions;

example: toral auts; conj to parallelogram model

–defn; mapping fam (cts along cpt metric)

–conjugacy/ stable sets.

–hyperbolicity/ uniqueness

–additive/ mult family

20.3. **Random transformations.** –random ergodic theorem
–Markov operator erg thm
–harmonic projection !!!
–subsequence erg thm (Furst; Bourgain !)

21. Adic transformations

(TO DO: Introduce edge shifts and nsfts)

21.1. **Sturmian sequences and the golden rotation.** What we have defined above in 4.11 is a **stationary** substitution dynamical system; one can more generally have a *non*stationary situation, built from an infinite sequence of substitutions $\dots \rho_{-1}, \rho_0$.

The precise general definition will be given below, but we mention an important example, the **golden** substitution dynamical system (because of its relation to the golden number, see below...). We define $\rho_0(0) = 01, \rho_0(1) = 1$ and $\rho_0(0) = 0, \rho_0(1) = 01$, and then alternate ρ_0 and ρ_1. Since these switch periodically, we can replace then by a single substitution $\rho = \rho_1 \circ \rho_0$ which sends 0 to 010 and 1 to 10; the sequence $\underline{x}^+ = .010100101001010010\dots$ is a fixed point for ρ, and its $\omega-$limit set, or rather that of $(\dots 0000.x^+)$, defines Ω. We mention that this sequence is a particular case of a **Sturmian sequence**. General Sturmian sequences can be built in a similar way but now we really have to allow ρ_0 and ρ_1 to be chosen in a nonperiodic way. See [AF01].

21.2. **Cutting and stacking: Interval exchange transformations; Rauzy induction.**

22. Group actions and the Cayley graph.

To get a feeling for more general infinite group actions, we need a geometric approach to the groups themselves, based on the twin notions of Cayley graph, factor space, and the related idea of the boundary at infinity.

We recall Definition 2.1: group action, and the orbit of an element.

The orbit of a point should be thought of as a copy of the group (or semigroup) itself, wrapped around inside the space on which it acts. So to visualize an orbit, we should first visualize the (semi)group itself.

We begin with a finitely generated group G or semigroup S, and a list of generators, $\mathcal{G} = (g_1, \dots g_n)$. The **Cayley graph** of G or S consists of one vertex for each element, connected by edges labelled by the generators. For a semigroup draw an edge labelled by $g_i \in \mathcal{G}$ from vertex g to h iff $g_i g = h$. For the case of a group, we do the same for the augmented list of generators together with their inverses, $\widetilde{\mathcal{G}} = (g_1, g_1^{-1}, \dots g_n, g_n^{-1})$.

A **word** is a finite string of generators. We consider a finite collection $\mathcal{R}$ or words, with $\widehat{\mathcal{R}}$ denote the subgroup generated by $\mathcal{R}$. A **relation** is an element of $\widehat{\mathcal{R}}$.

We denote by F_n the free group on n generators, and form the factor group $F_n/\widehat{\mathcal{R}}$.

For semigroups we proceed similarly: we write FS_m for the free semigroup on m generators (also called **letters**); we can get from this construction the free group as follows: beginnign with $m = 2n$ generators, labelled $(g_1, g_1^{-1}, \dots g_n, g_n^{-1})$, we factor by

the collection of relations $\widehat{\mathcal{R}}$ generated by $\mathcal{R} = \{g_i g_i^{-1} :, 1 \leq i \leq n\}e$. That is, we mod out by the relations $gg^{-1} = e$.

Conversely, any finitely generated group G can be represented in this way, as a factor group of F_n, and so as a facto semigroup of FS_{2n}. The relations $\widehat{\mathcal{R}}$ are, geometrically, the words which form closed loops starting at e in the Cayley graph.

For the case of an abelain group, the law $ab = ba$ is achieved by including the relation $f^{-1}g^{-1}fg$.

The Cayley graph in the case of a group is **homogeneous** in that its geometry everywhere is the same, and is just like that n the identity e.

A homomorphism from a group G to a group H can be visualized by a continuous map of the Cayley graphs; a good example to keep in mind is the homomorphism from the free group F_2 on two generators (a, b) to the free abelian group on two generators, $\mathbb{Z}^2$, and from $\mathbb{Z}^2$ to $\mathbb{Z}_6 = \mathbb{Z}_2 \oplus \mathbb{Z}_3$. See Fig. ??.

Remark 22.1. From the point of view of Category Theory, the directed edges ...??

factor groups, free semigroup, free group, free abelian group, finite abelian group
fundamental domain; lattice subgroup
random walk
boundary at infinity
hitting measure
example: Parry measure
normal subgroup
left/right actions; free semigroup boundary and IFS/ Cantor set
Free semigroup and group automorphisms
Kleinian limit set, Patterson measure

22.1. **A Markov partition for the doubling map on the torus.** Sinai and Bowen proved the existence of Markov partitions for hyperbolic toral automorphisms in any dimension; however when the stable and unstable manifolds no longer have dimension one, as in the case of the two-torus, the partition boundaries are no longer pieces of these smooth submanifolds but rather are constructed by an approximation procedure using a geometric series. In fact, as shown by Bowen, for dimension ≥ 3 for automorphisms and ≥ 2 for expanding endomorphisms of the torus, Markov partition boundaries can never be smooth. But remarkably, they can still be found with a wonderful *fractal* geometry. These developments are due in the first instance to Bedford in his thesis [Bed86b], [Bed86a], building on Dekking's work on L-systems and substitution dynamical systems, and on Gilbert's study of complex number bases [Dek82], [Gil82].

We illustrate this with the most basic example, the 2-to-1 endomorphism we call the doubling map on the torus,

$$T : z \mapsto (1 + i)z.$$

In Fig. 41 we see a Markov partition with fractal boundaries for this map, the **twindragon** Markov partition, which has two partition elements and codes the map

as a Bernoulli coin-tossing shift. So this partition is not just Markov, it's the simplest case of Markov: it is truly independent!

This can be understood as *arithmetic expansions* with a complex base. For the first, the base is $(1+i)$. We are following Misutani and Ito [MI87]. That is, any point in the torus can be expressed as $x = \sum_0^\infty x_k(1+i)^{-k}$, for $\underline{x}^+ = (.x_0x_1\dots) \in \sigma^+ = \Pi_0^\infty\{0,1\}$. This expression is unique except on the partition boundaries, where there may be two or, exceptopnally three, expressions, as is clear from the figure. We return to consider this example more deeply in §22.2.

22.2. **Dynamics and construction of fractal Markov partitions for the doubling map on the torus; a free semigroup homomorphism and a free group automorphism.** In §22.1 we described a Markov partition for the the doubling map on the torus

$$T : z \mapsto (1+i)z,$$

defined via arithmetic expansions with (complex) base $(1+i)$, with digits $\underline{x}^+ = (.x_0x_1\dots)$ where $x_i \in \{0,1\}$. The union of the two partition elements gives a fundamental domain for the action of $\mathbb{Z}^2$ on $\mathbb{R}^2$, better adapted to studying the dynamics than the usual fundamental domain (the unit square).

Here we take a different approach. First we describe the dynamics of the map geometrically with the help of a re-coding of the map as a renewal shift. Next we illustrate how to draw the boundary of the new fundamental domain by a limiting procedure. After that we explain how a similar limiting procedure produces a non-stationary Markov partitions which converges to the limiting stationary partition.

We recall that a renewal shift is a countable state Markov shift, see ???

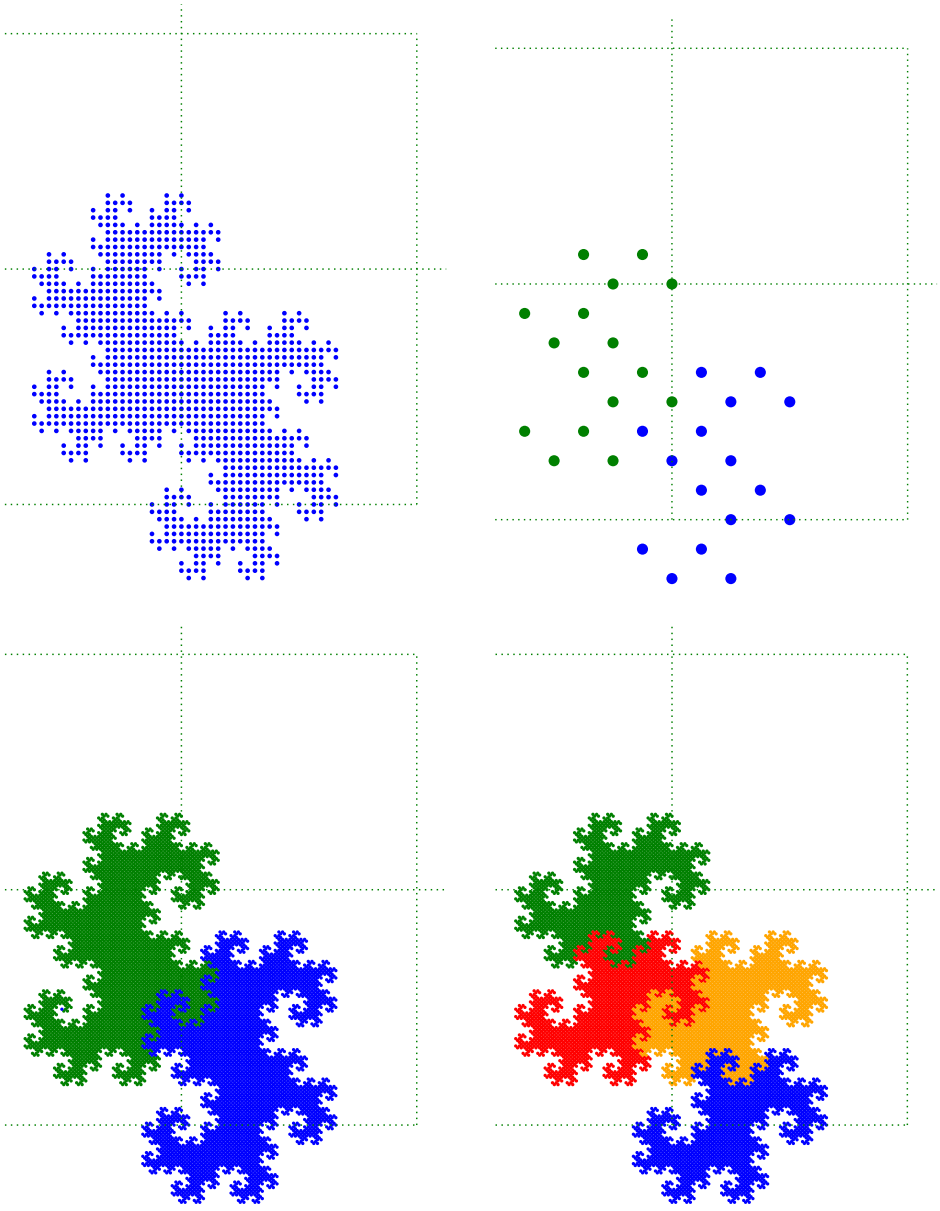


FIGURE 41. On left: complex base expansion for the base $1+i$, giving a fundamental domain for the torus acted on by the map $T : z \mapsto (1+i)z$, modulo the lattice $\mathbb{Z} + i\mathbb{Z}$. Thus, the region shown tiles the complex plane by action of this lattice. For the first figure on the left each point shown corresponds to one of the 2^{10} cylinder sets of the Bernoulli 2-shift of length 10. The second figure shows an approximation to the Markov partition $\mathcal{P} = \{P_0, P_1\}$ for $a_0 = 0, 1$, using the cylinder sets of length 5. Next is the Markov partition with 13 digits, and lastly $\mathcal{P} \vee T^{-1}(\mathcal{P})$.

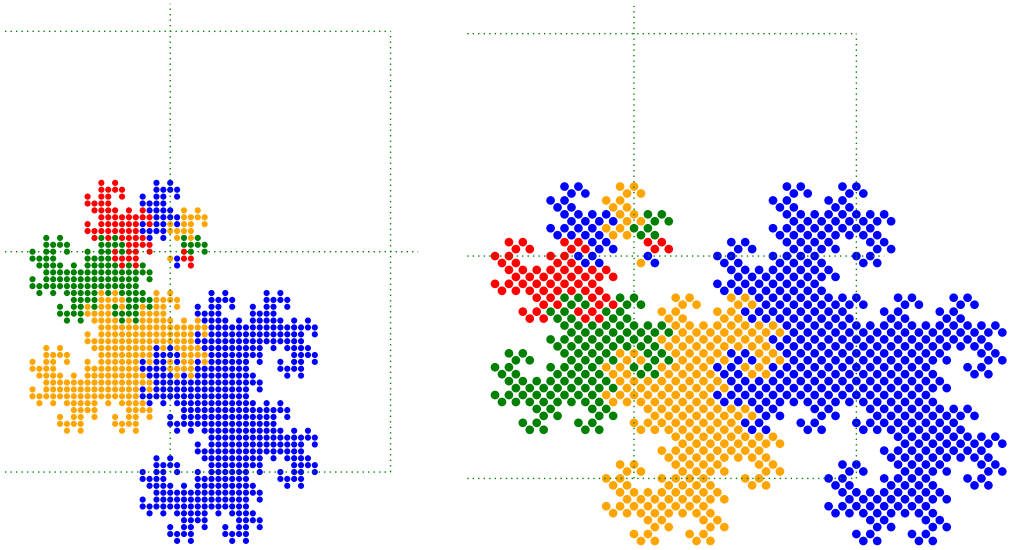


FIGURE 42. A renewal process coding for the doubling map on the torus, showing the dynamics of the map $z \mapsto (1+i)z$ on the twindragon fundamental domain of Fig. 41, on the left with a point for each of the 2^{10} cylinders of length 10. The regions correspond in the shift space to $[.000000001] \mapsto [.000000001] \mapsto \cdots \mapsto [.001] \mapsto [.01] \mapsto [.1]$ which maps to the whole space Σ^+, as the large blue region on the left maps to that on the right, which translates by the element -1 of the lattice $\mathbb{Z}[i]$ to the fundamental domain.

The approximation construction of a fractal boundary Markov partition for the map T can be motivated as an actual nonstationary Markov partition sequence, for this stationary map. See Fig. 44. For this, beginning with a pair of rectangles at time 0, we pull this back by the inverse image to four sets at time -1, 8 at time -2 and so on. The result is a commutative diagram of finite "shift" spaces, for time 0 being one set (the square) and hence a combinatorial space with a single point, $\{\emptyset\}$ (the semiconjugating sequence of maps φ_i are defined everywhere but the partition boundaries):

$$\begin{array}{ccccccccc} \cdots & & \mathbb{T} & \xrightarrow{T} & \mathbb{T} & \xrightarrow{T} & \mathbb{T} & \xrightarrow{T} & \mathbb{T} \\ & & \downarrow{\scriptstyle \varphi_{-3}} & & \downarrow{\scriptstyle \varphi_{-2}} & & \downarrow{\scriptstyle \varphi_{-1}} & & \downarrow{\scriptstyle \varphi_0} \\ \cdots & \Pi^0_{-2}\{0,1\} & \xrightarrow{\sigma} & \Pi^0_{-1}\{0,1\} & \xrightarrow{\sigma} & \{0,1\} & \xrightarrow{\sigma} & \{\emptyset\} & \end{array}$$

Next, following Misutani and Ito in [MI87], we show how to construct this sequence of partition boundaries directly, and hence algorithmically by computer.

As explained by Misutani and Ito [MI87], this construction can be seen as a two-dimensional analogue of a substitution dynamical system. We consider the free group on two generators F_2 with generating set $\{a, b\}$. Extending this to include the inverses, we take $\mathcal{A} = \{a, b, a^{-1}, b^{-1}\}$ and consider this as an alphabet, with $\mathcal{A}^*$ denoting the

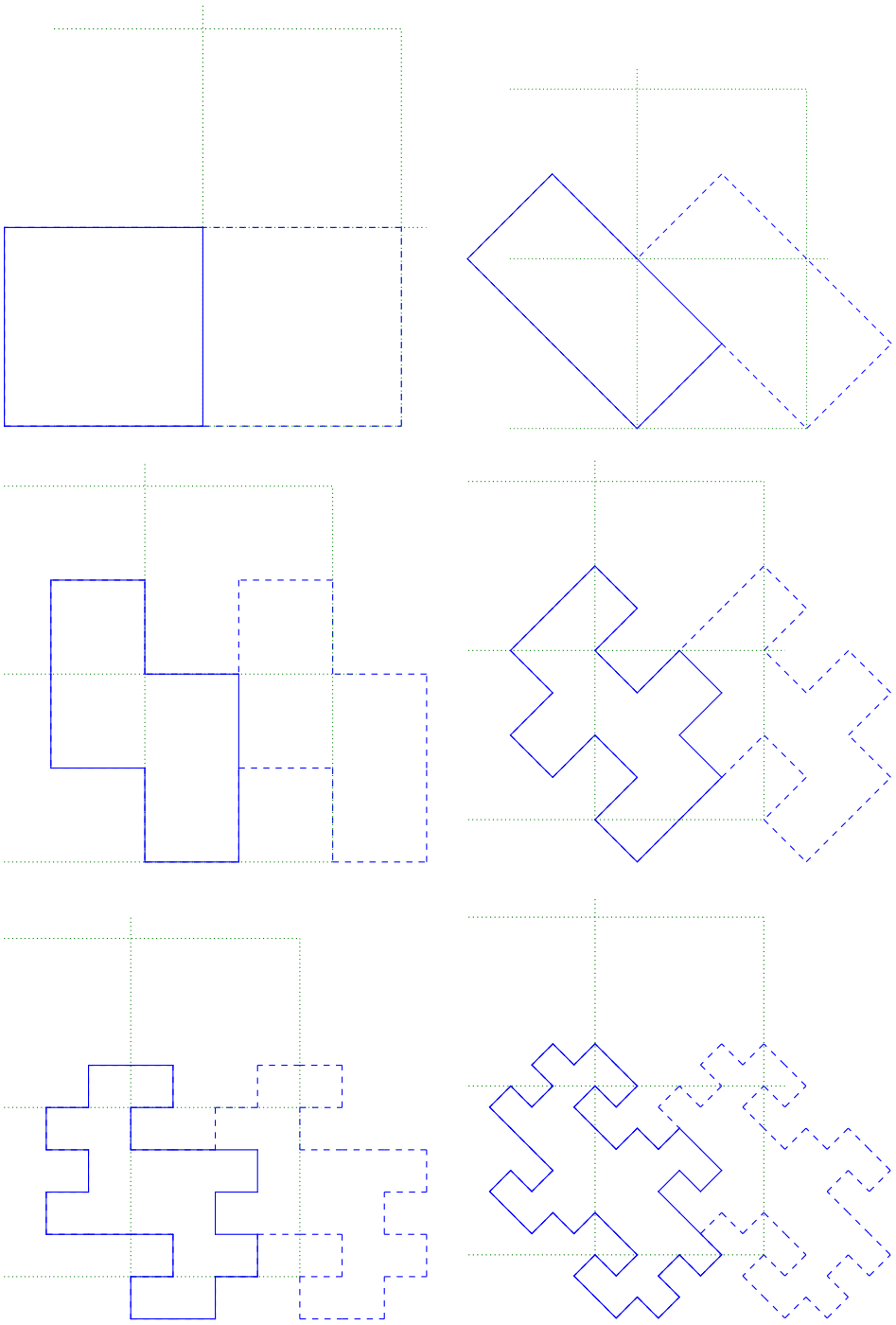


FIGURE 43. To draw the boundary curve of the invariant fundamental domain of Fig. 41, we begin at time 0 with the square to the lower left of the point $0 \in \mathbb{C}$. Then we double it by adding 1, and take the inverse image by the map $z \mapsto (1+i)z$ on the plane, giving the domain for time -1. Repeating this procedure gives a nonstationary sequence of fundamental domains for the torus, shown here for times $0, -1, \ldots, -5$. Each element of the tiling at stage n is mapped onto a union of two pieces at stage $n-1$. Modulo the lattice this image is a single copy, with the map giving a double cover. The limiting fundamental domain is stationary and so works for infinite future times as well.

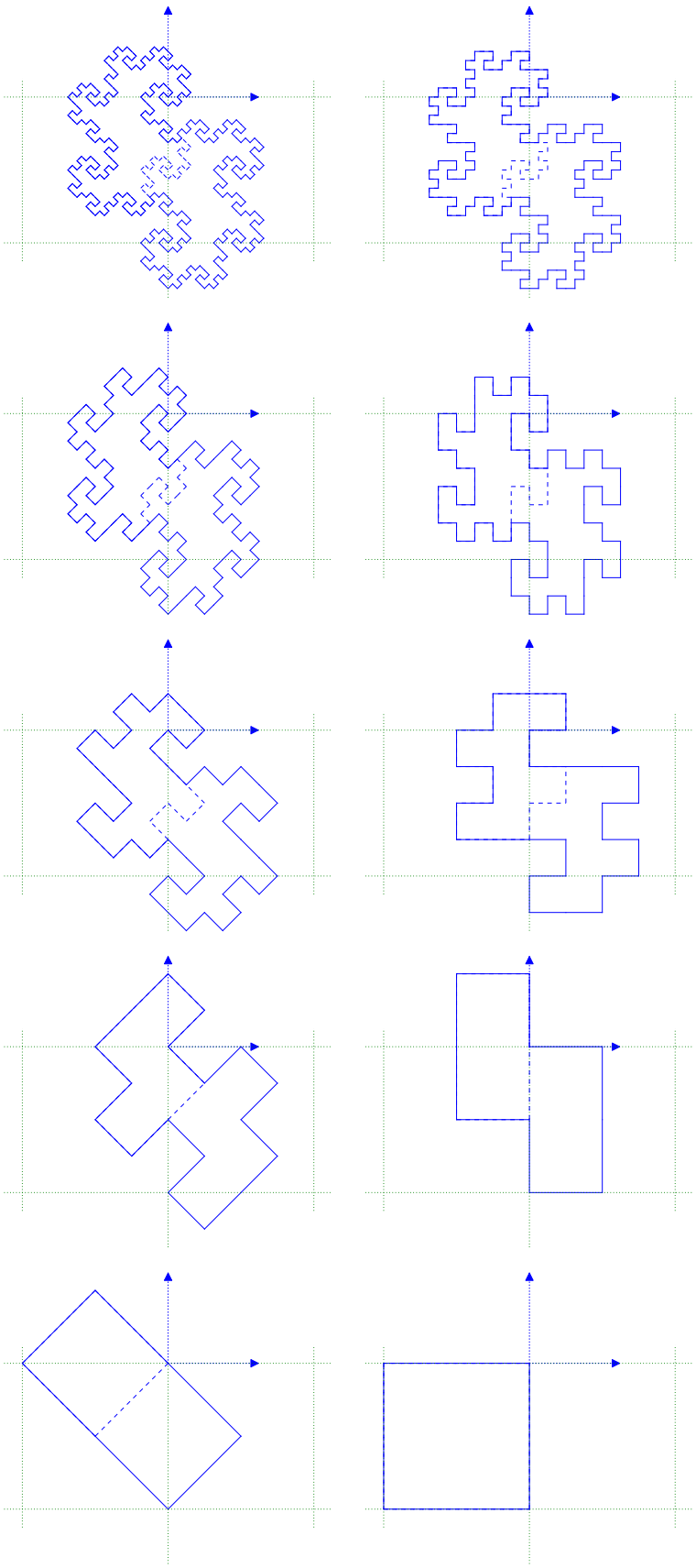


FIGURE 44. A nonstationary Markov partition sequence mapping forward by the stationary map $T : z \mapsto (1+i)z$; the partition at time $n-1$ consists of a pair of inverse images from time n. Shown for times $-9, -8, \ldots, -1, 0$; the limit at $-\infty$ is the stationary fractal Markov partition of Fig. 41.

collection of finite (non-reduced) words in these four letters, including the **empty word** denoted $\emptyset$; this is the free semigroup on four generators, FS_4. Then $F_2 = \mathcal{A}^*/\sim$ where $\sim$ is the equivalence relation generated by $aa^{-1} = a^{-1}a = e$ (with e the identity element) and similarly for b. Note that $\mathcal{A}^* \to \mathcal{A}^*/\sim$ is the semigroup homomorphism from FS_4 to F_2 given by moving from non-reduced to reduced words.

We define a substitution $\rho : \mathcal{A} \to \mathcal{A}^*$ by

$$\rho(a) = ab;\ \rho(b) = a^{-1}b$$

and extending to the inverses by $\rho(g^{-1}) = (\rho(g))^{-1}$. Then ρ extends via concatenation to $\rho :\ \mathcal{A}^* \to \mathcal{A}^*$. Setting $\rho(e) = e$, then ρ is a homomorphism of the free semigroup. Passing to $\mathcal{A}/\sim$ we have a homomorphism of F_2. (Note that conversely *any* homomorphism of F_2 is defined by a substitution satisfying the property $\rho(g^{-1}) = (\rho(g))^{-1}$.

Next we define a homomorphism φ from FS_4 to $(\mathbb{C}, +)$, the additive subgroup of $\mathbb{C}$, by defining it on $\{a, b\}$ via

$$\varphi(a) = 1, \varphi(b) = i.$$

Moreover, as in that key paper, we describe a second **Tetradragon** Markov partition, which codes the map as a subshift of finite type with an alphabet of four symbols, as contrasted to the first **twindragon** Markov partition, which codes the map as a Bernoulli shift.

The figures show convergence to a space-filling curve which fills in Markov partition elements of the tetradragon. Here is how the curve is defined. We are following [MI87], which in turn builds on Dekking [Dek82].

Considering a word $(a_0, a_1, \dots a_n) \in \mathcal{A}^*$, we define a curve K in $\mathbb{C}$ which connects the "dots", the images by φ of $e, a_0, a_0a_1, \dots, a_0a_1 \cdots a_n$, by line segments, as follows. Writing $z_0 = 0$, $z_i = \varphi(a_0a_1 \cdots a_i)$, then we set inductively $K(t) = z_i + t(z_{i+1} - z_i)$ for $t \in [i, i+1]$. This defines a continuous curve $K : [0, n] \to \mathbb{C}$.

Writing K_n for this curve for the words $\rho^n(a)$, we have that $K_n :\ [0, 2^n] \to \mathbb{C}$. The figures depict these curves as n increases. (For clarity, the square corners have been rounded off.) Recalling from §11.6 the scaling flow τ_s of exponent $\alpha > 0$ on $\mathcal{C}(\mathbb{R}, \mathbb{R})$, we now extend this idea:

Definition 22.1. Given $\alpha > 0$, we topologize $\mathcal{C}$ by an extension of the geometric topology: we use uniform convergence on compact subsets of time in the domain and the geometric topology in the range. We now define the **scaling flow** of exponent α on the space of continuous functions $\mathcal{C}$ from $\mathbb{R}$ to $\mathbb{C}$ by

$$\tau_s : f(t) \mapsto \frac{f(e^s t)}{e^{s\alpha}}.$$

We claim the following:

Proposition 22.1. *$K_n \subseteq K_{n+1}$. Denoting $\widehat{K} = \cup_{n\geq 1} K_n$, then for τ_t the scaling flow of exponent $\alpha = 1/2$, there exists a periodic point $\widehat{K}$ of period $\log 2$ for τ_t, such that $\tau_{t_k}(\widehat{K}) \to K$ as $k \to \infty$ in the geometric topology, where $t_k = 8 \cdot 2^k$. In fact, taking $G_k = (1+i)^{-k}\widehat{K}(2^k t)$, then G_k is a Cauchy sequence, uniform in time, with*

$$|G_{k+1}(t) - G_k(t)| \leq 2^{-k/2}$$

for all t.

Convergence of G_k is shown in Fig. 45.

Restricting the domain of the space-filling curve K to $[0,1]$, the image of the interval is the dragon fractal shown in Fig. 46. This ilustrates four rotated copies of the dragon, which give a fundamental domain for the torus and so tile the plane when translated by the integer lattice $\mathbb{Z}\oplus\mathbb{Z}$. The large square shown here is $[-1,1]\times[-1,1]$ so the torus is $\mathbb{R}^2/(2\mathbb{Z})^2$, or in complex notation $\mathbb{C}/2\mathbb{Z}[i]$.

This defines a second fractal Markov partition for the doubling map of the torus, the Ito-Misutani Tetradragon.

To study its coding, and to understand the relationship to the construction of the curve K, we begin with alphabet $\mathcal{A}=\{0,1,2,3\}$ and transition matrix

$$M=\begin{bmatrix}1&1&0&0\\0&1&1&0\\0&0&1&1\\1&0&0&1\end{bmatrix}.$$

We consider the vertex subshift of finite type Σ_M^+ and a string $x=(.x_0x_1\dots)\in\Sigma_M^+$. We also consider the edge paths $e=(.e_0e_1\dots)$ with edge alphabet labelled by two-blocks: $\mathcal{E}=\{e=e_{(ab)}:,a,b\in\mathcal{A},M_{ab}=1\}$. An edge path determines, and is determined by, a vertex path via $e_i^-=x_i,e_{i+1}=x_{i+1}$ where $e_i=e_{(ab)}$ and $x_i=a,x_{i+1}=b$.

We can realize the expanding Markov map defined by M as a discontinuous map of the interval or circle, see Fig. 47, or as a continuous, piecewise differentiable map of the *cloverleaf* $T:\mathcal{CL}\to\mathcal{CL}$ (topologically a *bouquet of four circles*), see Fig. 48, represented by the substitution with alphabet $\mathcal{A}=\{a,b,c,d\}$. and $\rho(a)=ab,\rho(b)=cb,\rho(c)=cd,\rho(d)=ad$. Here the substitution maps from left to right in the (stationary) Bratteli diagram. The space-filling curve is then given by the continuous map $\gamma:\mathcal{CL}\to\mathbb{T}$. This is a semiconjugacy from the cloverleaf to the doubling map of the torus, and moreover is a measure-preserving transformation, indeed it takes one-dimensional Lebesgue measure to two-dimensional Lebesgue measure.

The tetradragon curve factors onto the twindragon via the map given by rotation R of $\pi/2$. What does this do to the torus? It maps it to the quotient space of $\mathbb{R}^2$ with factor group generated by $\{\mathbb{Z}^2,R\}$. A fundamental domain is the triangle which is $1/4$ of the unit square.

23. Hyperbolic space and the Hilbert and projective metrics

Now we take an excursion through some basic complex analysis and hyperbolic geometry (the theory of Möbius transformations on the upper half space); this will serve two purposes in these notes. First, it will give us the background for discussing the important examples of geodesic and horocycle flows on Riemann surfaces; second, it will bring us to the Hilbert metric on a convex set, and the related projective metric on a positive cone. These provide powerful tools in dynamics, leading in particular to a second, contraction-mapping proof of the Perron-Frobenius theorem, which will prepare the way for our proof of the Ruelle Perron-Frobenius theorem.

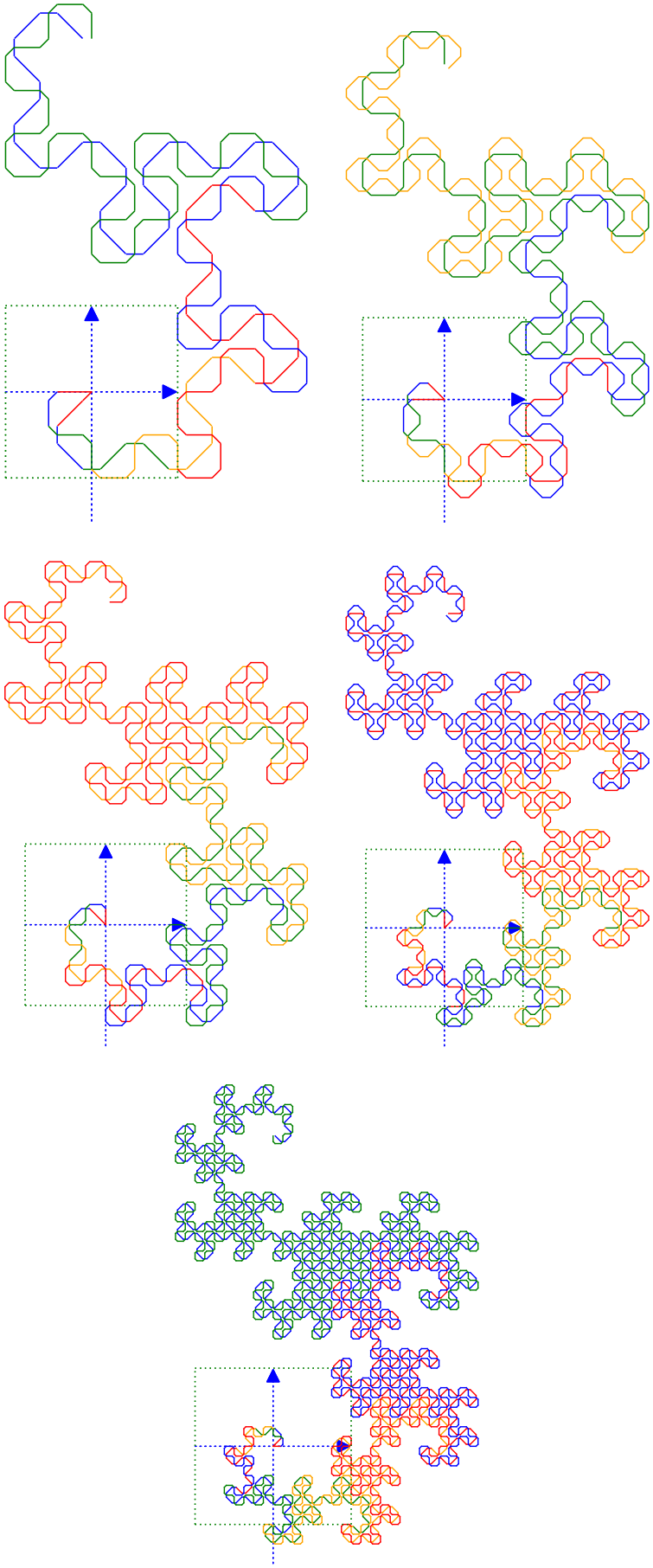


FIGURE 45. Convergence to a self-similar path in $\mathbb{C}$ with exponent $\alpha = 1/2$: a periodic orbit of period $8 \cdot \log 2$ of the scenery flow. The figures depict (parts of) G_k, G_{k+1} for $k = 5, 6, 7, 8, 9$. Note that the inital step rotates through eighth roots of unity.

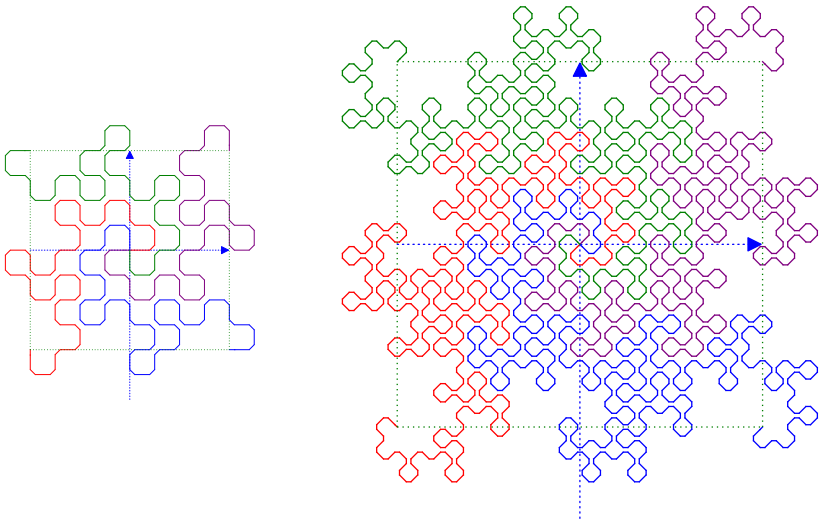


FIGURE 46. Approximating the Tetradragon, the image of four space-filling curves, each a rotated copy of the dragon.

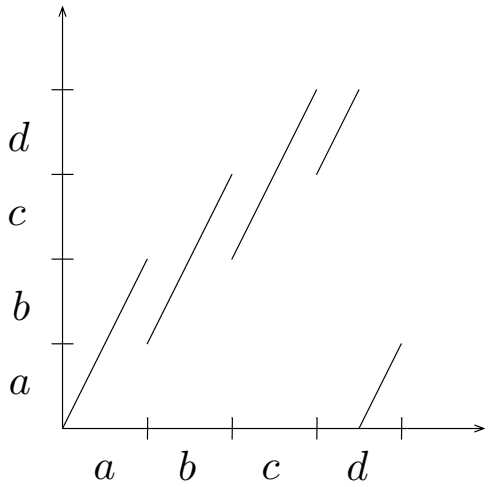


FIGURE 47. Markov map of the interval corresponding to substitution ρ.

23.1. **Complex Möbius transformations and the cross-ratio.** We shall work with the extended complex plane $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, or equivalently the Riemann sphere S^2 with the complex structure coming from its embedding as the unit sphere in $\mathbb{R}^3$. The map $\varphi : S^2 \to \widehat{\mathbb{C}}$ called *stereographic projection* describes this correspondence explicitly: one projects radially from the north pole $N = (0, 0, 1)$ of S^2 to $\mathbb{C}$ embedded in $\mathbb{R}^3$ as the $xy-$ plane. Sending N sent to ∞ gives a bijection from S^2 to $\widehat{\mathbb{C}}$, see Fig. 50. See [Ahl66] for a proof of the following:

Proposition 23.1. *Stereographic projection $\varphi : S^2 \to \widehat{\mathbb{C}}$ is biholomorphic and sends circles to lines and circles.* □

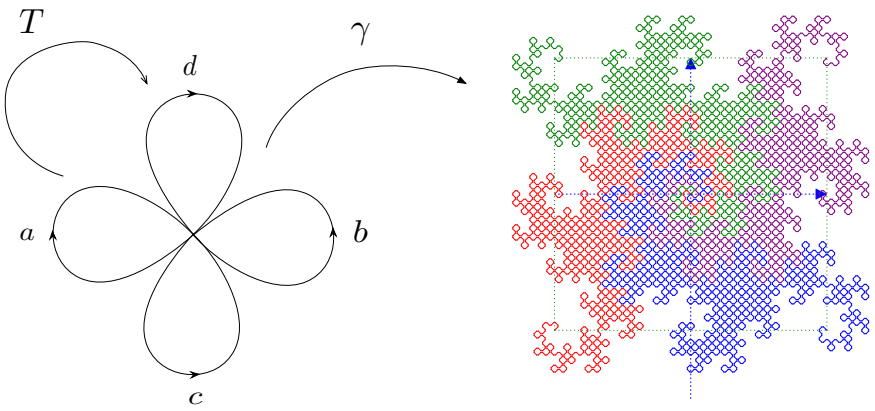


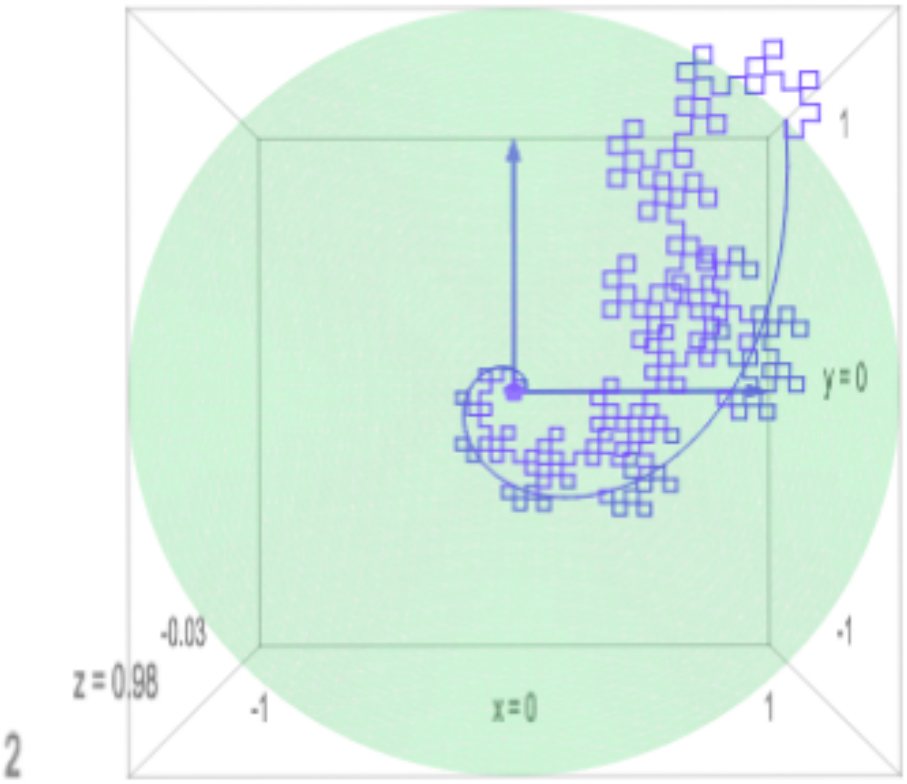
FIGURE 48. On the left a train-track map on the "cloverleaf", on the right a fundamental domain for the the Ito-Misutani Tetradragon, showing the four elements of a Markov partition for the doubling map $z \mapsto (1+i)z$ on the torus $\mathbb{C}/\mathbb{Z}[i]$ (the vectors shown are 1/2 of the standard basis vectors). The space-filling curve γ gives a measure-preserving semiconjugacy from this Markov map on the cloverleaf to the doubling map on the torus.

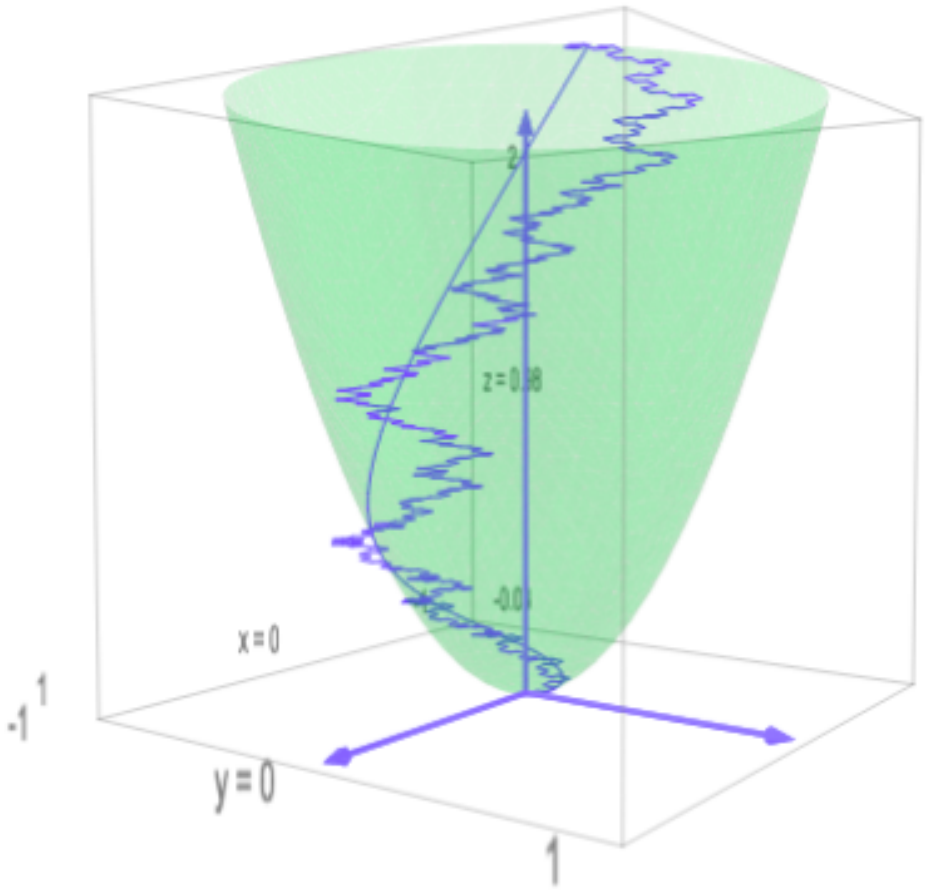
The extended complex plane is also identified with one-dimensional complex projective space, by definition the space of lines through $\mathbf{0}$ in $\mathbb{C}^2 \setminus \{\mathbf{0}\}$.

To explain this, we begin with the general setting of a vector space V over a field F. *Projective space* PV is then the collection of lines through the origin in V with the quotient topology. That is, we define an equivalence relation on $V \setminus \mathbf{0}$ by $\mathbf{v} \sim \mathbf{w} \iff \mathbf{v} = \lambda \mathbf{w}$ for some $\lambda \neq 0$ in F. For example, $P\mathbb{R}^2$, *one-dimensional real projective space* or the *real projective line*, is homeomorphic to a circle. We can see this in three different ways: first, a collection of representatives for these equivalence classes is given by the upper half circle; the endpoints are identified, resulting in a topological circle. Secondly, let S^1 denote the circle of radius 1 with center $(0,1)$ in $\mathbb{R}^2$. A line through $(0,0)$ passes through a unique point of the circle, giving our map. Note that the $x-$axis is sent to the point $S = (0,0)$ which is the "south pole" of the circle. Thirdly, this same line passing through a point (x,y) with $y > 0$ passes through a unique point in the line $y = 1$, mapping (x,y) to $(\tilde{x}, 1)$ where $\tilde{x} = x/y$; we then send $(\tilde{x}, 1)$ to $\tilde{x} \in \mathbb{R}$. We include the $x-$axis by sending it to ∞, giving a map from $P\mathbb{R}^2$ to $\widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$. This is the one-point compactification of $\mathbb{R}$, which we know is the circle. This defines *homogeneous coordinates* for projective space $P\mathbb{R}^2$: the coordinates for the point $(x,y)/\sim$ being $(\tilde{x}, 1)$ in the line $y = 1$.

A fourth correspondence comes out of this same picture: define a map from the circle S^1 about $(0,1)$ to $\widehat{\mathbb{R}}$ by sending a point $(x,y) \in S^1$ to its homogeneous coordinates. Indeed this is like the stereographic projection defined above for $\widehat{C}$, but using the south rather than north pole. See Fig. ??

Now a general projective space PV is not just a topological space; it also comes equipped with a collection of natural transformations, the *projective transformations*, which are induced from linear maps on V.


1
y = 0
-0.03
z = 0.98
-1
x = 0
1
-1
2


2
z = 0.98
-0.0
x = 0
-1
1
y = 0
1

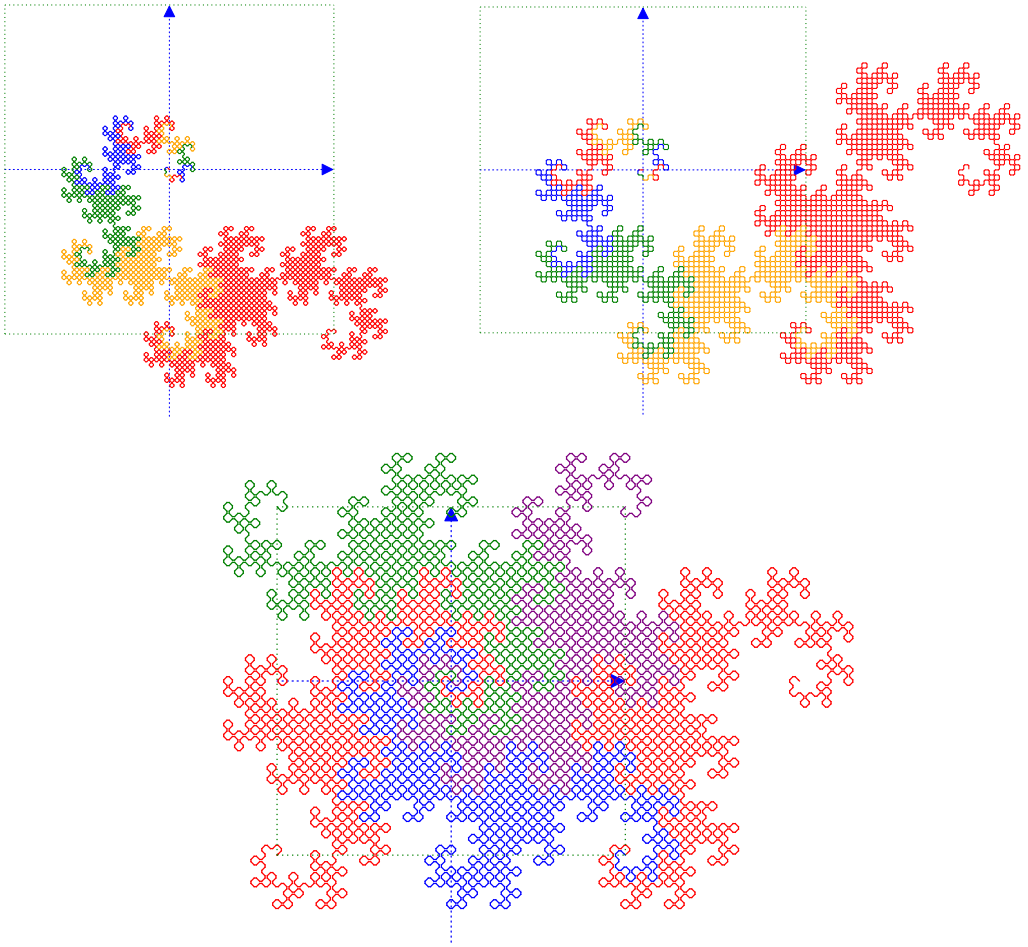


FIGURE 49. Dynamics on the blue Dragon: after applying $z \mapsto (1+i)z$, the red region translates by the element -1 of the lattice $\mathbb{Z}[i]$ to the red Dragon.

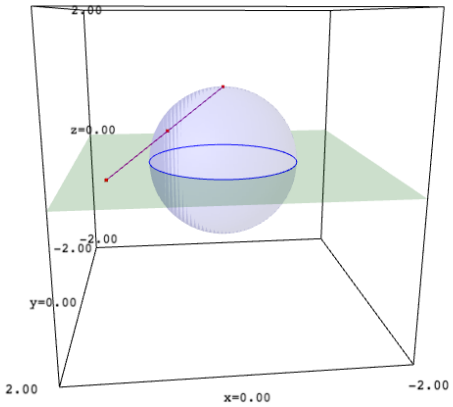


FIGURE 50. Stereographic projection

We shall examine this for one- dimensional complex projective space, $P\mathbb{C}^2$, which as we explain below can be identified with $\widehat{C}$. But first we recall some basics from complex analysis.

Definition 23.1. One has these three definitions:
(i) a map is *holomorphic* iff it is complex differentiable, i.e. its derivative, given by the usual limit, exists and is a complex number. If this number is $z \in \mathbb{C}$, then writing $z = re^{i\theta}$ for $r \geq 0$, since by Euler's formula $e^{i\theta} = \cos\theta + i\sin\theta = c + is$, we see that the multiplication map $w \mapsto z \cdot w$ is in real coordinates

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \mapsto r \begin{bmatrix} c & -s \\ s & c \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

In other words the function has a very special type of $\mathbb{R}^2$-derivative: a dilation composed with a rotation.
(ii) This implies the map is *conformal*: angles and orientation are preserved infinitesimally. By contrast, an *anti*conformal map preserves angles but *reverses* orientation; the simplest example is $z \mapsto \overline{z}$ where for $z = a+ib$, its *complex conjugate* is $\overline{z} = a-ib$. A general antiholomorpic map is given by a holomorphic map preceded or followed by complex conjugation, so the $\mathbb{R}^2$-derivative is a rotation composed with a reflection in a line through $(0,0)$. Note that for both conformal and anticonformal maps, infinitesimal circles are taken to infinitesimal circles (not ellipses, which is the general case).
(iii) A function is *(complex) analytic* iff it has a power series expansion near a point.

The first remakable fact from complex analysis is that all three defintions are equivalent. In particular, knowing a function has one continuous complex derivative, i.e. in $\mathcal{C}^1$, implies, very differently from the real case, it is not only infinitely continuously differentiable ($\mathcal{C}^\infty$) but has a power series (is $\mathcal{C}^\omega$).

The most basic examples are the biholomorphic maps of the Riemann sphere; these form a group under composition, the *Möbius transformations*. Given a disk in the Riemann sphere, its complement is another disk, the simplest examples being the two hemispheres of the sphere S^2, or the upper and lower half-planes which make up $\mathbb{C}$.

Each such disk has an interesting hyperbolic metric defined on its interior; the Möbius transformations are essential for understanding this geometry.

These maps are defined as follows. Given a (2×2) matrix with complex entries $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, such that $\det(A) = ad - bc \neq 0$, we define a map of $\widehat{\mathbb{C}}$ by

$$f_A(z) = (az + b)/(cz + d),$$

extending by continuity to ∞, so $f_A(\infty) = a/c$. This is a **complex Möbius transformation**; we write Möb($\mathbb{C}$) for the collection of all such maps. Another name is **linear fractional transformation**. This comes from viewing f_A as a map of projective space. The connection with projective space shall also allow us to see exactly what this collection of maps is algebraically, as we next explain.

Recall that $GL(2,\mathbb{C})$ is the group of (2×2) complex invertible matrices; one writes $SL(2,\mathbb{C})$ for the subgroup of matrices with determinant 1, and $PGL(2,\mathbb{C})$, $PSL(2,\mathbb{C})$

for the projective spaces, that is for the quotient groups given by the equivalence relation $A \sim \lambda A$ for $\lambda \neq 0$.

We let $GL(2,\mathbb{C})$ act on $\mathbb{C}^2$ by multiplication of column vectors on the left. This induces an action on the quotient space $P\mathbb{C}^1 = \{\text{lines through the origin in}\mathbb{C}^2\}$ (as above, the projective line, or *one-dimensional complex projective space*) since linear maps preserve lines and the origin. Defining a projection $\pi : \mathbb{C}^2 \setminus \{(0,0)\}$ onto $\widehat{\mathbb{C}}$ by

$$\pi : \ (z,w) \mapsto z/w, \ \text{(where for } z \neq 0,\ z/0 = \infty)$$

we see that π induces a bijection of $P\mathbb{C}^1$ with $\widehat{\mathbb{C}}$. (Note that the homogeneous coordinates of (z,w) are $(z/w,1)$, a point in a complex plane inside of $\mathbb{C}^2$, which we then map to simply $z/w \in \widehat{\mathbb{C}}$).

$$\begin{array}{ccc} \mathbb{C}^2 \setminus \{\mathbf{0}\} & \xrightarrow{A} & \mathbb{C}^2 \setminus \{\mathbf{0}\} \\ \downarrow \pi & & \downarrow \pi \\ \widehat{\mathbb{C}} & \xrightarrow{f_A} & \widehat{\mathbb{C}} \end{array} \tag{70}$$

We have:

Proposition 23.2. *The action of the matrix A on the left on the collection of lines in $\mathbb{C}^2$ is isomorphic via π to the Möbius transformation f_A of $\widehat{\mathbb{C}}$. The map $A \mapsto f_A$ from $GL(2,\mathbb{C})$ onto Möb$(\mathbb{C})$ takes matrix multiplication to composition: $f_{AB} = f_A \circ f_B$. In particular, each f_A is invertible, with inverse $f_{A^{-1}}$. The Möbius transformations form a group under composition; it is a factor group of $GL(2,\mathbb{C})$.*

Proof. The composition of π with matrix multiplication by A applied to the vector (w_1,w_2) is the extended complex number $(aw_1+bw_2)/(cw_1+dw_2) = (a(w_1/w_2)+b)/(c(w_1/w_2)+d) = f_A(\pi(w_1,w_2))$. Thus the action on projective space is given by f_A via the identification π of $P\mathbb{C}^1$ with $\widehat{\mathbb{C}}$. Automatically, the action of GL on lines is a group; so this passes over to Möbius transformations. The rest follows immediately. □

Remark 23.1. The picture in Fig ?? simultaneously shows the map from $P(\mathbb{R}^2)$ to S^1 and $\widehat{\mathbb{R}}$ via homogeneous coordinates. It is tempting to think this same thing works for $P(\mathbb{C}^2)$, but that is not true. We do have the sterographic projection of Fig. 50 from S^2 to $\widehat{\mathbb{C}}$ but clearly we cannot realize $\mathbb{C}^2$ in this drawing since we would need four real dimensions. Indeed $P(\mathbb{C}^2)$ is the collection of complex lines through the origin in $\mathbb{C}^2 \setminus \{\mathbf{0}\}$, and each line itself is a copy of $\mathbb{C}$. Now the drawing of Figg.?? suggests looking instead at $P(\mathbb{R}^3)$, but this space is *not* topologically the same as the sphere; it is a more complicated topological space called a crosscap: a disk with opposite points identified. This is because it is the upper hemisphere, with opposite points on the boundary circle identified. Now the boundary itself is therefore $\mathbb{R}P^1$, that is a circle; it can be visualized as twisting and folding over itself a rubber band to make a circle half as long. Fig.???, which we learned from [Thu97], compares the torus, Klein bottle and crosscap, the last two being nonorientable surfaces. It is easy to picture the first two, but it seems hard to visualize the crosscap surface! Spivak

nevertheless tries to draw it, on p. 17 of Vol. I of his Differential Geometry series, [Spi79].

Proposition 23.3.
(i) There exists a unique $h \in M\ddot{o}b(\mathbb{C})$ that fixes the points $0, 1$ and ∞: the identity map.
(ii) Given three distinct points $x, y, w \in \widehat{\mathbb{C}}$, there exists a unique $f \in M\ddot{o}b(\mathbb{C})$ that takes x to 0, y to 1, and w to ∞.
(iii) Given three distinct points $x, y, w \in \widehat{\mathbb{C}}$, there exists a unique $f \in M\ddot{o}b(\mathbb{C})$ (the identity map) that fixes these points.
(iv) Given two triples x, y, w and $\tilde{x}, \tilde{y}, \tilde{w}$ of distinct points in $\widehat{\mathbb{C}}$, there exists a unique $f \in M\ddot{o}b(\mathbb{C})$ that takes x to $\tilde{x}$, y to $\tilde{y}$, and w to $\tilde{w}$.

Proof. The proof is broken down into these simple steps; note that case (iv) includes all the others as the triples need not be disjoint sets.

For (i), if $h(z) = (az+b)/(cz+d)$ then $h(0) = b/d$ so if h fixes 0 then $b = 0$. Next, $h(\infty) = a/c$ so we have that $c = 0$. Hence $h(1) = (a+b)/(c+d) = a/d$ and so $a = d$, and $h = f_A$ for $A = \lambda I$ for some $\lambda \in \mathbb{C}$, where I is the identity matrix; and h is therefore the identity map.

To prove (ii), setting

$$f(z) = \frac{x-z}{x-y} \cdot \frac{y-w}{z-w}, \tag{71}$$

we see that f takes x, y, w to $0, 1, \infty$ and is Möbius, so $f = f_A$ for some $A \in GL$. This proves existence.

Now if there is another such map f_B, then $f_B^{-1} \circ f_A$ is Möbius and fixes $0, 1, \infty$ so equals the identity map by part (i), hence $f_A = f_B$.

To prove (iii), let f be the map from part (ii) and suppose g is a map that fixes the points x, y, w. Then the conjugate $f \circ g \circ f^{-1}$ fixes $0, 1, \infty$ so is the identity by part (ii), hence $g = f^{-1} \circ \mathrm{id} \circ f = \mathrm{id}$ also.

To prove (iv), let f be as in (ii) and let $\tilde{f}$ be the unique map which takes $\tilde{x}, \tilde{y}, \tilde{w}$ to $0, 1, \infty$. Then considering $\tilde{f}^{-1} \circ f$ proves existence. Now let g, h be two such maps; then $h^{-1} \circ g$ fixes x, y, w and so from part (iii) it follows that $g = h$. □

Corollary 23.4. *The group $M\ddot{o}b(\mathbb{C})$ is isomorphic to the factor group $PGL(2, \mathbb{C})$, which is naturally identified with $PSL(2, \mathbb{C})$ and $SL(2, \mathbb{C})/\{\pm I\}$.*

Proof. Let $\sim$ denote the equivalence relation $A \sim \lambda A$ for $\lambda \in \mathbb{C} \setminus \{0\}$, so $PGL(2, \mathbb{C}) = GL(2, \mathbb{C})/\sim$. Now given $A \in GL(2, \mathbb{C})$, the maps f_A and $f_{\lambda A}$ are equal; conversely, as in (i) of Proposition 23.3, if f_A is the identity map then $A \sim I$; it follows that if $f_A = f_B$, then $A \sim B$. Hence $PGL(2, \mathbb{C})$ is isomorphic to Möb$(\mathbb{C})$. Now for $A, B \in SL(2, \mathbb{C})$, if $A \sim B$ then $A = \lambda B$ but since $1 = \det A = \det(\lambda B) = \lambda^2 \det B = \lambda^2$, $\lambda = \pm 1$. Thus $PSL(2, \mathbb{C}) = SL(2, \mathbb{C})/\{\pm I\}$. Lastly, given $A \in GL(2, \mathbb{C})$, since $\det \lambda A = \lambda^2 \det A$, we can find an equivalent matrix with determinant 1, and so $PGL = PSL$. □

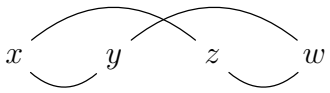
Remark 23.2. We shall write an element of $PSL(2, \mathbb{C})$ as a matrix A of determinant one though actually it is the equivalence class $\{A, -A\}$.

Definition 23.2. The **cross-ratio** of $x, y, z, w \in \widehat{C}$ where x, y and w are distinct points is

$$[x, y, z, w] = \frac{x - z}{x - y} \cdot \frac{y - w}{z - w}. \tag{72}$$

Thus, defining a function by $f(z) = [x, y, z, w]$, this is by the above the unique Möbius transformation that takes x, y, w to $0, 1, \infty$.

The name cross-ratio perhaps comes from the following mnemonic for the above formula:



Proposition 23.5. *Möbius transformations preserve cross-ratios.*

Proof. Let $g \in \text{Möb}(\mathbb{C})$; we are to show that, given three distinct points x, y, w, then for any $z \in \widehat{C}$,

$$[g(x), g(y), g(z), g(w)] = [x, y, z, w].$$

We define two further Möbius transformations, the first by

$$f(z) = [x, y, z, w]$$

and the second by

$$h(z) = [g(x), g(y), z, g(w)].$$

We wish to show that $h(g(z)) = f(z)$.

Now $h \circ g$ agrees with f on the three points x, y, w. And part (iv) of Proposition 23.3 says that a Möbius transformation is determined by where it sends three distinct points. Therefore $h(g(z)) = f(z)$ for all z, as desired. □

Lemma 23.6. *A Möbius transformation f can be written either as a composition $T_\gamma \circ M_\beta$ or as $T_\gamma \circ M_\beta \circ J \circ T_\alpha$ where T_γ is translation $T_\gamma(z) = z + \gamma$, M_β is complex multiplication $M_\beta(z) = \beta z$, and J is* multiplicative inversion *$J(z) = 1/z$.*

Proof. We start with the general form for a Möbius transformation

$$f(z) = (az + b)/(cz + d).$$

If $c = 0$, we have $f(z) = (a/d)z + (b/d)$ and we are in the first case with $f = T_{b/d} \circ M_{a/d}$. If $c \neq 0$, we can assume that $c = 1$, so

$$f(z) = \frac{az + b}{z + d};$$

comparing this with the equation

$$f(z) = T_\gamma \circ M_\beta \circ J \circ T_\alpha(z) = \gamma + \frac{\beta}{z + \alpha} = \frac{\gamma z + (\gamma\alpha + \beta)}{z + \alpha}$$

we see that taking $\alpha = d, \gamma = a$ and $\beta = b - ad$, or equivalently $b = \gamma\alpha + \beta$, we can pass from one form to the other, so the two representations for a Möbius transformation with $c \neq 0$ are also equivalent. □

anti

Remark 23.3. We call J multiplicative inversion to distinguish it from geometric inversion (inversion in a circle or line), which we call simply *inversion*. The formula for inversion in the unit circle is $z \mapsto 1/\overline{z} = \overline{J(z)} = J(\overline{z})$. Inversion in the real axis is conjugation, i.e. reflection, $z \mapsto \overline{z}$. Unlike J, these are *anti*conformal maps; see Example 23 below.

Proposition 23.7. *Möbius transformations preserve* $\{circles, lines\} \subseteq \mathbb{C}$.

Remark 23.4. Equivalently, from Proposition 23.1, $f \in \text{Möb}(\mathbb{C})$ takes circles to circles in the Riemann sphere.

Proof. It is clear that translations, rotations and real dilations preserve lines and preserve circles (separately), so it remains to show that inversion preserves the collection of circles and lines. We follow [MH87]; essentially the same proof, but written in purely complex notation, is given in [JS87].

We know a line or circle in the plane can be written as the solution to

$$Ax + By + C(x^2 + y^2) = D$$

where not all three of the constants A, B, C are zero, the lines corresponding to $C = 0$. For $z = x + iy$, note that $1/z = \overline{z}/|z| = u + iv$ with $u = x/(x^2 + y^2)$ and $v = -y/(x^2 + y^2)$. Since $u^2 + v^2 = 1/(x^2 + y^2)$, we have $x = u/(u^2 + v^2)$ and $y = -v/(u^2 + v^2)$. Thus the previous equation is equivalent to

$$Au - Bv - D(u^2 + v^2) = -C$$

and the condition that not all of A, B, C are zero is equivalent to that not all three of A, B, D are zero. This is a circle for $D \neq 0$. (Note that lines through zero go to lines through zero via inversion, the case with both C and D equal to 0.) □

23.2. **Real Möbius transformations and central projection.** Let us first note that:

Lemma 23.8. *For any* $n \geq 1$, $GL(n, \mathbb{C})$ *is pathwise connected.*

Proof. Given $A, B \in GL(n, \mathbb{C})$, the simple idea to connect them by a path $tA+(1-t)B$ for $t \in [0, 1]$ doesn't work as the determinant may be zero along the way. However, working instead in the Lie algebra, which is $\mathcal{M}_n(\mathbb{C})$ (the collection of all $(n \times n)$ complex matrices) does the trick: the exponential map $A \mapsto \exp(A)$ (defined for matrices by the power series for numbers, see 35.5) sends $\mathcal{M}_n(\mathbb{C})$ (the collection of all $(n \times n)$ complex matrices) onto $GL(n, \mathbb{C})$. For this proof we do need the basic fact that exp is onto the largest connected subgroup containing the identity element, which in this case is the whole group. □

We define Möb($\mathbb{R}$), the **real Möbius transformations**, to be the subgroup of Möb($\mathbb{C}$) such that the matrix A has real entries. Using the canonical embeddings of $\mathbb{R}^2$ in $\mathbb{C}^2$ and $\widehat{\mathbb{R}} = \mathbb{R} \cup \infty$ in $\widehat{\mathbb{C}} = \mathbb{C} \cup \infty$, the commutative diagram (70) becomes:

$$\begin{array}{ccc} \mathbb{R}^2 \setminus \{\mathbf{0}\} & \xrightarrow{A} & \mathbb{R}^2 \setminus \{\mathbf{0}\} \\ \downarrow \pi & & \downarrow \pi \\ \widehat{\mathbb{R}} & \xrightarrow{f_A} & \widehat{\mathbb{R}} \end{array} \tag{73}$$

We write $\mathbb{H} = \{z = x + iy : \ y > 0\}$ for the **upper half plane**. We have:

Proposition 23.9. *Möb*$(\mathbb{R})$ *is isomorphic to* $PGL(2,\mathbb{R})$ *via the map* $A \mapsto f_A$*. These are the complex Möbius transformations which preserve the real line.* $PGL(2,\mathbb{C})$ *is pathwise connected, while* $PGL(2,\mathbb{R})$ *has two connected topological components, those with determinant* > 0 *and* < 0*. Those with positive determinant form a normal subgroup* $PGL^+(2,\mathbb{R})$ *which is isomorphic to* $PSL(2,\mathbb{R})$*; these correspond via* f_A *to the subgroup* $Möb^+(\mathbb{R})$ *of maps in* $Möb(\mathbb{C})$ *which preserve the orientation of* $\mathbb{R}$ *and preserve the upper half plane.*

Proof. Certainly a matrix in $PGL(2,\mathbb{R})$ gives a Möbius transformation which maps $\mathbb{R}$ to $\mathbb{R}$. For the converse, let x, y, w be the images by f of the points $0, 1, \infty$; the formula for the inverse of f is given in formula (71), and the matrix has real entries.

From Corollary 23.4 therefore Möb$(\mathbb{R})$ is isomorphic to $PGL(2,\mathbb{R})$.

Given $A \in PGL(2,\mathbb{C})$ with $\det(A) > 0$, then iA has determinant < 0, and $\gamma(t) = tA + (1-t)iA$ is a path between them with nonzero determinant for each time t. By contrast, since $\det(\cdot)$ is a continuous function on the matrix entries, in $PGL(2,\mathbb{R})$ such a path must pass through a point with the value zero.

□

Remark 23.5. By definition the Lie algebra $\mathfrak{g}$ of a Lie group G is the tangent space at the identity; the exponential map takes $\mathfrak{g}$ to G but may not be onto: $\exp(\mathbf{0}) = \mathbf{e}$ the identity in G, and the image of $\mathfrak{g}$ is the largest connected subgroup containing $\mathbf{e}$. Indeed, the proof just given of Lemma 23.8 shows that the image of the exponential map from the Lie algebra is connected. See §35.5 and 35.15.

The simplest nonconnected example is the multiplicative subgroup $G(1,\mathbb{R}) = G^- \cup G^+ = \mathbb{R}^{*,-} \cup \mathbb{R}^{*,+}$ of $\mathbb{R}$, where $\mathbb{R}^{*,-} = (-\infty, 0$ and $\mathbb{R}^{*,+} = (0, +\infty)$; then $\mathfrak{g}_{\mathbb{R}} = \mathbb{R}$ and of course the image $\exp(\mathbb{R}) = \mathbb{R}^{*,+} = G^+(1,\mathbb{R})$. Note that by contrast for the multiplicative group $G(1,\mathbb{C}) = \mathbb{C} \setminus \{\mathbf{0}\}$, which is connected, the Lie algebra is $\mathfrak{g}_{\mathbb{C}} = \mathbb{C}$ and $\exp(\mathbb{C}) = G(1,\mathbb{C})$. Now the coset (but not subgroup) $G^-(1,\mathbb{R})$ also has to be the image of something in $\mathfrak{g}_{\mathbb{C}}$, and indeed, $\exp(t + \pi i) = -e^t$.

One way to understand that with the real field the image of the map $\exp(\cdot)$ is the collection of orientation-preserving matrices comes from the formula $\exp(\mathrm{tr}A) = \det(\exp(A))$, where tr is the trace.

Remark 23.6. In contrast to the complex case, $PSL(2,\mathbb{R})$ and $PGL(2,\mathbb{R})$ are not the same: consider the matrix $A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$, which has determinant -1 so is not in $SL(2,\mathbb{R})$, whereas in $PGL(2,\mathbb{C})$, A is equivalent to iA which has determinant 1. Thus A is an element of $PSL(2,\mathbb{C})$ and $PGL(2,\mathbb{R})$ though not of $PSL(2,\mathbb{R})$. Here

$f_A : z \mapsto -z$. Another example is $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, again with determinant -1; now $f_A : z \mapsto 1/z$.

For perspective drawing in art, objects in $\mathbb{R}^3$ are centrally projected from the point of view of the artist onto a plane (think of a window, or a canvas); an example we have already encountered is sterographic projection from the sphere to $\widehat{\mathbb{C}}$, though in that case the "window" lies beyond the object! Another example is homogeneous coordinates in $P(\mathbb{R}^n)$. Here we consider such projections in $\mathbb{R}^2$.

Let l_1, l_2 be two extended real lines in $\mathbb{R}^2$ (that is, each includes a point ∞ which gives the one-point compactification) which do not pass through the origin. To be more precise, these are lines in RP^2, the real plane with a projective circle at infinity; each line has a distinct point at infinity, which are equal if and only if the lines are parallel; see Remark 23.7 below.

We define a map $P_{l_2,l_1} : l_1 \to l_2$ by sending $\mathbf{v}_1 = (x_1, y_1) \in l_1$ to the unique point $\mathbf{v}_2 = (x_2, y_2)$ in l_2 which is projectively the same. That is, they have the same homogeneous coordinates $(x_1/y_1, 1) = (x_2/y_2, 1)$ in the line $y = 1$.

We call P_{l_2,l_1} the **central projection** from l_1 to l_2.

We note that:

Lemma 23.10. *Central projections compose: given three lines l_1, l_2, l_3 in RP^2 which miss the origin, $P_{l_1,l_2} \circ P_{l_2,l_3} = P_{l_1,l_3}$. In particular, $P_{l_1,l_2}^{-1} = P_{l_2,l_1}$.*

Proof. The first statement is immediate from the picture; the second is then a corollary. □

Next, we identify each line with $\widehat{\mathbb{R}}$ by choosing a point of origin, an orientation and a scale. These three are determined by specifying two distinct points $\mathbf{v}, \mathbf{w} \in l$: the first is the origin, while the second plays the role of $1 \in \widehat{\mathbb{R}}$, indicating both the positive direction and choice of scale. We write $l(\mathbf{v}, \mathbf{w})$ for the line with this choice of origin, orientation and metric. We call the points $\mathbf{v}, \mathbf{w}$ **base points**.

Given two lines l_1, l_2, which miss the origin and base points $\mathbf{v}_1, \mathbf{w}_1$ and $\mathbf{v}_2, \mathbf{w}_2$, the central projection P_{l_2,l_1} thus induces a map from $\widehat{\mathbb{R}}$ to $\widehat{\mathbb{R}}$. We will characterize such maps:

Proposition 23.11. *A central projection P_{l_1,l_2} from line l_2 to line l_1, both of which miss the origin and with base points $\mathbf{v}_2, \mathbf{w}_2$, $\mathbf{v}_1, \mathbf{w}_1$ induces a real Möbius transformation f of $\widehat{\mathbb{R}}$, and conversely all $f \in$ Möb$(\mathbb{R})$ arise in this way.*

Proof. We begin with l_1 the horizontal line $y = 1$, with base points $\mathbf{v}_1 = (0, 1)$ and $\mathbf{w}_1 = (1, 1)$. Let φ denote the isometry from l_1 to $\mathbb{R}$ determined by this choice of base points. Thus, $\varphi(x, y) = x$ and $\varphi^{-1}(x) = (x, 1)$.

Let l_2 be a second line which misses the origin, with a choice of base points $\mathbf{v}_2, \mathbf{w}_2$. We define A to be the unique real (2×2) matrix such that $\mathbf{v}_2 = A(\mathbf{v}_1)$ and $\mathbf{w}_2 = A(\mathbf{w}_1)$. Here A acts by multiplying column vectors on the left. Since A is invertible it has nonzero determinant. We write $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

Note that $A : l_1 \to l_2$ is an (origin- and orientation-preserving) isometry, by linearity plus the definition of metric on the two lines.

We consider the map from l_1 to itself defined by sending l_1 to l_2 via A, followed by central projection from l_2 to l_1. We denote by f_A the map of $\widehat{\mathbb{R}}$ induced from this via the identification of l_1 with $\widehat{\mathbb{R}}$. That is, $\widehat{f_A} = \varphi \circ P_{l_1,l_2} \circ A \circ \varphi^{-1}$. We claim this is exactly the Möbius transformation f_A of $\widehat{\mathbb{R}}$ defined as in diagram (130).

Now as noted, A and φ are isometries; therefore the map P_{l_1,l_2} from $l_2(\mathbf{v}_2, \mathbf{w}_2)$ to $l_1(\mathbf{v}_1, \mathbf{w}_1)$ is isometrically conjugate to f_A, and

$$\widehat{f_A}(x) = \varphi \circ P_{l_1,l_2} \circ A \circ \varphi^{-1}(x) = \varphi \circ P_{l_1,l_2} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix} = \frac{ax+b}{cx+d} = f_A(x)$$

This handles the case of central projection from a general line with a pair of base points to the specific line l_1 with base points $(0, 1)$ and $(1, 1)$.

Next suppose we have two general lines l_2, l_3 with arbitrary chosen base points. But by the lemma, the central projection from l_2 to l_3 is $P_{l_3,l_2} = P_{l_3,l_1} \circ P_{l_1,l_2} = P^{-1}_{l_1,l_3} \circ P_{l_1,l_2}$ and each of these is Möbius. So we are done: any central projection is Möbius.

We note that changing the base points on a line $l(\mathbf{v}.\mathbf{w})$ is conjugate to a composition of a translation, multiplication and possibly an inversion of $\mathbb{R}$.

For the converse, begin with a real (2×2) matrix A with nonzero determinant, and reverse the above procedure: for l_2 we take the line $A(l_1)$; we define the base points to be the images $\mathbf{v}_2 = A(\mathbf{v}_1)$ and $\mathbf{w}_2 = A(\mathbf{w}_1)$. Since A is invertible, l_2 also misses the origin and these image points are also distinct.

The previous argument now shows that $P_{l_1,l_2} : l_2(\mathbf{v}_2, \mathbf{w}_2) \to l_1(\mathbf{v}_1, \mathbf{w}_1)$ is isometrically conjugate to the Möbius transformation f_A, finishing the proof. □

For an example, note that the central projection from the line $\tilde{l}$: $x = 1$ with base points $(1, 0), (1, 1)$ to l, the line $y = 1$, with base points $(0, 1), (1, 1)$, is inversion $x \mapsto 1/x$, and indeed the above construction gives us the matrix $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, whose Möbius transformation f_A is inversion. Indeed, central projection sends $(-1, 1)$, the point corresponding to -1 in $\widetilde{l}$, to $(1, -1)$, which corresponds to -1 in l; $(1, 1)$ is fixed by central projection, and that corresponds to 1 in both lines.

Remark 23.7. Above we said that given two lines inside $\mathbb{R}^2$, then when considered as extended real lines, their points at infinity should be equal iff the lines are parallel. To understand this precisely we need a projective space of one higher dimension, the projective plane RP^2, in addition to the projective line RP^1. In fact, we shall model RP^1 as lines through the origin in an extended plane which is a model for RP^2: the plane $z = 1$ sitting inside $\mathbb{R}^3$, compactified by a circle at infinity.

To explain this, recall that $RP^2 = P(\mathbb{R}^3)$, the space of lines through the origin. We want to understand lines as well as points in this space, and for this we consider several models.

First is the unit sphere S^2 in $\mathbb{R}^3$ with antipodal points identified, writtten $S^2/\sim$. The second is the upper hemisphere, with antipodal points identified on its circle boundary. The third is the plane $z = 1$, with a circle added at infinity: for instance the

unit circle in the plane $z = 0$ with antipodal points identified, as for the hemisphere model.

This describes the points in RP^2 and the topology. Now a *line* in RP^2 is by definition a *plane* through $\mathbf{0}$ in $\mathbb{R}^3$ modulo the projective equivalence relation $\sim$, where $\mathbf{v} \sim \lambda\mathbf{v}$ for $\lambda \neq 0$. Let us see what this gives in each of our models.

A plane through the origin meets S^2 in a great circle, so these are the lines in $S^2/\sim$. Note that any two distinct points determine a line and that any two distinct lines meet in a single point.

In the hemisphere model, a line is either half of a great circle with endpoints identified, or the equator with antipodal points identified. In the plane model, a line is either a straight Euclidean line or the circle at infinity c.

Note that in the first case a line l meets c in one point. Note also that two distinct Euclidean parallel lines l_1, l_2 in the plane meet in one point at infinity. Thus if we denote the point at infinity of l by $l_\infty = l \cap c$, then $(l_1)_\infty = (l_2)_\infty$.

Now we consider any two distinct extended lines l_1, l_2 in the Euclidean plane $\mathbb{R}^2$. These correspond to two lines (but not the line at infinity c) in the plane model for RP^2. Then l_1, l_2 meet in a unique point, and this is their point at infinity iff they are parallel.

Next we consider a line l in the plane model of RP^2 such that $l \neq c$ (the line at infinity) and also l does not pass through $\mathbf{0}$.

Then this extended line is in bijective correspondence with the collection of lines through the origin in this plane, since each such line m meets l in a unique point. Note that if m is the unique line through the origin parallel to l then this is the point at infinity of l.

Lastly we consider the central projection $P_{l_1 l_2}$ from l_2 to l_1, where these lines do not pass through the origin, from this viewpoint. Since each l_i is in bijective correspondence with RP^1, the map $P_{l_1 l_2}$ simply sends a point in the extended line l_2 to the corresponding point in l_1.

In conclusion, our lines l_1, l_2, which do not pass through the origin of $\mathbb{R}^2$, should be thought of as lines in a specific model of RP^2, that is the (extended) plane model $z = 1$.

Regarding RP^2, a nice observation is made on p. 18 of Thurston [Thu97]: if we identify opposite sides of the unit square, preserving orientation of the boundary segments as we do so, we of course get the torus. If we identify two opposite sides in this way, but change the direction of one of the other two segments, we get the Klein bottle. If, now, we change the other also, we have the projective plane. See Remark 23.1. In all cases, geodesics are (locally) straight lines. The last two are nonorientable- follow a frame (a pair of orthonormal vectors) along a geodesic!

23.3. **The hyperbolic and Hilbert metrics.** We begin with the Euclidean metric. Let us say that a metric $d(\cdot, \cdot)$ on $\mathbb{R}$ is **additive** if for $x < y < z$, the triangle inequality is *exact*, i.e. if $d(x, z) = d(x, y) + d(y, z)$. Then:

Lemma 23.12. *The Euclidean metric $d(x, y) = |x - y|$ is the unique additive metric on $\mathbb{R}$ which is translation-invariant, up to change of scale. It is also invariant for the additive inversion $x \mapsto -x$.*

Proof. The metric is nonzero since otherwise it would be a pseudometric. Since it is additive this extends to a measure on the Borel sets. But there is a unique translation-invariant measure on $\mathbb{R}$, up to multiplication by a constant. □

Exercise 23.1. Find a *non*additive translation-invariant metric on $\mathbb{R}$.
(Hint: consider a helix embedded in $\mathbb{R}^3$!)

Now we transport this metric to $\mathbb{R}^{>0} \equiv (0, +\infty)$ via the exponential map, defining $d_{0,+\infty}(x, y) = d(\log x, \log y) = |\log x - \log y|$. This defines (up to a constant multiple) the **hyperbolic metric** on a half-line.

Next, given an open segment $(\alpha, \beta) \subseteq \mathbb{R}$, we map this to the half-line via a Möbius transformation f, which sends α, x, β to $0, 1, \infty$. Hence f is defined by the cross-ratio: $f(y) = [\alpha, x, y, \beta]$ and so our metric is:

$$d_{\alpha,\beta}(x, y) = |\log[\alpha, x, y, \beta]|. \tag{74}$$

Given a circular arc γ in $\widehat{C}$ with endpoints α, β (so possibly a straight line segment) we define $d_{\gamma,\alpha,\beta}$ on γ by the same formula (74).

Proposition 23.13. *The hyperbolic metric $d_{0,+\infty}$ on $\mathbb{R}> 0$ is the unique (up to multiplication by a constant) additive metric which is dilation-invariant; it is also invariant for the multiplicative inversion $x \mapsto 1/x$. The hyperbolic metric $d_{\alpha,\beta}$ on $(\alpha, \beta) \subseteq \mathbb{R}$ and more generally on the circular arc $\gamma \subseteq \widehat{\mathbb{C}}$ is the unique (up to multiplication by a constant) additive metric which is invariant for the group of Möbius transformations which preserve this arc.*

Proof. Both statements are a consequence of the lemma, since the group of Möbius transformations which preserve $\mathbb{R}> 0$ is generated by the dilations (those which fix the endpoints) plus inversion (which interchanges them). □

Note that formula (74) works also for the first case $\alpha = 0, \beta = +\infty$, and that a half-line in $\widehat{\mathbb{C}}$ is a circular arc in the Riemann sphere.

Remark 23.8. Multiplying the metric by a constant corresponds to taking the logarithm with respect to a different base. That is, for $a > 0$, then for $s = e^{1/a}$, then $a \cdot d_{\alpha,\beta}(x, y) = a|\log[\alpha, x, y, \beta]| = |\log_s[\alpha, x, y, \beta]|$ since $\log_s(t) = \log t / \log s$.

Now we define the hyperbolic metric on half-space $\mathbb{H}$. We should really say "a" hyperbolic metric since the definition could differ from this by a constant multiple, as above; on a segment this doesn't make much difference, but in higher dimensions it does: for instance for the Poincaré disk, taking a multiple will change the curvature constant.

We need:

Lemma 23.14. *Given w, z distinct points in the interior of $\mathbb{H}$, there is a unique half-line or circular arc containing them which meets $\partial\mathbb{H}$ orthogonally.*

Proof. If w, z lie on the imaginary axis in $\mathbb{H}$ then this is clear; for any other choice we move them to this position via a Möbius transformation, and we know the line is taken to a circle, which meets the boundary orthogonally, since conformal maps preserve angles. Alternatively, the center of the circle can be constructed with compass and

straightedge: this is the point where the bisector of the segment from w to z meets the $x-$axis. □

Definition 23.3. For $\mathbf{x}.\mathbf{y} \in \mathbb{H}$, we define $d_{\mathbb{H}}(\mathbf{x}, \mathbf{y}) = c \cdot d_\gamma(\mathbf{x}, \mathbf{y})$ where γ is the unique circle or vertical half-line which passes through $\mathbf{x}$ and $\mathbf{y}$ and meets the boundary $\mathbb{R}$ orthogonally. This is the **hyperbolic metric** on $\mathbb{H}$ determined by the choice of c. Taking $c = 1/2$ is the traditional choice for $\mathbb{H}$, as it gives constant curvature -1. This choice is called the *Poincaré metric,*. Then $\mathbb{H}$ together with this metric is known as the *upper half space model* of the *hyperbolic plane*.

The usual approach to this metric encountered in the literature is via the infinitesimal formula for arc length

$$\mathrm{d}s^2 = \frac{\mathrm{d}x^2 + \mathrm{d}y^2}{y^2}$$

which means the following: $\mathrm{d}s^2 = (\mathrm{d}s)^2$, where $\mathrm{d}s$ is the *line element*, i.e. it gives the (infiinitesimal) length of a tangent vector to which it is applied. This means that we calculate the length of a curve $\gamma : [a, b] \to \mathbb{H}$ by

$$l(\gamma) = \int_\gamma \mathrm{d}s \equiv \int_a^b ||\gamma'(t)|| \mathrm{d}t.$$

Note that if we change this to the Euclidean line element

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2$$

then this formula gives the usual arc length.

Now let us for example consider the curve along the $y-$ axis in $\mathbb{R}^2$ $\gamma(t) = (0, t)$ for $t \in [a, b]$. Then for the previous formula,

$$\int_\gamma \mathrm{d}s = \int_a^b \frac{1}{t} \mathrm{d}t = \log(b) - \log(a)$$

which is the same as $\mathrm{d}_{\mathbb{H}}(ai, bi)$.

Next, let us consider the curve $\eta(t) = (t, y)$ for $t \in [a, b]$. Then

$$\int_\eta \mathrm{d}s = \int_a^b t \frac{1}{y} \mathrm{d}t = (b - a)/y.$$

Writing $y = e^s$, thus parametrizing the $y-$axis by arc length in the Poincaré metric, we have

$$l(\eta) = e^{-s}(b - a)$$

.

See Fig. **??**.

One shows in differential geometry that any such infinitesimal formula, given by a Riemannian metric on a manifold, i.e. a smoothly varying inner product on each tangent space, also called the first fundamental form, gives an arc length and thence a metric (defined to be the infimum of the curve lengths between two points), since the triangle inequality is automatically satisfied. The work then comes in showing that this locally equals arc length along geodesics, the existence of which has to be proved. We take a different approach below, by measuring distance along circular arcs, which

will be the geodesics. We then first prove the triangle inequality in Klein model, in much more general circumstances, and then relate this to the Poincaré model.

There is also a hyperbolic metric for 3−dimensional half-space $\mathbb{H}^3 \subseteq \mathbb{R}^3$, defined as follows. There is a unique plane passing through $\mathbf{x}$ and $\mathbf{y}$ which is perpendicular to the boundary plane, and restricting to this plane there is as before a unique such circle, and the definition is identical. A similar definition can be given for $\mathbb{H}^n$.

Let B be the open unit ball in $\mathbb{R}^d$ with $n \geq 2$. As above for the case $n = 2$, two distinct points $\mathbf{x}, \mathbf{y} \in B$ determine a unique circle which meets ∂B perpendicularly. Taking this arc as γ, we define $d_P(\mathbf{x}, \mathbf{y}) = c \cdot d_\gamma(\mathbf{x}, \mathbf{y})$. This is a hyperbolic metric on the **Poincaré ball**; for $B \subseteq \mathbb{R}^2$, this is isometrically taken to $\mathbb{H}$ by a Möbius transformation, defining the **Poincaré model** for the hyperbolic plane.

Now we move on to describe the Hilbert metric on a convex set. Let X be a vector space (or more generally an affine space) and let $C \subset X$ be a convex subset which satisfies this property:

(no-line property): C contains no complete copies of the real line. (75)

For instance, if X is Euclidean space, C could be compact convex, or could be an unbounded set such as $\{(x, y) : xy \geq 1; x, y > 0\}$. Thus given two distinct points $\mathbf{x}, \mathbf{y} \in C$, the line which passes through $\mathbf{x}$ and $\mathbf{y}$ meets C in a line segment l which is either an interval or a half-line denoted γ (the endpoints may be included or not). Then if neither $\mathbf{x}$ nor $\mathbf{y}$ is an endpoint we define $d_C(\mathbf{x}, \mathbf{y}) = d_\gamma(\mathbf{x}, \mathbf{y})$ to be the hyperbolic distance on that segment or half-line; if $\mathbf{x} = \mathbf{y}$ we set $d_C(\mathbf{x}, \mathbf{y}) = 0$; if one is an endpoint and the other not then we define $d_C(\mathbf{x}, \mathbf{y}) = \infty$. We call d_C the **Hilbert metric**. Note that reflexivity and symmetry, being properties that depend only on two points hence only on the segment γ, is automatic since we already know that d_γ is a metric. So to verify that d_C is indeed a metric, we need only check:

Proposition 23.15. *The triangle inequality holds for d_C.*

Proof. There is a beautiful proof (mostly geometric) in de la Harpe's article [DLH93]; our argument is inspired by that diagram and is purely geometrical.

Since the triangle inequality refers to three points $\mathbf{x}, \mathbf{y}, \mathbf{z}$, all of the proof reduces to the convex region in the plane containing these points which meets C. Everthing reduces further to the convex hexagon determined by where the line segments extending the triangle meet the boundary of C, see Fig. ??.

The proof can be motivated by a geometrical proof of the triangle inequality is the Euclidean plane. Consider a triangle with sides a, b, c of length $|a|, |b|, |c|$. We place the center of a compass at the endpoints of side c, and mark on that side intervals a', b' also of lengths $|a|, |b|$. These intervals overlap, hence $|a| + |b| \geq |c|$.

We do the same here except replace our compass with central projection, given by two different central points.

We shall use (two different) central projections to project sides a, b to the line containing side c.

To project side a we choose as central point the point where two lines intersect: those that are determined by the endpoints of the segments s_a and s_c containing the two involved sides, a and c.

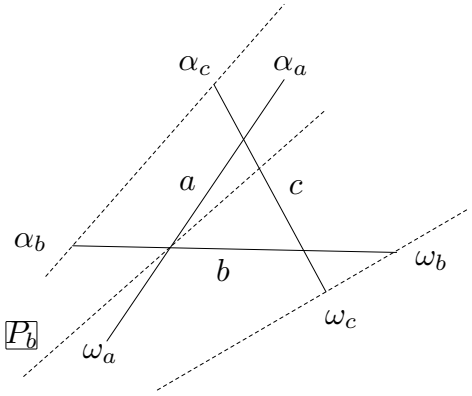


FIGURE 51. Proof of the triangle inequality for the Hilbert metric, showing how side b is central-projected to a subset b' of side c.

Thus for instance writing α_a, ω_a for the endpoints of the segment s_a, then projection P_a sends α_a and ω_a to α_c and ω_c.

That is, the endpoints of s_a are taken to the endpoints of s_c.

We write a' for the image of a; this is a subinterval of segment s_c which has one endpoint in common with c: where the two sides a and c meet. We call this the fixed endpoint of side a, as it is fixed by the projection.

We write d_a, d_b, d_c for the hyperbolic metrics on the open line segments s_a, s_b, s_c.

Now we know from Prop. 23.11 that a central projection induces a Möbius transformation of the two lines (defined up to base points, but change of base points is also a Möbius transformation). Hence by Prop. 23.5 this preserves cross-ratios. Since it maps the endpoints of s_a to those of s_c, it therefore is an isomtery from s_a with metric d_a to s_c with metric d_c. In particular, a and a' have the same length relative to these hyperbolic metrics.

We do the same for b, producing interval $b' \subseteq c$ and which includes the other endpoint of c as fixed point. The non-fixed endpoints of a', b' are the images of the third vertex of the triangle, by the two different projections.

We claim that the two segments overlap, which will prove that $|a|+|b| = |a'|+|b'| \geq |c|$ as desired. Here $|a|$, $|b|$ denote the d_a, d_b-lengths of a and b, while $|a'|$, $|b'|$ denotes the d_c-length of a', b'.

To show they overlap, consider the effect on the projection a', b' of moving one endpoint in the hexagon, corresponding to side c, farther in along the same line. Moving in α_c changes the non-fixed endpoint of a', shortening that segment, and has the same effect on b'. The same is true when we move ω_c. This shortens the interval where they overlap.

By moving both, finally a moment is reached where the overlap is a single point. This is the worst-case senario, as now the two central projection points have coalesced. The reason is that the hexagon has become a quadrilateral; the points α_c, ω_c are no longer extreme points.

And here is where we use the convexity: moving beyond this would destroy that hypothesis. See Fig. 51. □

The projective metric on a convex cone. Next we present a more abstract approach to the same metric.

Let V be a real vector space (infinite dimension is permitted). A subset C is a **cone** iff $\alpha C \subseteq C$ for all $\alpha \geq 0$. A cone is **positive** if $C \cap -C = \{\mathbf{0}\}$. It is **convex** if $C + C \subseteq C$. An **affine line** in V is $\{\mathbf{x} + t\mathbf{y} : \ t \in \mathbb{R}\}$ for $\mathbf{x}, \mathbf{y} \in V$ with $\mathbf{y} \neq 0$. We will say a cone C satisfies **Furstenberg's condition** if no affine line in V is completely contained in C.

Proposition 23.16.
(i) A cone is positive iff it contains no complete lines which pass through the origin.
(ii)Furstenberg's condition implies positivity.
(iii)It is a convex cone iff it is a cone which is a convex set, and equivalently,
(iv) iff it is a cone and for any affine subspace H, then $H \cap C$ is a convex set.
(v)For a convex cone, Furstenberg's condition is eqivalent to: for any affine subspace H, then the convex set $H \cap C$ satisfies the no-line property (75).

Proof. Part (i) follows directly from the definitions. For (ii), Furstenberg's condition implies in particular that C contains no complete lines which pass *through the origin*, which is positivity. Part (iii) is clear. For part (iv), if C is convex and $\mathbf{v}, \mathbf{w} \in H \cap C$, then for $p, q \geq 0$ with $p + q = 1$, then $p\mathbf{v} + q\mathbf{w} \in H$ while also $p\mathbf{v}, q\mathbf{w} \in C$ whence $p\mathbf{v} + q\mathbf{w} \in C$. Conversely, let $\mathbf{v}, \mathbf{w} \in C$, and define H to be some affine subspace containing these points; we are assuming that $H \cap C$ is convex. Then taking $p = q = 1/2$, $p\mathbf{v} + q\mathbf{w} \in H \cap C$ whence $\mathbf{v} + \mathbf{w} = 2 \cdot (p\mathbf{v} + q\mathbf{w}) \in C$. For (v), assuming Furstenberg's condition, then $H \cap C$ is a subset of C so can contain no complete lines. Conversely, if the no-line property holds for each $H \cap C$, then we can take for H in particular any affine line; $H \cap C$ cannot contain a complete line so H cannot be completely contained in C, proving Furstenberg's condition. □

Definition 23.4. Given a positive convex cone $C \subseteq V$, we say a vector $\mathbf{x} \in V$ is **positive** iff $\mathbf{x} \in C$. For $\mathbf{x}, \mathbf{y} \in V$, we define $\mathbf{x} \leq \mathbf{y}$ iff $(\mathbf{y} - \mathbf{x})$ is positive.

Proposition 23.17. *This defines a partial order on V. We have $\mathbf{x} \leq \mathbf{y}$ iff $\mathbf{y} \in \mathbf{x} + C$.*

Proof. The second statement is clear. The properties reflexivity $\mathbf{x} \leq \mathbf{x}$, symmetry $\mathbf{x} \leq \mathbf{y}$ and $\mathbf{y} \leq \mathbf{x} \implies \mathbf{y} = \mathbf{x}$, and transitivity $(\mathbf{x} \leq \mathbf{y},\ \mathbf{y} \leq \mathbf{z}) \implies (\mathbf{x} \leq \mathbf{z})$ follow respectively from the cone property, positivity and convexity. □

Definition 23.5. Given vector spaces V, W containing positive convex cones C, D, a linear transformation $f : V \to W$ is **positive** iff $f(C) \subseteq D$.

One has immediately:

Proposition 23.18. *A linear transformation is positive iff it preserves the partial order:* $(\mathbf{x} \leq \mathbf{y}) \implies (f(\mathbf{x}) \leq f(\mathbf{y}))$. □

Given C a positive convex cone and $\mathbf{x}, \mathbf{y} \in C \setminus \{\mathbf{0}\}$, we define two numbers

$$\alpha(\mathbf{x}, \mathbf{y}) = \sup\{\alpha \in \mathbb{R} : \alpha\mathbf{y} \leq \mathbf{x}\}$$

and

$$\beta(\mathbf{x}, \mathbf{y}) = \inf\{\beta \geq 0 : \mathbf{x} \leq \beta\mathbf{y}\}.$$

Thus $0 \leq \alpha(\mathbf{x}, \mathbf{y}) \leq \beta(\mathbf{x}, \mathbf{y}) \leq +\infty$.

We then define the **projective metric** on $C \setminus \{\mathbf{0}\}$ by:

$$d_C(\mathbf{x}, \mathbf{y}) = \log(\beta/\alpha). \tag{76}$$

We define the equivalence relation $\sim$ on $C \setminus \{\mathbf{0}\}$ by $\mathbf{x} \sim \lambda\mathbf{x}$ for $\lambda > 0$, and write $P(C) = (C \setminus \{\mathbf{0}\})/\sim$, the projective space of the cone.

Proposition 23.19. *d_C is a pseudometric on $P(C)$. It is a metric iff the cone satisfies Furstenberg's condition.*

Proof. Note first that it is well-defined on $P(C)$, as for $\mathbf{x}, \mathbf{y} \in C \setminus \{\mathbf{0}\}$, then given some $\lambda > 0$, $d_C(\mathbf{x}, \mathbf{y}) = d_C(\lambda\mathbf{x}, \mathbf{y})$ since both $\alpha(\mathbf{x}, \mathbf{y})$ and $\beta(\mathbf{x}, \mathbf{y})$ are multiplied by the same constant.

We next check that $d_C(\mathbf{x}, \mathbf{y}) = d_C(\mathbf{y}, \mathbf{x})$: we have $\alpha\mathbf{y} \leq \mathbf{x} \leq \beta\mathbf{y}$ so $(1/\beta)\mathbf{x} \leq \mathbf{y} \leq (1/\alpha)\mathbf{x}$; since $\alpha^{-1}/\beta^{-1} = \beta/\alpha$ we are done.

To show we have a pseudometric, it remains to verify the triangle inequality. Given three vectors $\mathbf{x}, \mathbf{y}, \mathbf{z} \in C$, let us write α_1 for $\alpha(\mathbf{x}, \mathbf{y})$, α_2 for $\alpha(\mathbf{y}, \mathbf{z})$ and α_3 for $\alpha(\mathbf{x}, \mathbf{z})$, and similarly for β. We have

$$\alpha_1\mathbf{y} \leq \mathbf{x} \leq \beta_1\mathbf{y}$$

and

$$\alpha_2\mathbf{z} \leq \mathbf{y} \leq \beta_2\mathbf{z}$$

and so

$$\alpha_1\alpha_2\mathbf{z} \leq \alpha_1\mathbf{y} \leq \mathbf{x} \leq \beta_1\mathbf{y} \leq \beta_1\beta_2\mathbf{z}$$

and therefore $\alpha_3 \geq \alpha_1\alpha_2$, $\beta_3 \leq \beta_1\beta_2$ and so

$$d_C(\mathbf{x}, \mathbf{z}) = \log(\beta_3/\alpha_3) \leq \log(\beta_1\beta_2/\alpha_1\alpha_2) = d_C(\mathbf{x}, \mathbf{y}) + d_C(\mathbf{y}, \mathbf{z}).$$

Next, with Furstenberg's condition, we prove it is a metric. Here we will follow Furstenberg's wonderful little proof of Lemma 15.1(*iii*) in [Fur61]. Assume that $d_C(\mathbf{x}, \mathbf{y}) = 0$. Then we have $\alpha(\mathbf{x}, \mathbf{y})$ and $\beta(\mathbf{x}, \mathbf{y})$, the sup and inf of numbers with $\alpha\mathbf{y} \leq \mathbf{x} \leq \beta\mathbf{y}$, and since $\log(\beta(\mathbf{x}, \mathbf{y})/\alpha(\mathbf{x}, \mathbf{y})) = 0$ we have $\alpha(\mathbf{x}, \mathbf{y}) = \beta(\mathbf{x}, \mathbf{y})$. Defining $\mathbf{w} = \alpha\mathbf{y}$, we have for each $\varepsilon > 0$ that (using the cone property)

$$(1 - \varepsilon)\mathbf{w} \leq \mathbf{y} \leq (1 + \varepsilon)\mathbf{w}. \tag{77}$$

Therefore, $\mathbf{y} - (1 - \varepsilon)\mathbf{w} \in C$ and so

$$(\mathbf{y} - \mathbf{w}) + \varepsilon\mathbf{w} \in C.$$

Thus

$$\frac{1}{\varepsilon}(\mathbf{y} - \mathbf{w}) + \mathbf{w} \in C,$$

for all $\varepsilon > 0$.

In other words,

$$t(\mathbf{y} - \mathbf{w}) + \mathbf{w} \in C,$$

for all $t > 0$. Using the right-hand side of (77), we have that this also holds for all $t \leq 0$.

If $\mathbf{y} \neq \mathbf{w}$, this is an affine line. Thus, $\mathbf{y} = \mathbf{w}$.

To prove the converse, we shall show that if Furstenberg's condition fails, then there exist two points $\mathbf{x} \neq \mathbf{y}$ with $d_C(\mathbf{x}, \mathbf{y}) = 0$.

Let $\mathbf{x} \neq \mathbf{y}$ be in an affine line $l \subseteq C$. Consider the two-dimensional subspace S of V spanned by $\mathbf{x}, \mathbf{y}$, so $l \subseteq S$. There exists a linear map $T : S \to \mathbb{R}^2$ which sends $\mathbf{x}$ to $(1, 1)$, $\mathbf{y}$ to $(1, 0)$ and hence l to the line $x = 1$. So we assume we are in this situation, with these points and this line.

Now let $\varepsilon \in (0, 1)$. We claim that $\varepsilon\mathbf{y} \leq \mathbf{x}$. We have $\mathbf{x} - \varepsilon\mathbf{y} = (1, 1) - (\varepsilon, 0) = (1 - \varepsilon, 1) \in (1 - \varepsilon)l \subseteq C$.

Thus $\alpha(\mathbf{x}, \mathbf{y}) = 1$, and similarly $\beta(\mathbf{x}, \mathbf{y}) = 1$, whence $d_C(\mathbf{x}, \mathbf{y}) = 0$, so d_C is not a metric. □

Example 20. Let $V = \mathbb{R}^2$ and $C = \{(x, y) : x > 0\}$. Then C is a positive convex cone but does not satisfy Furstenberg's condition. We have just shown that this cone will give a pseudometric but not a metric. Note also that as in the proof just given, we do need here to use sup and inf in the definiton of α, β, essentially since the cone boundary (the $y-$axis) is not in C. In the literature to avoid this issue it is often assumed that C is a closed cone in a Banach space. Furstenberg's condition instead isolates exactly what is needed, averting the need for topological considerations.

In the next result, we see (easily) that positive mappings always give a weak contraction. The much stronger statement of Birkhoff waits until §24.1.

Proposition 23.20. *Let V be a vector space and $C \subseteq V$ a convex cone satisfiying Furstenberg's condition. Write d_C for the projective metric on C. Let $L : V \to V$ be a linear transformation. Then:*
(a) if $\mathbf{v} \leq \mathbf{w}$ in the C-partial order then $L(\mathbf{v}) \leq L(\mathbf{w})$ in the $L(C)$-order.
(b) if L is invertible, it is an isometry from d_C to $d_{L(C)}$. In any case it is a weak contraction, i.e.

$$d_{L(C)}(L(\mathbf{v}), L(\mathbf{w})) \leq d_C(\mathbf{v}, \mathbf{w}).$$

(c) Let L be a positive linear transformation. Then it is a weak contraction in the C-metric.

Proof.
(a) For $\mathbf{v}, \mathbf{w} \in V$, $\mathbf{v} \leq \mathbf{w}$ iff there is a $\mathbf{z} \in C$ such that $\mathbf{w} = \mathbf{v} + \mathbf{z}$, iff $L(\mathbf{w}) = L(\mathbf{v}) + L(\mathbf{z})$.
(b) By part (a), $\alpha\mathbf{v} \leq \mathbf{w} \leq \beta\mathbf{v}$ in the C-order iff $\alpha L(\mathbf{v}) \leq L(\mathbf{w}) \leq \beta L(\mathbf{v})$ in the $L(C)$-order. Hence distance is preserved. For L not necessarily invertible, the distance may become 0, but in any case we have a weak contraction from the C- to the $L(C)$-metric.
(c) Since $L(C) \subseteq C$, this distance can only decrease further, proving the claim. □

23.4. **From the projective to the Hilbert metric.** Here we make the connection with the Hilbert metric, making use of an alternate definition often encountered in the dynamics literature.

Proposition 23.21.
(i) Let V be a vector space and $C \subseteq V$ a convex cone satisfying Furstenberg's condition. Let H be an affine subspace of V. Then for $d_{H \cap C}$ the Hilbert metric on

$H \cap C$ and d_C the projective metric on rays in C, we have for $\mathbf{w}, \mathbf{v} \in H \cap C$, $d_{H\cap C}(\mathbf{v}, \mathbf{w}) = d_C(\mathbf{v}, \mathbf{w})$.
(ii) Conversely, let $C \subseteq V$ be a convex set satisfying the no-line property. Let us imbed V as an affine hyperspace not containing $\mathbf{0}$ in a vector space $\widehat{V}$. Write $\widehat{C}$ for the cone generated by C (the smallest cone containing C). Then the metrics d_C and $d_{\widehat{C}}$ are equal.
(iii) For the hyperbolic metric on an interval $J = [a, b] \subseteq \mathbb{R}$, the distance between two points x, y with $a \leq x \leq y \leq b$ is given by

$$d_J(x, y) = \log\left(\frac{L+M}{L} \cdot \frac{M+R}{R}\right),$$

where L, M, R are the lengths or the left, middle and right subintervals.

Proof. Note first that by Proposition 23.16, $H \cap C$ is a convex set which satisfies the no-line property. Therefore the metric $d_{H\cap C}$ is defined.

Consider two vectors $\mathbf{v} \neq \mathbf{w} \in H \cap C$. The plane which contains $\mathbf{v}, \mathbf{w}$ and $\mathbf{0}$ is depicted in Fig. 52. This meets C in a two-dimensional cone. The line containing $\mathbf{v}$ and $\mathbf{w}$ is not a complete line hence is either a half-line or a segment.

Without loss of generality, in our figure we draw the cone as the positive quadrant in $\mathbb{R}^2$, and can take $\mathbf{v}$ to the left of $\mathbf{w}$ on the segment. This segment is divided into three, with lengths equal to L, M, R (for left, middle, right).

We copy this segment isometrically to an interval $[x^*, y^*] \subseteq \mathbb{R}$ with points x, y corresponding to the vectors $\mathbf{v}, \mathbf{w}$; that is, such that $x^* \leq x \leq y \leq y^*$ with $x - x^* = L$, $y - x = M$, $y^* - y = R$, so $y^* - x^* = L + M + R$.

We shall use, twice, the fact from elementary geometry that three parallel lines cut two transverse segments proportionally. See Fig. 52.

First we consider three vertical lines: the $y-$axis and the lines passing through $\mathbf{w}$ and through $\mathbf{v}$. The line through $\mathbf{v}$ meets the ray through $\mathbf{w}$ at a point $\alpha\mathbf{w}$ where we define $\alpha = \alpha(\mathbf{v}, \mathbf{w}) = \sup\{\alpha : \alpha\mathbf{w} \leq \mathbf{v}\}$. Since these parallels divide the line segments of the vectors $\mathbf{v}$, $\mathbf{w}$ in proportional segments, we have that $\alpha/1 = L/(L+M)$. Next, we consider three horizontal lines, the $x-$axis and the lines through $\mathbf{w}$ and $\mathbf{v}$. This horizontal line through $\mathbf{v}$ meets the ray through $\mathbf{w}$ at a point $\beta\mathbf{w}$. We have $\beta = \beta(\mathbf{v}, \mathbf{w}) = \inf\{\beta : \mathbf{v} \leq \beta\mathbf{w}\}$. The three parallels now divide the segment $M \cup R$ and the segment of the vector $\beta\mathbf{w}$ proportionally, giving $\beta/1 = (M + R)/R$.

Now the distance $d_{H\cap C}(\mathbf{v}, \mathbf{w})$ is by definition

$$d_{H\cap C}(\mathbf{v}, \mathbf{w}) = \log([x^*, x, y, y^*] = \log\left(\frac{x^* - y}{x^* - x} \cdot \frac{x - y^*}{y - y^*}\right) =$$
$$\log\left(\frac{L+M}{L} \cdot \frac{M+R}{R}\right) = \log\frac{\beta}{\alpha} = d_C(\mathbf{v}, \mathbf{w}).$$

This proves parts (i) and (iii).

For (ii), note first that $\widehat{C}$ is indeed a positive convex cone. From Proposition 23.16, it satisfies Furstenberg's condition. Now we are in the situation of part (i), so the two metrics are equal. □

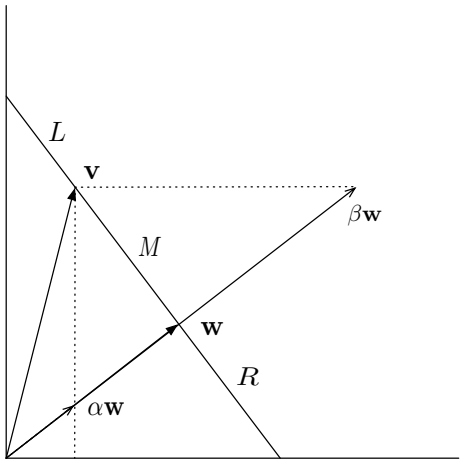


FIGURE 52. The Hilbert metric on an interval

Example 21. Consider $V = \mathbb{R}^d$ and C the usual positive cone

$$\mathbb{R}^{n+} = \{\mathbf{v} : v_i \geq 0, 1 \leq i \leq n\}.$$

Then:

Proposition 23.22.

$$d(\mathbf{v}, \mathbf{w}) = \left| \log \sup_{i,j} \frac{v_i w_j}{w_i v_j} \right|$$

Proof. $\alpha = \sup\{\tilde{\alpha} : \tilde{\alpha}\mathbf{v} \leq \mathbf{w}\} = \inf\{v_i/w_i\}$ and similarly, $\beta = \sup\{v_i/w_i\}$. Therefore,

$$\beta/\alpha = \sup_{i,j} \left\{ \frac{v_i}{w_i} / \frac{v_j}{w_j} \right\} = \sup_{i,j} \frac{v_i w_j}{w_i v_j}.$$

□

From this we see immediately the nice little fact that diagonal matrices act as isometries; we shall need this in the proof of Lemma 41.6: (to do: mixing conditions)

Corollary 23.23. *Let $\boldsymbol{z} = (z_1, \ldots z_n) \in \mathbb{R}^d$ with $z_i \neq 0$ for all i, and write Z for the diagonal matrix with entries $Z_{ii} = z_i$. Then $Z : \mathbb{R}^{n+} \to \mathbb{R}^{n+}$ is an isometry in the projective metric on $\mathbb{R}^{n+}$.*

Remark 23.9. We have included part (*iii*) here because this terminology is used sometimes in the literature; compare [dMvS93].

Furstenberg's condition seems to us to provide the most natural and general framework for defining projective metrics; it isolates what is necessary to have a metric while avoiding all mention of topology on V. Alternatively, one can for instance assume that V is a Banach space and closure of the cone is positive, or other conditions, all of which imply Furstenberg's condition. Compare the references cited below for a variety of approaches and much interesting additional information. [Bus73], [Sen81], [Woj86], [KP82], [Bir57], [Bir67]. [DLH93].

23.5. **An example: the ellipse, hemisphere and hyperboloid Klein models for hyperbolic n-space.** Let B be the open unit ball in $\mathbb{R}^n$ with $n \geq 2$. Above we defined the Poincaré metric $d_B(\mathbf{x}, \mathbf{y}) = c \cdot d_\gamma(\mathbf{x}, \mathbf{y})$ where γ is the unique circular arc which passes through $\mathbf{x}, \mathbf{y}$ and encounters the boundary ∂B orthogonally. With the constant $c = 1/2$, then this normalizes the curvature to be -1, which is the standard choice.

We consider the Hilbert metric d_H on B. Writing $d_K = c \cdot d_H$, then (B, d_K) is the **Klein model** for n-dimensional hyperbolic space.

This is also isometric to (c times) the Hilbert metric on an ellipsoid E.

With the help of the projective metric, we can describe other Klein-type models for two-dimensional hyperbolic space. Consider a standard solid cone $\mathbb{R}^3$, $C = \{(x, y, z) : x^2 + y^2 \leq z^2, z \geq 0\}$; choose a horizontal cross-section, say by the plane $z = 1$, giving a disk D. The cone C with the projective metric is isometric to the disk by central projection.

The same works for any elliptical conic section E. As we remarked earlier, the disk is the Klein model (times a constant); E straight lines for geodesics.

The isometry between the ellipsoid and ball Klein models is easy to describe: there exists an affine map from one to the other; this preverves line segments, and on each line segment gives a Möbius transformation to the image line segment, so preserves distance.

We note that while in D and E geodesics are segments in $\mathbb{R}^3$, in P each boundary point has exactly one direction where the geodesic emanating from it is a half-line (the vertical ray), all other angles giving segments, while for the hyperbola model there is a cone of such half-lines at every boundary point, parallel to the asymptotes.

We mention that the other conic sections (a parabola or a hyperbola) will not work here, as the projective map from the hyperboloid to those sections is not onto.

Now we know that the Poincaré and Klein models Δ, K for $\mathbb{H}^2$ have to be isometric, and it is a reasonable guess that a geodesic in Δ (a circle meeting the boundary orthogonally) will go to a geodesic in K (a Euclidean line) with the same endpoints. Nevertheless it seems hard to visualize this map. Thurston in [Thu97] gives a way to do this, which we describe.

Consider stereographic projection φ from the south pole S of the unit sphere $S^2 \subseteq \mathbb{R}^3$ to $\widehat{\mathbb{C}}$ embedded as the $xy-$plane. Consider a geodesic with endpoints ξ, η in the equatorial circle, the unit circle in this plane. This meets the circle at right angles, and the map φ is conformal, preserving angles and circles. Hence the image of this arc is a circular arc in S^2 which likewise meets the boundary perpendicularly.

This map, φ^{-1}, restricted to the disk gives a map from Δ to the upper hemisphere, with an inherited hyperbolic metric. This is called the *hemisphere model* for hyperbolic space. The geodesics are these circles which meet the boundary orthogonally, and hence are the intersection of the hemisphere with a vertical plane.

Next, we project downward to the disk via this vertical plane, giving a Euclidean line of the Klein model.

In summary, the map is the inverse of stereographic projection followed by vertical projection downwards, with the hemisphere model serving as the intermediary space!

FIGURE 53. Geodesics in the hemisphere, Klein and Poincaré models of the hyperbolic plane

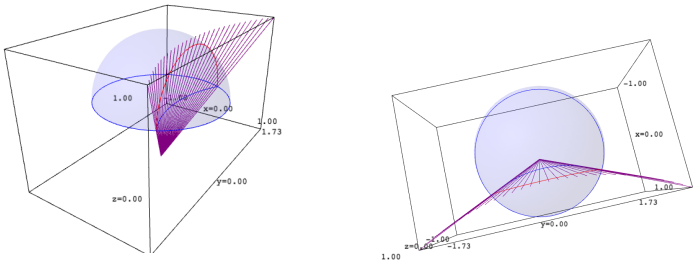


FIGURE 54. Stereographic projection from the hemisphere to the Poincaré model

See Fig. 2.12 of [Thu97]. This amazing book has much more on these remarkable spaces.

In particular he also discusses the hyperboloid model. Note that the metric on this model is that given by the projective metric on the cone, restricted to the hyperboloid, and same for the Klein model on a disk on any ellipse which is a planar section of the cone. This gives a direct isometry bteween the Klein and hyperboloid models, via the projective metric.

We remark that we have used here central projections twice, once to map the hemisphere model to the Poincaré model, and once to map the Klein model to the hyperboloid model. However, the metric in the first two cases is *not* that induced from the projective metric- otherwise Euclidean straight lines in the Poincaré disk would be geodesics, which they are not. The utility of the central projection in the first case is rather that since it is stereographic projection, it is conformal, so we can directly see what the geodesics are, as explained above.

Remark 23.10. The Hilbert metric is defined for any convex subset C of a vector space V which satisfies the no-line property. In the special case where V is finite dimensional and C is a ball (or more generally, as explained above, an ellipsoid), this is isometric to the Poincaré ball model and so to a Riemannian manifold. As noted above, this means a smooth manifold with a Riemannian metric, and by definition

that is an inner product on each tangent space, which varies smoothly. A more general notion in differential geometry is that of *Finsler metric*, where one has a smoothly varying *norm*. As Wojtkowski shows in [Woj83], in general, the Hilbert metric gives a Finsler but not a Riemannian metric; that is the case e.g. for the standard positive cone in $\mathbb{R}^d$.

The Hilbert metric or projective metric provides but one way of generalizing the notion of hyperbolic space. Another generalization is to assume a Riemannian metric but with variable negative curvature, and one can try to imagine further generalizations. Examples come e.g from the deeply insightful work of Gromov, Thurston and many others.

The Hilbert or projective metrics are especially natural and useful because of the connection with linear maps, while having a Riemannian metric of constant negative curvature as for the Poincaré model allows one to bring in all the powerful tools of Riemannian geometry and of Lie groups. See e.g. [Mas88], [Bea83], [Thu97].

?? elem pf of constant!!

24. A projective metric proof of the Perron-Frobenius theorem

Returning to the projective and Hilbert metrics, we next describe a beautiful approach to the Perron-Frobenius theorem due (independently, at about the same time) to Birkhoff and Samelson. We mention that Furstenberg, as a small part of his thesis and also at the same time, introduced related ideas. Birkhoff's first publication on this is dated 1957, [Bir57] while Furstenberg's thesis was submitted in 1958 (published by Princeton University Press in 1960). Samelson's paper is dated 1956 [Sam56]). All seem to be indpendent. The essential difference is that while Samelson used the Hilbert metric, Birkhoff used the projective metric, as did Furstenberg. Birkhoff gives a sharp bound on the contraction, which is especially useful when studying a sequence of matrices; see [Fis09]. And moreover, Birkhoff gets an upper boud of the contraction of the complementary subspace, that is, for the modulus of the other eigenvalues. Furstenberg also gives a boound; see §16.2 of [Fur60]; we have not yet checked to see how this compares with Birkhoff's result or presentation. Apparently all three approaches were developed independently. (Furstenberg's result represents a small part of his thesis.)

24.1. **Birkhoff's contraction estimate.** In [Bir57] and again, with more details, in §XVI of [Bir67], Birkhoff gives a sharp bound on the amount of contraction for a Möbius transformation, measured in the hyperbolic metric. His 1967 proof is a tour-de-force of elementary algebra and calculus, however a lot of the technical difficulty is due to the fact that he calculates for the general case. We manage to simplify this proof considerably by reducing first to the most symmetric possible case.

This result is stated in terms of the hyperbolic metric on half-lines and segments, from §23.3; to apply it, in Theorem 24.3, we shall then switch to the cone point of view.

Remark 24.1. Let us recall that

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

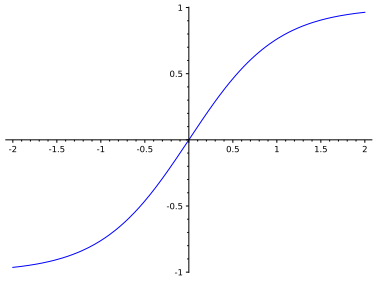


FIGURE 55. tanh(x)

This is an increasing odd function with value 0 at 0, $-1 \leq \tanh(x) < 1$ and with limits ± 1 at $\pm\infty$. We note that

$$\frac{e^{2x}-1}{e^{2x}+1} = \frac{t^2-1}{t^2+1} = \frac{\lambda-1}{\lambda+1},$$

for $\lambda = e^{2x}$. Thus any time we encounter a formula of these latter types, we have come across a hyperbolic tangent.

Lemma 24.1. *(Birkhoff* [Bir57]*) Let $d_{(0,\infty)}$ be the hyperbolic metric on $(0,\infty)$ and suppose f is a real Möbius transformation which maps $(0,\infty)$ inside itself. Then*

$$\sup_{x,y>0} \frac{d_{(0,\infty)}(f(x),f(y))}{d_{(0,\infty)}(x,y)} = \tanh(\Delta/4)$$

where Δ is the diameter of the image interval in the metric $d_{(0,\infty)}$.

For $f(x) = (ax+b)/(cx+d)$ this quantity is $\Delta = \log(ad/bc)$. The weakest contraction occurs at $x = (bd/ac)^{\frac{1}{2}}$, which is the hyperbolic midpoint of the image interval.

As we see below, since $\tanh(\Delta/4) < 1$, this will give us the contraction needed to prove a fixed point theorem!

Proof. We are given

$$f_A(x) = \frac{ax+b}{cx+d}$$

for the matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ with $a,b,c,d \in \mathbb{R}$ such that $0 \leq f(0) \leq f(\infty) \leq \infty$; here, $f(0) = b/d$ and $f(\infty) = a/c$ so indeed

$$\Delta = \log\left(\frac{a/c}{b/d}\right) = \log\left(\frac{ad}{bc}\right).$$

Setting

$$\lambda = \left(\frac{ad}{bc}\right)^{1/2},$$

then $\log(\lambda) = \Delta/2$ is the hyperbolic radius of the image interval.

We claim that there is a dilation of the reals, $D_\alpha(x) = \alpha x$, such that $D_\alpha \circ f_A = f_B$ has the form

$$f_B(x) = \frac{\lambda x + 1}{x + \lambda};$$

this corresponds to the matrix $B = \begin{bmatrix} \lambda & 1 \\ 1 & \lambda \end{bmatrix}$.

This resulting map is the most symmetric possible case as $f_B(0) = 1/\lambda, f_B(\infty) = \lambda$ and $f_B(1) = 1$; its hyperbolic midpoint is 1 and the intervals on either side of this point have length $\log \lambda$.

So to find this dilation, we let p be the hyperbolic midpoint of the image interval $(b/d, a/c)$ and take $\alpha = 1/p$.

To find the midpoint p we solve the equation

$$\frac{1}{p}\frac{b}{d} = \left(\frac{1}{p}\frac{a}{c}\right)^{-1}$$

and so

$$\frac{ab}{cd} = p^2$$

and

$$p = \left(\frac{ab}{cd}\right)^{1/2}.$$

Since the compostion $D_\alpha \circ f_A$ sends 0, 1 and ∞ to $1/\lambda, 1$ and λ, it must equal f_B.

The dilations are exactly the orientation-preserving isometries of $(0, \infty)$ (with the hyperbolic metric), so proving the theorem for this one-parameter family of maps will be enough. The dilations are, from another point of view, the translations in the multiplicative group of positive reals, $(\mathbb{R}^{>0}, \cdot)$. We note that the function f_B is an odd function on the multiplicative reals, that is: $f_B(x^{-1}) = (f_B(x))^{-1}$. This fact reflects the symmety of f_B.

Next we conjugate from the multiplicative reals to the additive reals, by the exponential $\exp : \mathbb{R} \to (0, \infty)$; this is an isometry from the Euclidean metric to the hyperbolic metric. We define $F : \mathbb{R} \to \mathbb{R}$ by $F = \log \circ f_B \circ \exp$.

Now note that since the multiplicative group identity 1 was fixed for f_B, the point 0 is a fixed point for F. Note that the image interval is $F(\mathbb{R}) = (-\log \lambda, \log \lambda)$, and that F is an odd function, $F(-x) = -F(x)$, since f_B is (multiplicatively).

Following Birkhoff, the reason for moving to the reals is that we can do our calculations in the Euclidean metric.

Since we have conjugated by isometries, our contraction constant for f is the supremum of $(F(y) - F(x))/(y - x)$. By the Mean Value Theorem, there exists $c \in [x, y]$ such that $(F(y) - F(x))/(y - x) = F'(c)$. So the contraction constant is equal to $\sup_{t \in \mathbb{R}} F'(t)$. The equality is what proves Birkhoff's bound is sharp.

We calculate F':

$$F(x) = (\log \circ f_B \circ \exp)(x) = \log\left(\frac{\lambda e^x + 1}{e^x + \lambda}\right)$$

so

$$F'(x) = \frac{\lambda e^x}{\lambda e^x + 1} - \frac{e^x}{e^x + \lambda} = \frac{(\lambda^2 - 1)e^x}{\lambda e^{2x} + (\lambda^2 + 1)e^x + \lambda} \tag{78}$$

Our bound is the maximum of this function $G(x) = F'(x)$, which is an even function since F is odd. If it has a single maximum, this must, by this symmetry, occur at the point $x = 0$.

To verify this fact we take the derivative:

$$G'(x) = F''(x) = \frac{(\lambda^2 - 1)e^x(-\lambda e^{2x} + \lambda)}{(\lambda e^{2x} + (\lambda^2 + 1)e^x + \lambda)^2}$$

We know that $\lambda^2 - 1 > 0$ so indeed the top equals zero only for $\lambda = \lambda e^{2x}$ 1.e. for $x = 0$.

Since $\lambda = e^{\Delta/2} > 0$, the value of $G = F'$ at this point is

$$\frac{\lambda^2 - 1}{\lambda^2 + 2\lambda + 1} = \frac{\lambda - 1}{\lambda + 1} \tag{79}$$

This is of the form in Remark 24.1; so this equals $\tanh(x)$ for $x = \frac{1}{2}\log\lambda$ and since $\log\lambda = \Delta/2$, this equals $\tanh(\Delta/4)$ as claimed. (I would love to see a purely geometrical explanation of this beautiful formula -no doubt complex analysts know how to do this!) □

Corollary 24.2. *Let l_1 be a subset of $\mathbb{R}$ which is a segment or half-line, with its hyperbolic metric d_1. Let l_2 be a subsegment with diameter $0 < \Delta < \infty$. Then*

$$\sup_{x,y\in l_2} \frac{d_1(x,y)}{d_2(x,y)} = \tanh(\Delta/4).$$

Proof. Write a_1, b_1 for the endpoints of l_1 and a_2, b_2 for the endpoints of l_2, with $a_1 < a_2 < b_2 < b_1$.

Choose a real Möbius transformation g with $g(a_2) = a_1$ and $g(b_2) = b_1$.

Then g is an isometry from l_2, d_2 to l_1, d_1. Write $f = g^{-1}$. Then for $x, y \in l_2$ and $\tilde{x} = g(x), \tilde{y} = g(y)$, we have

$$\frac{d_1(x,y)}{d_2(x,y)} = \frac{d_1(f(\tilde{x}), f(\tilde{y}))}{d_1(\tilde{x}, \tilde{y})}$$

and so by Lemma 24.1,

$$\sup_{x,y\in l_2} \frac{d_1(x,y)}{d_2(x,y)} = \sup_{\tilde{x},\tilde{y}\in l_1} \frac{d_1(f(\tilde{x}), f(\tilde{y}))}{d_1(\tilde{x}, \tilde{y})} = \tanh(\Delta/4).$$

□

We saw in Proposition 23.20 above that one always has a weak contraction, by a very easy proof. For strict contraction, all the hard work has been done in Lemma 24.1, and we now conclude:

Theorem 24.3. (Birkhoff's cone contraction estimate) *Let V, W be vector spaces and $C \subseteq V$, $D \subseteq W$ convex cones satisfiying Furstenberg's condition, with d_C, d_D the projective metrics on these cones. Let $L : V \to W$ be a positive linear*

transformation. Write Δ for the D- diameter of the image of C. Then if $\Delta < \infty$, L is a strict contraction, with coefficient

$$\sup_{\mathbf{v},\mathbf{w}\in C} \frac{d_D(L(\mathbf{v}), L(\mathbf{w}))}{d_C(\mathbf{v}, \mathbf{w})} = tanh(\Delta/4).$$

This bound is sharp.

Proof. Let $\mathbf{v}, \mathbf{w} \in C$ be be projectively distinct (i.e. they are linearly independent vectors), so $d_C(\mathbf{v}, \mathbf{w}) > 0$. Let l denote the line segment or a half-line in C determined by these two points. If $L(\mathbf{v})$ and $L(\mathbf{w})$ are not linearly independent, the distance is zero and the upper bound holds trivially. So assume that they are independent; then $l_2 \equiv L(l)$ is a line segment or half-line in $L(C)$ which is not projectively trivial. We embed each of these segments into a real line, by choice of an origin, unit length and positive direction.

Since L is a linear map, it preserves convex combinations of the endpoints $\mathbf{v}$ and $\mathbf{w}$ of l, so identifying the above real lines to $\mathbb{R}$ with the Euclidean metric $d(x, y) = |x-y|$, L induces a map of $\mathbb{R}$: a translation composed with a dilation by the factor $|l_2|/|l|$. This is a Möbius transformation of $\mathbb{R}$. Now give l its hyperbolic metric, and do the same for l_2; these are not affected by the choice of embeddings into $\mathbb{R}$. These metrics are defined from the Euclidean metric on the real line, as in (*iii*) of Proposition 23.21, so the restriction $L : l \to l_2$ is an isometry for these hyperbolic metrics. (An alternative argument is to note that this restriction is a Möbius transformation of the segments hence an isometry).

Now l_2 extends to a segment or half-line l_1 in the cone D, and $l_2 = l_1 \cap L(C)$.

The hyperbolic length Δ_{l_2} of l_2 as a subsegment of l_1 is bounded above by Δ, the diameter of $L(C)$ in D. Applying Corollary 24.2, we have

$$\sup_{x,y\in l_2} \frac{d_{l_1}(x, y)}{d_{l_2}(x, y)} = \tanh(\Delta_{l_2}/4).$$

Since L is an isometry from l to l_2, $d_C(\mathbf{v}, \mathbf{w}) = d_l(\mathbf{v}, \mathbf{w}) = d_{l_2}(L(\mathbf{v}), L(\mathbf{w}))$. And $d_D(L(\mathbf{v}), L(\mathbf{w})) = d_{l_1}(L(\mathbf{v}), L(\mathbf{w}))$. Thus

$$\sup_{\mathbf{v},\mathbf{w}\in l} \frac{d_D(L(\mathbf{v}), L(\mathbf{w}))}{d_C(\mathbf{v}, \mathbf{w})} = \tanh(\Delta_{l_2}/4) \leq \tanh(\Delta/4).$$

This proves the upper bound.

To show sharpness of the bound, consider a segment or half-line in D where the ratio of where it meets $L(C)$ is close to the diameter Δ; taking the inverse image, we apply the above argument and are arbitrarily close to that bound. □

We next recall:

Lemma 24.4. *Let (X, d) be a complete metric space and $f : X \to X$ a strict contraction, i.e. there exists $c \in [0, 1)$ such that $d(f(x), f(y)) \leq cd(x, y)$. Then f has a unique fixed point.*

Proof. Choose $x \in X$; the iterates $f^n(x)$ form a Cauchy sequence, since the contraction gives a geometric series, and this has a limit point which is the unique fixed point. □

The next statement is basically Theorem 1 of [Bir57] (we have replaced his condition "C is a bounded closed convex cone of a real Banach space" by the more general formulation of having Furstenberg's condition for the cone, in a real vector space; and "L is a bounded linear transformation" simply by L linear).

Theorem 24.5. *(Birkhoff's Perron-Frobenius Theorem) Let V be vector space and $C \subseteq V$ a convex cone satisfiying Furstenberg's condition, with d_C the projective metric on C. Write $\sim$ for the projective equivalence relation $\mathbf{v} \sim \lambda\mathbf{v}$ for $\lambda \neq 0$, defining the projective space $P(V) = V/\sim$. Assume that the metric space $(C/\sim, d_C)$ is complete. Let $L : V \to V$ be a positive linear transformation, and assume that for some power $m \geq 1$, the d_C- diameter of $L^m(C)$, $\Delta \equiv diam(L^m(C))$, is finite. Define $c = tanh(\Delta/4) < 1$ and $\rho = 1 - e^{-\Delta} < 1$.*
Then:
(i) there exists (up to multiplication by a positive constant) a unique positive eigenvector $\mathbf{w}$, with eigenvalue $\lambda > 0$.
(ii) All other nonnegative rays are attracted to this direction, exponentially fast in the projective metric: for all $\mathbf{v} \in C$, $d(L^{km}\mathbf{v}, \mathbf{w}) \leq c^k \to 0$.
(iii)This can be expressed as a contraction in the complementary subspace. Precisely, there exists a positive linear functional $M : V \to \mathbb{R}$ such that for every $\mathbf{f} \in C$, writing $M = M(\mathbf{f})$ and $\widetilde{L} \equiv L^m$, we have a constant $K = K(\mathbf{f})$ with

$$d(\widetilde{L}^k(\mathbf{f}), M \cdot \widetilde{L}^k(\mathbf{w})) \leq K(\mathbf{f})\rho^k.$$

Moreover, let E_p denote the one-dimensional space generated by $\mathbf{w}$, and let E_c denote the kernel of M. Then we have an $L-$ invariant splitting $V = E_p \oplus E_c$, such that for every $\mathbf{u} \in E_c$, $d(L^n\mathbf{u}, \mathbf{0})$ is nonincreasing, with

$$d(\widetilde{L}^k\mathbf{u}, \mathbf{0}) \leq \rho^k \to 0.$$

(iv) Let $\mathbf{v}$ be an eigenvector for L with eigenvalue $\mu \in \mathbb{C}$ and $\mathbf{v}$ not a multiple of $\mathbf{w}$. Then $|\mu| < \rho \cdot \lambda < \lambda$.

Proof. To prove (i), for L^m, from Birkhoff's estimate we have the (sharp) contraction coefficient $c = \tanh(\Delta/4) < 1$. Then by Lemma 24.4, there exists a unique fixed ray in the positive cone, and this gives the eigenvector $\mathbf{w}$. Since $L\mathbf{w} \in C$, its eigenvalue λ is positive, proving (i).

To prove (ii), we assume without loss of generality that $m = 1$, and also that $\lambda = 1$, otherwise replacing L by $(1/\lambda)L$. For the proof we largely follow §7 of [Bir57]. See also Theorem 8 of [Bir67].

Given $\mathbf{f} \in C$, for each $n \geq 1$, we set $\mathbf{f}_n = L^n\mathbf{f}$. Let a_n, b_n be the largest, smallest numbers such that

$$a_n\mathbf{w} \leq \mathbf{f}_n \leq b_n\mathbf{w}. \tag{80}$$

Applying L gives $a_n\mathbf{w} \leq \mathbf{f}_{n+1} \leq b_n\mathbf{w}$, whence $a_n \leq a_{n+1} \leq b_{n+1} \leq b_n$; we claim that $b_n - a_n \to 0$.

Since, for each $n \geq 1$, $C \supseteq L(C) \supseteq L^2(C) \supseteq \cdots \supseteq L^n(C)$ while (with our assumption $m = 1$) the diameter $L(C)$ is less than $\Delta < \infty$, we have $d(\mathbf{f}_n, \mathbf{w}) \leq \Delta$.

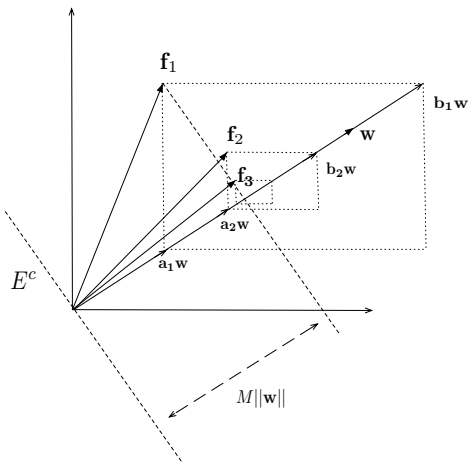


FIGURE 56. Convergence to the linear functional $M(\mathbf{f})$ (for the normalized case $\lambda = 1$)

From the definition of the projective metric (76),

$$0 \leq d(f_n, \mathbf{w}) = \log \frac{b_n}{a_n} \leq \Delta.$$

Thus

$$1 < b_n/a_n \leq e^{\Delta} < \infty \tag{81}$$

whence

$$0 < a_n < b_n \leq e^{\Delta} a_n < \infty \tag{82}$$

Defining for any $n \geq 0$ vectors

$$\begin{aligned} \mathbf{r}_n &= \mathbf{f}_n - a_n \mathbf{w}, \\ \mathbf{s}_n &= b_n \mathbf{w} - \mathbf{f}_n \end{aligned} \tag{83}$$

so $\mathbf{r}_n, \mathbf{s}_n \in C$, (these vectors represent the sides of the boxes in Fig. 56) then, letting α_n, α_n' be the greatest, least numbers such that

$$\alpha_n \mathbf{w} \leq L\mathbf{r}_n \leq \alpha_n' \mathbf{w},$$

and following the logic for the estimate (82), from (76) we have

$$d(L\mathbf{r}_n, \mathbf{w}) = \log \frac{\alpha_n'}{\alpha_n} \leq \Delta < \infty$$

whence

$$1 < \alpha_n'/\alpha_n \leq e^{\Delta} \tag{84}$$

and

$$0 < \alpha_n < e^{\Delta} \alpha_n' \leq e^{\Delta} \alpha_n < \infty \tag{85}$$

Similarly, defining $0 < \beta_n, \beta_n'$ to be the greatest, least numbers such that so that $\beta_n \mathbf{w} \leq L\mathbf{s}_n \leq \beta_n' \mathbf{w}$, we have in conclusion that

$$\begin{aligned} \alpha_n \mathbf{w} \leq L\mathbf{r}_n \leq e^{\Delta} \alpha_n \mathbf{w} \\ \beta_n \mathbf{w} \leq L\mathbf{s}_n \leq e^{\Delta} \beta_n \mathbf{w} \end{aligned} \tag{86}$$

Now from (83), $L(\mathbf{r}_n + \mathbf{s}_n) = (b_n - a_n)\mathbf{w}$ so

$$(b_n - a_n) \leq e^{\Delta}(\alpha_n + \beta_n) \tag{87}$$

which we shall need in a moment. Also from (83)

$$a_n\mathbf{w} + \mathbf{r}_n = \mathbf{f}_n = b_n\mathbf{w} - \mathbf{s}_n \tag{88}$$

whence (applying L)

$$a_n\mathbf{w} + L\mathbf{r}_n = \mathbf{f}_{n+1} = b_n\mathbf{w} - L\mathbf{s}_n. \tag{89}$$

From (86) therefore

$$(a_n + \alpha_n)\mathbf{w} \leq \mathbf{f}_{n+1} \leq (b_n - \beta_n)\mathbf{w} \tag{90}$$

Thus by (80) (with index $n+1$), $a_n + \alpha_n \leq a_{n+1} \leq b_{n+1} \leq b_n - \beta_n$. So by (87)

$$b_{n+1} - a_{n+1} \leq b_n - a_n - (\beta_n + \alpha_n) \leq (1 - e^{-\Delta})(b_n - a_n) \tag{91}$$

This is true for all $n \geq 1$. So

$$b_{n+1} - a_{n+1} \leq (1 - e^{-\Delta})^n(b_1 - a_1) \to 0$$

as claimed.

Now it follows that for any $f \in C$, there is a positive number $M = M(f)$ with $a_n \uparrow M$ and $b_n \downarrow M$. See Fig. 56.

We claim that this extends to a linear functional $M : V \to \mathbb{R}$. For the proof, write for $f \in \mathcal{V}$, $f = f^+ - f^-$ where $f^+, f^- \in C$. That is to say, $f^+ = \max(f, \mathbf{0})$ and $f^- = \min(f, \mathbf{0})$. This depends on the idea of *vector lattice*. See Def. 24.1 and e.g. [Wic02].

Definition 24.1. A vector space V with a positive cone C is a **vector lattice**: there is an operation $\mathbf{v} \wedge \mathbf{w}$ (min) and $\mathbf{v} \vee \mathbf{w}$ (max). And indeed if $\mathbf{v} \wedge \mathbf{w} = \mathbf{u}$ then $\mathbf{u}$ is the greatest element of V which is $\leq \mathbf{v}, \mathbf{w}$ while $\mathbf{v} \vee \mathbf{w} = \mathbf{u}$ then $\mathbf{u}$ is the least element of V which is $\geq \mathbf{v}, \mathbf{w}$. The *positive part* of $\mathbf{v}$ is $\mathbf{v}^+ = \mathbf{v} \vee \mathbf{0}$ and the *negative part* is $\mathbf{v}^- = -\mathbf{v} \wedge \mathbf{0}$. We note that $\mathbf{v} = \mathbf{v}^+ - \mathbf{v}^-$.

The rest of parts (*iii*) and (*iv*) then follow from this. □

Corollary 24.6. (Perron-Frobenius Theorem, Birkhoff-Samelson proof) *Let M be an $(d \times d)$ matrix with nonnegative real entries, such that M is primitive. Then there exists a unique positive right eigenvector; its eigenvalue λ is positive, and is greater in modulus than all the other eigenvalues.*

Proof. Our vector space is the column vectors of $\mathbb{R}^d$ and the cone C is the usual positive cone, i.e. the set of column vectors with nonnegative entries. This satisfies Furstenberg's condition, and is projectively complete. M primitive says that for some $m > 0$, M^m has entires all nonzero. The image $D = M^m(C)$ is contained in the interior of C hence because we are in a finite dimensional space, it has a finite projective diameter Δ. Now we apply the theorem. The eigenvalue is positive since the matrix M^m and vector v have positive entries, and is of greatest modulus as proved in the theorem. □

Example 22. As in Example 21 we consider the usual positive cone in Euclidean space, and have an explicit formula for Δ. This is due to Birkhoff for the case of a square matrix; the same statement and proof holds for the rectangular case, which we shall need below. We include a proof as §XVI of [Bir67] is not so easy to follow, due to some misprints and skipped steps.

Definition 24.2. In this situation, a positive linear map $L : \mathbb{R}^m \to \mathbb{R}^n$ with the **standard cones** $\mathbb{R}^{m+}, \mathbb{R}^{n+}$, we call the projective diameter of the image the **opening** of the linear map L, written $\Theta(L)$.

Remark 24.2. The reason we now switch from Birkhoff's notation Δ (for diameter) to Θ is that in the next section, Δ will be reserved for denoting the unit simplex. We choose the name **opening** of a map as suggestive of the aperture of a photographic shutter.

Proposition 24.7. *Let V, W be $\mathbb{R}^n$,$W = \mathbb{R}^m$ with their standard cones $C = \mathbb{R}^{n+}$, $D = \mathbb{R}^{m+}$. Let L be a $(m \times n)$ nonnegative matrix. Then the opening of the map L, the d_D- diameter of $L(C)$, is:*

$$\Theta(L) = \sup_{i,j,k,l} \left| \log \frac{L_{il} L_{jk}}{L_{jl} L_{ik}} \right|. \tag{92}$$

Proof. We know from Proposition 23.22 that

$$d_C(\mathbf{v}, \mathbf{w}) = \left| \log \sup_{i,j} \frac{v_i w_j}{w_i v_j} \right|.$$

Similarly, therefore, for $\mathbf{v}, \mathbf{w} \in C$,

$$d_D(L\mathbf{v}, L\mathbf{w}) = \left| \log \sup_{i,j} \frac{\sum_{l=1}^n L_{il} v_l \sum_{k=1}^n L_{jk} w_k}{\sum_{l=1}^n L_{jl} v_l \sum_{k=1}^n L_{ik} w_k} \right|.$$

Now

$$\sum_{l=1}^n L_{il} v_l \sum_{k=1}^n L_{jk} w_k = \sum_{l,k} L_{il} v_l L_{jk} w_k.$$

Hence

$$d_D(L\mathbf{v}, L\mathbf{w}) = \left| \log \sup_{i,j} \frac{\sum_{l,k=1}^n L_{il} v_l L_{jk} w_k}{\sum_{l,k=1}^n L_{jl} v_l L_{ik} w_k} \right|.$$

Defining $\delta_{\mathbf{i}} \in \mathbb{R}^n$ to be the vector with 1 in the i^{th} coordinate, 0 elsewhere, then if we take in the right hand side of this equation $\mathbf{v} = \delta_{\mathbf{l}}$ and $\mathbf{w} = \delta_{\mathbf{k}}$, we get for general $\mathbf{v}, \mathbf{w}$ that all but one term of each sum is zero so:

$$d_D(L\mathbf{v}, L\mathbf{w}) \geq \left| \log \sup_{i,j} \frac{L_{il} v_l L_{jk} w_k}{L_{jl} v_l L_{ik} w_k} \right|.$$

We claim the supremum of these is equal to the supremum of the sums. Quoting Birkhoff [Bir67], p. 383, this "obviously cannot be exceeded since averaging (by positive weight factors $v_l w_k$) always makes ratios less extreme". □

Indeed we have:

Lemma 24.8. *Let $a, b, c, d > 0$. Then*

$$\frac{a}{b} < \frac{c}{d}$$

implies that for $\alpha, \beta > 0$,

$$\frac{a}{b} < \frac{\alpha a + \beta c}{\alpha b + \beta d} < \frac{c}{d}.$$

Proof. Draw vectors (a, b) and (c, d) in the positive quadrant of the plane. Then $(\alpha a + \beta c, \alpha b + \beta d)$ is a positive linear combination of these, and by the parallelogram law for vector addition, projectively this vector is between the other two.

This proof extends to n vectors; consider the two most extreme ones. □

Remark 24.3. The quantity

$$\frac{L_{il} L_{jk}}{L_{jl} L_{ik}}$$

can be visualized as the ratio of the product of the opposite corners of a rectangle of entries in the matrix, where the i, j rows and l, k columns meet.

As a consequence, for the (2×2) case (as Birkhoff notes), the formula simplifies to

$$\Theta(L) = \left| \log \frac{L_{11} L_{22}}{L_{21} L_{12}} \right|,$$

since inversion of the ratio does not change the absolute value of the log, and since this is the only nontrival rectangle in the matrix, all others giving ratio 1, and so, distance 0.

Remark 24.4. We note that if all of the columns of the matrix are roughly proportional, then the opening is small, and conversely. This should be the intuition behind the formula.

24.2. **Contraction for the dual cones.** From Remark 24.3 we have this immediate consequence, which shall be important below:

Corollary 24.9. *Let L be a $(m \times n)$ nonnegative real matrix, and let $V = \mathbb{R}^n$,$W = \mathbb{R}^m$ and $C = \mathbb{R}^{n+}$, $D = \mathbb{R}^{m+}$ as in the proposition. Then the opening of L equals that of its transpose, i.e. $\Theta(L) = \Theta(L^t)$; that is:*

$$\Delta_D(L(C)) = \Delta_C(L^t(D)).$$

□

However the proof is special to $\mathbb{R}^n$ with the standard cones. We shall see in this section how this useful fact can indeed be extended to a quite general setting.

Remark 24.5. Remark on the complementary subspace. We claim that Birkhoff's linear functional can be produced in a different way: as the Perron-Frobenius vector for the dual operator. Thus, for a $(d \times d)$ matrix, this is the row eigenvector $\mathbf{v}^t$ of our first proof.

Proof: we have $\mathbf{v}, \mathbf{w}$ with $M\mathbf{w} = \lambda \mathbf{w}, \mathbf{v}^t M = \lambda \mathbf{v}^t$, normalized so that $\mathbf{v}^t \mathbf{w} = 1$. wolog, take $\lambda = 1$. By Birkhoff, there is a linear functional φ on $V = \mathbb{R}^n$ such that for any $\mathbf{u} \in V$, $M^n(\mathbf{u} \to \varphi(\mathbf{u})\mathbf{w}$. We claim that $\varphi(\mathbf{u}) = \mathbf{v}^t \mathbf{u}$.

But $\mathbf{v}^t M^n \mathbf{u} = \mathbf{v}^t \mathbf{u}$ and also $\lim \mathbf{v}^t M^n \mathbf{u} = \mathbf{v}^t \lim M^n \mathbf{u} = \mathbf{v}^t \varphi(\mathbf{u}) \mathbf{w} = \varphi(\mathbf{u}) \mathbf{v}^t \mathbf{w} = \varphi(\mathbf{u})$.

It is interesting to consider this in the nonstationary situation of a matrix sequence $(M_i)_{i \in \mathbb{Z}}$, where $\mathbf{w}$ depends on the future $(M_i)_{i \geq 0}$, and v on the past $(M_i)_{i \leq 0}$. We claim that the same holds for the nonstationary version of Birkhoff's proof.

Remark 24.6. McMullen via de Faria: $Y = (0, +\infty)$; $X = f(Y) = (f(0), f(+\infty))$.

$$d = d_Y = d(0, \infty)$$

$$g = f^{-1}$$

And

$$\sinh s = (e^s - e^{-s})/2, \cosh s = (e^s + e^{-s})/2, \tanh(s) = \frac{\sinh}{\cosh} = \frac{e^s - e^{-s}}{e^s + e^{-s}}.$$

$$\text{for } \ x \in X, \ s(x) = \inf d(x, \{f(0), f(+\infty)\}) \leq \operatorname{diam}(X)/2 = \Delta/2$$

$$\Phi(s) = \sinh(s) \log \frac{1 + e^{-s}}{1 - e^{-s}}$$

$$|g'(x)| \geq (\Phi(s(x)))^{-1}$$

$$g(x) = y$$

$$f'(y) = 1/g'(x) \leq (\Phi(s(x)) = \sinh(s) \log \frac{1 + e^{-s}}{1 - e^{-s}} = \frac{e^s - e^{-s}}{2} \log \frac{1 + e^{-s}}{1 - e^{-s}}$$

So ....

25. Geodesic flows and the modular flow

Having just encountered hyperbolic geometry, we will take a break from the projective and Hilbert metrics (which we return to later) to study a central example for dynamics: the geodesic flow for a hyperbolic Riemann surface.

25.1. **The geometry of Möbius transformations.**

Example 23. (Examples of isometries)

We remark on the difference between the maps $F : z \mapsto 1/z$, and $K : z \mapsto \overline{z}$, which switch the upper and lower half planes, and $G : z \mapsto -1/z$ and $L : z \mapsto 1/\overline{z}$, which preserve them. K is an anticonformal map, given as a map of $\mathbb{R}^2$ by the linear action on column vectors by the matrix $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ with determinant -1, and so orientation-reversing; it fixes $(1, 0)$, in other words $1 \in \mathbb{C}$, and sends $(0, 1)$ to $(0, -1)$, that is $i \mapsto -i$. The map F also fixes 1 and sends i to $-i$, so one might at first think it also reverses orientation. Indeed as a Möbius transformation it comes from the matrix $M = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, also with determinant -1. However all complex Möbius transformations are orientation-preserving, since this is defined infinitesimally, and the derivative at z is the complex number $f'_M(z) = -z^{-2}$; the derivative acts on the tangent plane $\mathbb{C}$ at z as multiplication by this complex number, which is a rotation composed with a dilation, so again, orientation-preserving. The map $L : z \mapsto 1/\overline{z}$ is a

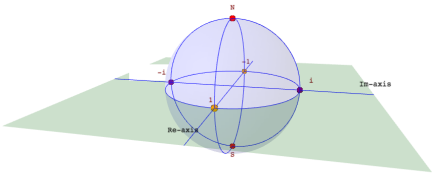


FIGURE 57. The maps $\varphi^{-1}\circ F\circ\varphi$ and $\varphi^{-1}\circ G\circ\varphi$ exchange $N=\varphi^{-1}(\infty)$ and $S=\varphi^{-1}(0)$, rotating about the real and imaginary axes of the Riemann sphere respectively. Note that the rotations go along the two great circles through the poles, which are the images by φ^{-1} of these axes.

composition of these two, hence is also anti-conformal. L is important geometrically as it defines *inversion in the unit circle*; each ray from 0 is mapped to itself, with reciprocal norm. It preserves the hyperbolic metric on $\mathbb{H}$, since it preserves cross-ratios and the collection of circles perpendicular to the real axis. All other circle inversions can be defined from this via conjugation with Möbius transformations. The map $G: z\mapsto -1/z$ is Möbius, with determinant one matrix $M=\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}$. This preserves $\mathbb{H}$ and maps the unit circle to itself via a reflection in the imaginary axis. See Fig.??

It is interesting to visualize the maps $F: z\mapsto 1/z$ and $G: z\mapsto -1/z$ transported to the Riemann sphere. For this let us label the and the north and south poles N, S and also the real and imaginary axes by the inverse images of the stereographic projection $\varphi: S^2\to\widehat{C}$. Now the points of the equator are fixed since this is the unit circle, so we also draw the points $\pm 1, \pm i$. Then $F: z\mapsto 1/z$ is a rotation of the sphere around the real axis in $\mathbb{C}$, while $G: z\mapsto -1/z$ rotates around the imaginary axis in $\mathbb{C}$. In other words these are rotations along two great circles through the poles, corresponding to the imaginary and the real axes respectively. See Fig. 57. Both maps interchange the poles, and so, $\mathbb{H}$ and $-\mathbb{H}$.

A second example of an antiholomorphic map of $\mathbb{H}$ which preserves the hyperbolic metric is $x+yi\mapsto -x+iy$. Further examples are the composition of this with any map in $\mathrm{Möb}^+(\mathbb{R})$).

We have noted that the antiholomorphic map $z\mapsto 1/\overline{z}$ is inversion in the unit circle; as a map of $\mathbb{R}^2$ this sends a vector $\mathbf{v}$ to $\mathbf{v}/||\mathbf{v}||^2$. (The same formula defines inversion in the unit sphere of any any Euclidean space; this allows one to define Möbius transformations in $\mathbb{R}^d$, as a composition of an even number of inversions; see [Bea83]!)

We can define inversion in any circle in the Riemann sphere (so, any line or circle C in $\widehat{\mathbb{C}}$) by conjugation with a Möbius transformation which takes $\mathbb{C}$ to the unit circle.

We claim that inversion in the real line is the map $z\mapsto\overline{z}$, that is, reflection in the real axis. To prove this, note that

$$g(z)=\frac{z-i}{z+i}$$

maps $\mathbb{H}$ to Δ. This follows from the fact that g maps $-1, 0, 1, \infty$ and i to $-1' = i, 0' = -1, 1' = -i, \infty' = 1$ and 0. See Fig. ?? We claim that $h(z) = g^{-1} \circ j \circ g(z) = \overline{z}$. Now the inversion $j(z) = 1/\overline{z}$ fixes all points on the unit circle whence h fixes all points on the real line. And j interchanges 0 and ∞ whence h interchanges i and $-i$, so indeed $h(z) = \overline{z}$.

In other words, inversion in the real line is just Euclidean reflection in the line. For this reason, the words *inversion* and *reflection* in circles (and lines) are used interchangeably.

Next we study the geometry and dynamics of Möbius transformations. In fact we consider the behavior of the map f_M acting on $\widehat{C}$ or (via conjugation) S^2 in parallel to the geometry of the matrix $M \in SL(2, \mathbb{C})$, as it acts on $\mathbb{C}^2$. There are two closely related quadratic equations which appear here: for the fixed points of f_M and for the eigenvalues of M. The geometry of M is determined by the eigenvectors and the eigenvalues, these being preserved by conjugation in $SL(2, \mathbb{C})$, while parallel to this the geometry of f_M is fixed points together with their *multipliers*, see below; these are preserved by conjugation in Möb($\mathbb{C}$). We have seen the beginning of this in Proposition 23.3, where we showed that f_M is determined by where it sends three distinct points, so if three points are fixed, it must be the identity map. The matrix point of view developed here allows for a second proof of this, in (ii) of Proposition 25.3, with no calculation.

Lemma 25.1. *Given $M \in SL(2, \mathbb{C})$, then either:*
(i) there are (up to nonzero multiples) two linearly independent eigenvectors $\mathbf{v}_0, \mathbf{v}_1$ or (ii) there is one eigenvector $\mathbf{v}$.
In case (i), the eigenvalues λ_0, λ_1 satisfy $\lambda_0 \cdot \lambda_1 = 1$, and M is similar to a diagonal matrix D with those entries. Thus $D = B^{-1}MB$ via the change-of-basis matrix $B \in SL(2, \mathbb{C})$ with columns $\mathbf{v}_0, \mathbf{v}_1$. $M = \pm I$ iff here $\lambda_0 = \lambda_1$.
In case (ii), there exists a rotation matrix R such that $R^{-1}MR = S$ with S the shear transformation $S = \pm \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix}$. Moreover M is conjugate in $PSL(2, \mathbb{C})$ (but perhaps not in $SL(2, \mathbb{C})$) to $H^+ = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and to $H^- = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.

Proof. The eigenvalues of M are the roots of the characteristic polynomial $p_M(\lambda)$. This is quadratic, so by the Fundamental Theorem of Algebra there are always two roots in $\mathbb{C}$, which are either distinct $\lambda_0 \neq \lambda_1$ or the same (a *double root*), $\lambda_0 = \lambda_1$. We recall that the eigenvectors for distinct eigenvalues are linearly independent, giving case (1). A double root can either give $M = \pm I$ (also case (i)) or case (ii).
(i): In the first case, with linearly independent eigenvectors $\mathbf{v}_0, \mathbf{v}_1$, build a matrix B with these as the columns. Multiplying these by nonzero complex numbers we can assume that $\det(B) = 1$ so $B \in SL(2, \mathbb{C})$. Then setting $\mathbf{e}_0 = (1, 0), \mathbf{e}_1 = (0, 1))$ we have for $D \equiv B^{-1}MB$ that $D\mathbf{e}_0 = B^{-1}M(B\mathbf{e}_0) = B^{-1}M\mathbf{v}_0 = \lambda_0 B^{-1}\mathbf{v}_0 = \lambda_0 \mathbf{e}_0$ and similarly for $\mathbf{e}_1$, whence D is diagonal with entries λ_i. From the similarity equation $\det D = \det M = 1$, whence $\lambda_0\lambda_1 = 1$ as claimed. In the special case $\lambda_0 = \lambda_1 \equiv \lambda$ then this gives $\lambda = \pm 1$, whence $D = \pm I$ and so also $M = \pm I$.

(*ii*): In the second case, where M has only one eigenvector $\mathbf{v}_0$, then it has only one (double) eigenvalue, as noted. We again form a change-of-basis matrix B with first column $\mathbf{v}_0$ and now with second column any $\mathbf{v}_1$ which is linearly independent of $\mathbf{v}_0$. Then for $S = B^{-1}MB$ we have as above that $S\mathbf{e}_1 = B^{-1}M(B\mathbf{e}_1) = \lambda\mathbf{e}_1$. This yields

$$S = \begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$$

which has eigenvalues $a, d = \pm 1$, whence

$$S = \pm \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix}$$

for some $b \in \mathbb{C} \setminus \{0\}$.

Multiplying by constants, we again can take $B \in SL(2, \mathbb{C})$. We can do slightly better than this: choosing $\mathbf{v}_1$ such that the basis $(\mathbf{v}_0, \mathbf{v}_1)$ is of positive orientation and normalized, then $B = R$ is a rotation matrix.

One further improvenment is possible. Let $C = \pm \begin{bmatrix} w & 0 \\ 0 & w^{-1} \end{bmatrix}$. Then $C^{-1}SC = \pm \begin{bmatrix} 1 & w^{-2}b \\ 0 & 1 \end{bmatrix}$. Hence for $w = b^{1/2}$ we have that

$$\widetilde{M} = (RB)^{-1}M(CR) = \pm \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

For the last statement, write $H^+ = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $H^- = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ and define $K = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Then $KH^+K = H^-$. (Indeed, multiplication on the right by K switches columns while multiplication on the left switches rows.) We remark that this is a conjugacy in $GL(2, \mathbb{C})$ but not $SL(2, \mathbb{C})$ since K has determinant -1; they are also conjugate in $PSL(2, \mathbb{C})$, since $K \sim iK$ which has determinant 1.

□

Lemma 25.2. *Given $M \in SL(2, \mathbb{R})$, then either:*
(i) there are (up to nonzero multiples) two linearly independent eigenvectors $\mathbf{v}_0, \mathbf{v}_1$ or
(ii) there is one eigenvector $\mathbf{v}$.
In case (i), the eigenvalues λ_0, λ_1 satisfy $\lambda_0 \cdot \lambda_1 = 1$, and M is similar to a diagonal matrix D with those entries. Thus $D = B^{-1}MB$ via a change-of-basis matrix $B \in SL(2, \mathbb{R})$ with columns $\mathbf{v}_0, \mathbf{v}_1$. $M = \pm I$ iff here $\lambda_0 = \lambda_1$.

There are two subcases:
(ia) $(trM)^2 - 4 > 0$, and the eigenvalues are real, and
(ib)$(trM)^2 - 4 < 0$, and they are imaginary.
In case (ii), $(trM)^2 - 4 = 0$ and there is a double real eigenvalue. Then there exists a rotation matrix R such that $R^{-1}MR = T$ with $T = \pm \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix}$. Moreover M is conjugate in $GL(2, \mathbb{C})$ to $H^+ = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and to $H^- = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.

Proof. The characteristic polynomial is

$$p_M(\lambda) = \det(M - \lambda I) = \lambda^2 - (\mathrm{tr}M)\lambda + \det M = \lambda^2 - (\mathrm{tr}M)\lambda + 1$$

with roots

$$\lambda^{\pm} \equiv \frac{\mathrm{tr}M \pm \sqrt{(\mathrm{tr}M)^2 - 4}}{2}.$$

Defining $\alpha = \sqrt{(\mathrm{tr}M)^2 - 4}$, then since the entries are real, if $(\mathrm{tr}M)^2 - 4 > 0$ we have $\lambda^{\pm} = (\mathrm{tr}M \pm \alpha)/2$. If $(\mathrm{tr}M)^2 - 4 < 0$ then writing $\beta = \sqrt{4 - (\mathrm{tr}M)^2} > 0$ we have $\lambda^{\pm} = (\mathrm{tr}M \pm \beta i)/2$. Lastly if $(\mathrm{tr}M)^2 - 4 = 0$ then $\lambda^{\pm} = (\mathrm{tr}M)/2$. Note that in all three cases, $\lambda^+ \cdot \lambda^- = 1$, verifying what we already know from Lemma 25.1.

We know from Lemma 25.1 that, defining as above B to have a columns the eigenvectors, this gives a change-of-basis matrix in $GL(2, \mathbb{C})$. We now show in fact we can take the eigenvectors to be real and B to be in $SL(2, \mathbb{R})$. Now for an eigenvector (without loss of generality) $\mathbf{v} = \begin{bmatrix} z \\ 1 \end{bmatrix}$ we have

$$M\mathbf{v} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} z \\ 1 \end{bmatrix} = \begin{bmatrix} az + b \\ cz + d \end{bmatrix} = \begin{bmatrix} \lambda z \\ \lambda \end{bmatrix}$$

so $cz + d = \lambda$ whence $\lambda \in \mathbb{R} \implies z \in \mathbb{R}$. Thus up to multiplication by a constant in $\mathbb{C}^*$, the eigenvectors are real, whence $B \in SL(2, \mathbb{R})$.

$D = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$

□

Remark 25.1. We state the next theorem for $M \in SL(2, \mathbb{C})$, as in Lemma 25.1. We recall from Corollary 23.4 that Möb($\mathbb{C}$) is isomorphic to $PGL(2, \mathbb{C})$, which is isomorphic to $PSL(2, \mathbb{C})$ and $SL(2, \mathbb{C})/\{\pm I\}$.

We remark that if $\widetilde{M} \in GL(2, \mathbb{C})$ is equivalent to M, that is, there exists $c \in C \setminus 0$ such that $\widetilde{M} = cM$, then $\widetilde{M}$ and M have the same eigenvectors, however the eigenvalues for $\widetilde{M}$ are multiplied by c.

In particular for $M \in SL(2, \mathbb{C})$, if $\mathbf{v}$ is an eigenvector for M with eigenvalue λ, then $\mathbf{v}$ is an eigenvector for $(-M)$ with eigenvalue $-\lambda$. In part (v) below, switching to $(-M)$ gives the same result as this does not affect the value of the multiplier λ^2.

We have:

Proposition 25.3.
(i) Eigendirections for $M \in SL(2, \mathbb{C})$ correspond bijectively to fixed points of f_M via the map $\pi : \mathbb{C}^2 \setminus \{\mathbf{0}\} \to \widehat{C}$.
(ii) f_M is the identity map iff it has three fixed points (second proof).
(iii) f_M has two fixed points iff M has two distinct eigenvalues. In this case we have an eigenvector $\mathbf{v} = (z, w) \in \mathbb{C}^2 \setminus \{\mathbf{0}\}$ with eigenvalue λ iff the fixed point $\widetilde{z} = \pi(\mathbf{v})$ has multiplier λ^2. The second multiplier is then λ^{-2}. The diagonalization $BMB^{-1} = D$ of Lemma 25.1 induces the conjugacy in Möb($\mathbb{C}$) of f_M to the map f_D. Since D is diagonal, f_D has fixed points $0, \infty$; interchanging the columns of B switches these fixed points.

(iv) If $M = \pm I$ then $f_M = \pm$ id. If $M \neq \pm I$ has a single eigenvalue λ (hence $\lambda = \pm 1$) then f_M has a single fixed point, with multiplier 1. The conjugacy of M to $\pm H^+ = \pm \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ induces a conjugacy of f_M to f_{H^+}, which has fixed point ∞. The conjugacy of M to $\pm H^-$ gives f_{H^-}, with fixed point 0.

Proof. (i): Let $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL(2, \mathbb{C})$.

Let $\mathbf{v} = (z, w) \in \mathbb{C}^2 \setminus \{\mathbf{0}\}$ with $M\mathbf{v} = \lambda\mathbf{v}$. Note that since M is invertible, $\lambda \neq 0$. Setting $\widetilde{z} \equiv \pi(z, w) = z/w \in \widehat{\mathbb{C}}$, then from the commutative diagram of Fig. 70

$$\begin{array}{ccc} \mathbb{C}^2 \setminus \{\mathbf{0}\} & \xrightarrow{M} & \mathbb{C}^2 \setminus \{\mathbf{0}\} \\ \downarrow \pi & & \downarrow \pi \\ \widehat{\mathbb{C}} & \xrightarrow{f_M} & \widehat{\mathbb{C}} \end{array} \tag{93}$$

we have that $f_M(\widetilde{z}) = \pi(M\mathbf{v}) = \pi(\lambda\mathbf{v}) = \widetilde{z}$. Conversely, if $f_M(z') = z'$ then for $\mathbf{v} = (z, w)$ with $\pi(\mathbf{v}) = z'$, we have $\pi(M\mathbf{v}) = z'$ whence $M\mathbf{v} = \lambda\mathbf{v}$ for some $\lambda \neq 0$.

(ii): Of course we already proved this basic fact in part (i) of Proposition 23.3; now we have a very different understanding of this. By part (i), f_M has three fixed points iff M has three eigenvectors which are not multiples of one another. We know that if n vectors in a vector space have distinct eigenvalues for a map T then they are linearly independent (see Lemma l:linearlyindep.) Since three vectors in $\mathbb{R}^2$ cannot be linearly independent, two of them must have the same eigenvalue. Since those two vectors form a basis, M must be a constant times the identity matrix. Thus $f_M =$ id.

(iii): The characteristic polynomial of M with $\det M = 1$ is quadratic so has two complex roots (possibly a double root). If it has two distict roots λ_1, λ_2 then there are two linearly independent eigenvectors $\mathbf{v}_1, \mathbf{v}_2$. Build a matrix B with these as the columns. Then $BMB^{-1} = D$ is diagonal with entries λ_i. Since $\det D = 1$ also, $\lambda_1\lambda_2 = 1$, so let us call these λ, λ^{-1}. For $D = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$ then $D \sim \widetilde{D} = \begin{bmatrix} \lambda^2 & 0 \\ 0 & 1 \end{bmatrix}$ which is the map $f_D = f_{\widetilde{D}}$ with $f_D : z \mapsto \lambda^2 \cdot z$, with fixed points $0, \infty$.

(iv) The multipliers of f_M are preserved by conjugation in Möb$(\mathbb{C})$. That is, for $f, g \in$ Möb$(\mathbb{C})$, if $\widetilde{f} = g^{-1} \circ f \circ g$, then if z is a fixed point for f with multiplier w, then $g^{-1}(z)$ is a fixed point for $\widetilde{f}$ with the same multiplier (by the chain rule). Similarly, the eigenvectors and values of M are preseved by conjugation by a matrix G.

For $f : z \mapsto w \cdot z$, the derivative at any $z \in \mathbb{C}$ is w, hence this is the derivative at the fixed point 0. Given f_M as in (ii), we conjugate it to f_D. Thus the multiplier of f_D at 0. and hence of f_M at z, is λ^2.

As noted above in Lemma ??, $H^+ = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $H^- = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ are conjugate via $K = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Note that $f_K(z) = 1/z$ which exchanges 0 and ∞.

If we exchange the columns of B to get $\widetilde{B} = BK$, then for the new matrix $\widetilde{D}$ we have $f_{\widetilde{D}} : z \mapsto \lambda^{-1}$. Now since f_K exchanges 0 and ∞ ??? $B^{-1}KB = \widetilde{B}^{-1}B$. In conclusion the multiplier of f_D at ∞ is the multiplier of $f_{\widetilde{D}}$ at 0, that is, λ^{-2}.

□

.....

Remark 25.2. We give a second proof of part (v) which is "elementary" in that it is a computation, because we find this interesting as well.

Given $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in GL(2,\mathbb{C})$, to calculate the fixed points for f_M we have $f_M(z) = (az+b)/(cz+d) = z$ so

$$cz^2 + (d-a)z - b = 0$$

This has roots

$$\frac{(a-d) \pm \sqrt{(d-a)^2 + 4cb}}{2c}$$

Since $(d-a)^2 + 4cb = (a+d)^2 - 4ad + 4bc$ then assuming without loss of generality that $M \in SL(2,\mathbb{C})$ so $\det M = 1$, and writing $\mathrm{tr} = \mathrm{tr}(M)$, the roots are

$$\alpha^{\pm} = \frac{(a-d) \pm \sqrt{(\mathrm{tr})^2 - 4}}{2c}.$$

We note that

$$\alpha^+\alpha^- = \frac{(a-d)^2 - (\mathrm{tr})^2 + 4}{4c^2} = \frac{-ad+1}{c^2} = -b/c.$$

To find the eigenvalues of M and then the eigenvectors. We want the roots of the characteristic polynomial $p_M(\lambda) = \det(M - \lambda I) = \lambda^2 - \mathrm{tr}M\lambda + \det M$.

Assuming $M \in SL(2,\mathbb{C})$ so $\det(M) = 1$, these are

$$\lambda^{\pm} \equiv \frac{\mathrm{tr} \pm \sqrt{(\mathrm{tr})^2 - 4}}{2}$$

We have

$$\lambda^+ \cdot \lambda^- = (\mathrm{tr}^2 - \mathrm{tr}^2 + 4)/4 = 1$$

so these are reciprocals.

An eigenvector for λ is $\mathbf{v} = \begin{bmatrix} x \\ 1 \end{bmatrix}$ where $x = (\lambda - d)/c$. Note that $\pi(\mathbf{v}) = (\lambda - d)/c$.

We then check that in fact this is our fixed point for f_M.

We calculate that for $f_M(z) = (az+b)/(cz+d)$ the derivative is

$$f'_M(z) = \det(M)/(cz+d)^2. \tag{94}$$

The derivative at the point ∞ is defined via the chart $F(z) = 1/z$, to be the derivative at zero of $\widetilde{f} = F^{-1} \circ f_M \circ F$; we have $\widetilde{f}(z) = (dz+c)/(bz+a)$ whence

$$\widetilde{f}'(z)) = \det(M)/(a+bz)^2. \tag{95}$$

The derivative at ∞ is independent of choice of chart, by the Chain Rule, giving

$$f'_M(\infty) \equiv \widetilde{f}'(0)) = \det(M)/a^2. \tag{96}$$

So if $z \neq \infty$ is a fixed point then assuming $\det M = 1$ the multiplier at z is $1/(cz+d)^2$ and if ∞ is a fixed point it is $1/a^2$.

For the fixed points we had

$$\alpha^{\pm} = \frac{(a-d) \pm \sqrt{(\mathrm{tr})^2 - 4}}{2c}.$$

Thus if $\alpha \neq \infty$ the multiplier there is
$1/(cz+d)^2$ and

$$(cz+d) = c\frac{(a-d) \pm \sqrt{(\mathrm{tr})^2 - 4}}{2c} = \frac{\mathrm{tr} \pm \sqrt{(\mathrm{tr})^2 - 4}}{2} = \lambda$$

whence
the multiplier is $\lambda^{-2} = (\lambda_1)^2$.
so in particular if ∞ is a fixed point then this is the multiplier there.

....
$1/(1/z+1) = 1/((1+z)/z) = z/(1+z)$
0 fixed
deriv= $((1+z)-z)/(1+z)^2 = (1+z)^{-2}$
at 0= 1 Specifically, with the trace of M being $\mathrm{tr}(M) = a+d$, then

$$p_M(\lambda) = \det(M - \lambda I) = \lambda^2 - \mathrm{tr}(M)\lambda + \det M = \lambda^2 - \mathrm{tr}(M)\lambda + 1,$$

(since $M \in SL(2,\mathbb{C})$ so $\det(M) = 1$), and the roots are, writing $\mathrm{tr} = \mathrm{tr}(M)$,

$$\lambda^{\pm} \equiv \frac{\mathrm{tr} \pm \sqrt{\mathrm{tr}^2 - 4}}{2}$$

We have

$$\lambda^+ \cdot \lambda^- = (\mathrm{tr}^2 - \mathrm{tr}^2 + 4)/4 = 1$$

so the two roots are reciprocals.

In particular, for a double root, $\lambda = \lambda_0 = \lambda_1 = \pm 1$.

There are two possibilities: either M has only one eigenvector or has more. If it has two linearly independent eigenvectors with the same eigenvalue λ, then every vector is an eigenvector so $M = \lambda I$. Thus from the above, $M = \pm I$.

If the characteristic polynomial of M has two distict roots λ_0, λ_1 then there are
If they are distinct, $\lambda^{\pm} = \rho^{\pm 1} e^{i\pm t}$ and either $|\lambda_1| < |\lambda_0|$ or $|\lambda_1| = |\lambda_0| = \rho = 1$.

An eigenvector for λ is $\mathbf{v} = \begin{bmatrix} x \\ 1 \end{bmatrix}$ where $x = (\lambda - d)/c$. Note that $\pi(\mathbf{v}) = (\lambda - d)/c$.
....

Hence for $w = b^{1/2}$ we have that

$$\widetilde{M} = (RB)^{-1} M (KR) = \pm \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Definition 25.1. A Möbius transformation $f_M \in \mathrm{Möb}^+(\mathbb{C})$ is called:
(i) *loxodromic* iff the map has two fixed points, and the multipliers $\alpha^{\pm}$ has modulus $\neq 1$ (if one does, the other does by ??); the fixed point with $|\alpha_1| < 1$ is called attracting (or *contracting*), and the other with $|\alpha^+| > 1$ *repelling* (or *repulsive* or *expanding* The loxodromic maps are further divided into those that are
(ia)*hyperbolic*: the multiplier is real;
(ib)*purely loxodromic*: the multiplier nonreal;

(ii) *elliptic*: the map has two *neutral* fixed points, meaning the multiplier has modulus one;
(iii) *parabolic*: the map has a single fixed point, with multiplier zero.

We note that the multiplier at a fixed point is just the eigenvalue of the derivative map there (since the tangent space to the complex one-manifold $\widehat{C}$ at every point is just $\mathbb{C}$ itself).

We next see how the geometry of the map is determined by this number of fixed points together with knowledge of the multipliers.

Definition 25.2. A Möbius transformation $f_M \in \text{Möb}^+(\mathbb{C})$ is called:
(i) *loxodromic*: the map has two fixed points, one *attracting* (i.e. contracting, the multiplier there has modulus < 1) and one *repulsive* or expanding (the multiplier has modulus > 1). The loxodromic maps are further divided into those that are
(ia)*hyperbolic*: the multiplier is real $\neq 1$, and the map moves anong circles connecting the fixed points, or
(ib)*purely loxodromic*: the multiplier is complex nonreal of modulus $\neq 1$, and points move along spirals of a chosen slope connecting the fixed points;
(ii) *elliptic*: the map has two *neutral* fixed points, meaning the multiplier has modulus one, thus neither expanding nor contracting; the map moves points along circles about these points.
(iii) *parabolic*: the map has a single fixed point, with multiplier zero there; points move along two families of circles tangent to a tangent vector at this point, one clockwise and one counterclockwise.

Example 24. (Rotations) We examined above the maps $F : z \mapsto 1/z$ and $G : z \mapsto -1/z$. These are elliptic maps, as they are conjugate by stereographic projection φ to rotations of the Riemann sphere around the real and imaginary axes respectively by angle π, interchanging the north and south poles, while fixing the points ± 1, $\pm i$ respectively; the multiplier at each of these points is -1. See Fig. 57. In $\widehat{\mathbb{C}}$, they interchange 0 and ∞.

See also Example 44.

To generalize these maps, if we try to think of a Möbius transformation which is the analogue of a rotation on the plane $\mathbb{R}^2$, we might come up with *two* natural candidates.

The first is $f_{M_t} : z \mapsto e^{it}z$; this rotates the plane $\widehat{C}$ counterclockwise by angle t. The second is the real matrix $R_t \in PSO(2, \mathbb{R})$ where

$$R_t = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

for $a = \cos(t)$ and $b = \sin(t)$, which rotates the plane $\mathbb{R}^2$ conterclockwise by angle t. So let us consider what each does as a Möbius transformation.

For the first, M_t is the diagonal matrix

$$M_t = \begin{bmatrix} e^{it} & 0 \\ 0 & 1 \end{bmatrix}$$

with the eigenvalues the diagonal entries $e^{it}, 1$ and eigenvectors $(1,0)$ and (0.1). Normailzing to a matrix with determinant one, we have

$$\widetilde{M_t} = \begin{bmatrix} e^{it/2} & 0 \\ 0 & e^{-it/2} \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

The Möbius transformation $f_{M_t} = f_{\widetilde{M_t}}$ has two fixed points, 0 and ∞. From (94), the multiplier at zero is $f'_{\widetilde{M}}(0) = 1/d^2 = e^{it}$ as we would expect, while from (96), the multiplier at ∞ is $f'_{\widetilde{M}}(\infty) = 1/a^2 = e^{-it}$.

Regarding the second example $R_t = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$, see Remark 44: this has complex eigenvectors $\begin{bmatrix} 1 \\ -i \end{bmatrix}$ and $\begin{bmatrix} 1 \\ i \end{bmatrix}$ with eigenvalues $e^{i\theta}$, $e^{-i\theta}$.

The Möbius transformation f_{R_t} preserves $\mathbb{H}$ and the real axis, since R_t has real entries. Since $f_{R_t}(z) = (ai - b)/(bi + a)$, it rotates around the two fixed points $\pm i$. The matrices M_t and R_t are conjugate by the change of basis matrix whose columns are the eigenvectors. Since R_t is a normal matrix, defining

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix}$$

this is a unitary change-of-basis matrix, yielding the diagonalization $U^* A U = D$ where $D = \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix}$ with commutative diagrams

$$\begin{array}{ccc} \mathbb{C}^2 & \xrightarrow{R_t} & \mathbb{C}^2 \\ \uparrow{\scriptstyle U} & & \downarrow{\scriptstyle U^*} \\ \mathbb{C}^2 & \xrightarrow{M_t} & \mathbb{C}^2 \end{array} \qquad \begin{array}{ccc} \widehat{\mathbb{C}} & \xrightarrow{f_{R_t}} & \widehat{\mathbb{C}} \\ \uparrow{\scriptstyle f_U} & & \downarrow{\scriptstyle f_U^{-1}} \\ \widehat{\mathbb{C}} & \xrightarrow{f_{M_t}} & \widehat{\mathbb{C}} \end{array} \tag{97}$$

and indeed, since $f_U(z) = (z+1)/(-iz+i)$, we have $f_U(0) = -i, f_U(\infty) = i$.

By (94), the multiplier of f_{R_t} at i is $f'_{R_t}(i) = 1/(bi+a)^2 = e^{-i2t}$ while at $-i$ it is $f'_{R_t}(-i) = 1/(b(-i)+a)^2 = e^{i2t}$.

Note that in the parabolic case, if the fixed point is ∞, then these circles tangent to ∞ are the familes of parallel lines in $\mathbb{C}$ with a chosen direction.

We shall in fact have three different (equivalent) ways of classifying these maps:
–geometrically, as just described, in terms of the behavior at fixed points;
–algebraically, in terms of the trace $a + d$ of the matrix $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$;
–in terms of a canonical form, that the map is conjugate to one of three types.

Definition 25.3. A *one-parameter subgroup* of a group G is $H = \{g_t : t \in \mathbb{R}\}$ satisfying $g_s \circ g_t = g_{t+s}$. In other words, this defines an action of the additive group $(\mathbb{R}, +)$, see Definition 2.1.

This is closely related to the flow property, since it states that the action of H on G on the left defines an action of $(\mathbb{R}, +)$, i.e. a flow (Definition 35.3) and more generally, if G acts on a space X, then the action of H on X defines a flow.

Proposition 25.4. *Each non-identity element M of $SL(2,\mathbb{C})$ embeds in a unique one-parameter subgroup. This is of the form $exp(tA)$ where A is an element of the Lie algebra $\mathfrak{sl}(2,\mathbb{C})$ of $SL(2,\mathbb{C})$. Up to conjugation in $SL(2,\mathbb{C})$, we can take $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ or $A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ where these are a basis for the Lie algebra.*

Each Möbius transformation not equal to id embeds in a unique one-parameter flow. Up to conjugation in Möb$(\mathbb{C})$, , this is of three types:

Proof. The exp map is onto, so exists a uiuque $A \in \mathfrak{sl}(2,\mathbb{C})$ with $M = \exp(A)$ Then $M^t = (\exp(A))^t = \exp(tA)$ is our one-parameter subgroup. Now express A in terms of the basis. NO..... □

In addition, we shall describe the classification:
–in terms of the elements A of the Lie algebra such that $M = e^A$. This will naturally embed each type in a one-parameter subgroup.

Proposition 25.5. *For $M \in PSL(2,\mathbb{R})$, f_M is either:*
(i) hyperbolic iff trace$(M) = trM = a + d > 4$, iff f_M is conjugate to the map
(ii) elliptic iff: iff f_M is conjugate to the map
(iii) parabolic iff: iff f_M is conjugate to the map...

Proof. As in Proposition 23.2, for $A \in GL(2,\mathbb{C}) \cong PSL(2,\mathbb{C})$ with $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ we write $f_A \in \text{Möb}^+(\mathbb{C})$ for the correponding Möbius transformation

We note that fixed points for f_A correspond bijectively to eigenvectors of A.

Indeed, for $\mathbf{v} \in \mathbb{C}^2 \setminus \mathbf{0}$ with $\mathbf{v} = (z_1, z_2)$ and $z = z_1/z_2$ we have $A\mathbf{v} = \lambda\mathbf{v}$ iff $f_A(z) = z$.

To find the fixed points, we could simply consider the equation $f_A(z) = z$, which is a quadratic equation, and use the quadratic formula, but it seems easier to find the eigenvalues of A and then the eigenvectors. We want the roots of the characteristic polynomial $p_A(\lambda) = \det(A - \lambda I) = \lambda^2 - \text{tr}A\lambda + \det A$.

Assuming $A \in PSL(2,\mathbb{C})$ so $\det(A) = 1$, these are

$$\lambda^{\pm} \equiv \frac{\text{tr}A \pm \sqrt{(\text{tr}A)^2 - 4}}{2}$$

Writing $\alpha = \text{tr}A$, we have

$$\lambda^+ \cdot \lambda_1 = (\alpha^2 - \alpha^2 + 4)/4 = 1$$

so these are reciprocals.
(i) two complex roots, $\lambda^{\pm} = r^{\pm}e^{\pm i\theta}$ when $(\text{tr}A)^2 \notin [0,4]$ with $0 < r^+, r_1$ and $r^+r_1 = 1$;
(ia) the subcase with two real roots $r^{\pm}$ when $(\text{tr}A)^2 \in (4, +\infty)$;
(ia) the subcase with two nonreal complex roots when $(\text{tr}A)^2 \in (4, +\infty)$;
(ii) two roots of modulus one $\lambda^{\pm} = e^{\pm i\theta}$ when $(\text{tr}A)^2 \in (0,4)$;
(iii) a double root $\lambda = \text{tr}(A)$ when $(\text{tr}A)^2 = 4$.

As we show, these correspond to the previous cases.

□

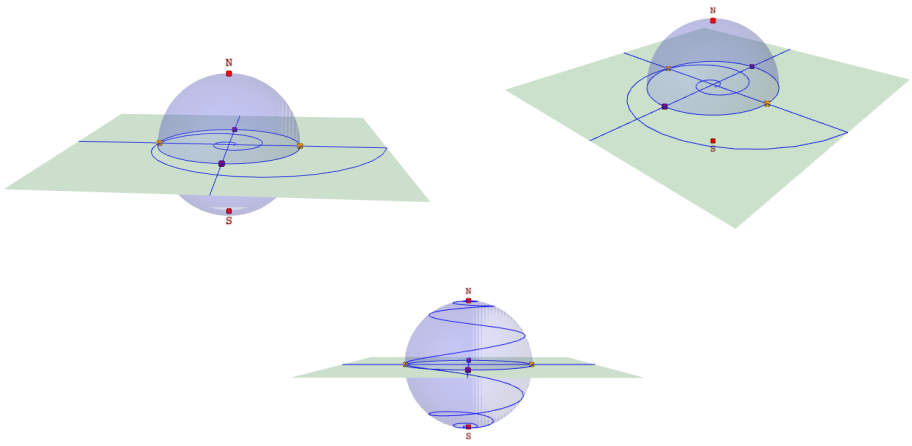


FIGURE 58. A loxodromic spiral in the plane, with fixed points $0, \infty$, and its stereographic projection to the Riemann sphere with fixed points N, S.

Proposition 25.6. *Each complex Möbius transformation embeds in a unique (up to change of time) one-parameter subgroup. For the canonical maps, these are of the form f_{G_t} for $G_t = \exp(tA)$ for the following elements A of the Lie algebra $PSL(2,\mathbb{C})$ =???, as follows:*
(i) loxodromic
(ia)hyperbolic
(ib) purely loxodromic iff:
(ii) elliptic iff:
(iii) parabolic iff:
In the general case, $G_t = \exp(tA)$ where....
Each real Möbius transformation embeds in a unique (up to change of time) one-parameter subgroup. For the canonical maps, these are of the form f_{G_t} for $G_t = \exp(tA)$ for the following elements A of the Lie algebra $PSL(2,\mathbb{C})$ =???, as follows:
(i) hyperbolic iff
(ii) elliptic iff:
(iii) parabolic iff:

In Fig. 59 we show a loxodromic spiral on the Riemann sphere. This is the orbit of the point $z = 1$ in $\widehat{\mathbb{C}}$ by the one-parameter subgroup of Möbius transformations $z \mapsto e^{t/6}e^{it}z$, transferred to the sphere by stereographic projection from the north pole N. (So S is the repelling fixed point and N the attracting.) In Fig. 63 we see orbits of the flows R_t, H_t^+ in the extended complex plane and Riemann sphere. Note that the parabolic flow is a geometric limit of a sequence of elliptic flows, as the pair of fixed points are brought together. It is also the geometric limit of a sequence of hyperbolic flows, as the pair of fixed points are brought together, in the orthogonal direction. It is appropriate that the parablic flow is on the boundary between hyperbolic and elliptic!

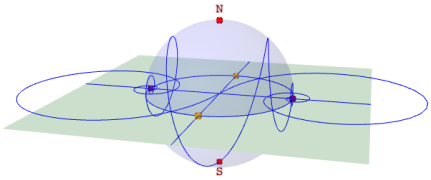


FIGURE 59. A loxodromic spiral in the sphere, and its stereographic projection to the plane, both with fixed points $\pm i$.

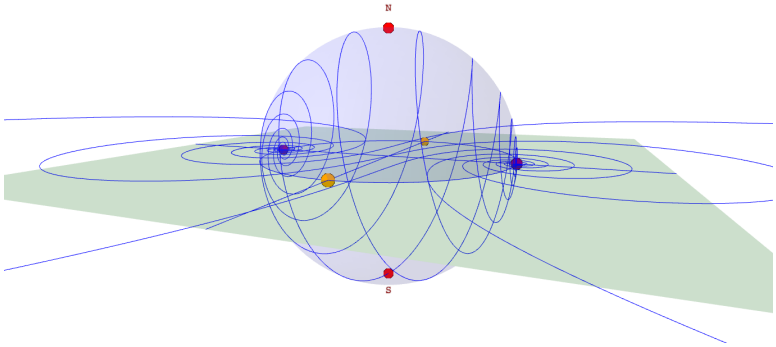


FIGURE 60. Three loxodromic spirals in the plane and sphere, fixed points $\pm i$.

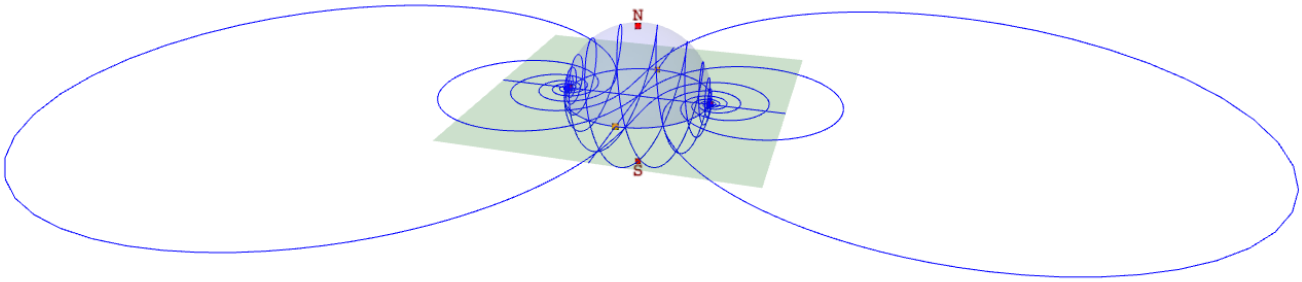


FIGURE 61. Another view of the three loxodromic spirals in the plane and sphere.

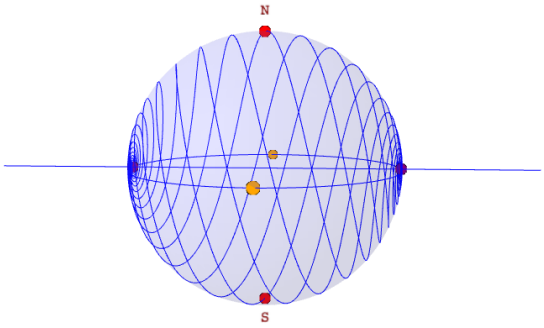


FIGURE 62. Six loxodromic spirals in the sphere.

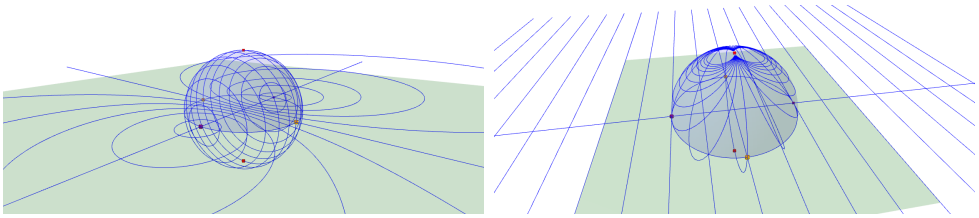


FIGURE 63. Orbits of the elliptic flow R_t in the plane, with fixed points $\pm i$, and its sterographic projection to the Riemann sphere S^2, as a rotation flow about the imaginary axis. Orbits of the parabolic flow H_t^+, with fixed point $\infty \in \widehat{\mathbb{C}}$ and north pole $N \in S^2$.

We have shown that $\text{Möb}^+(\mathbb{R})$ can also be characterized as the group of orientation-preserving isometries of the hyperbolic plane $\mathbb{H}$.

To motivate our next construction, we recall the definition of the torus $T = \mathbb{R}^2/\mathbb{Z}^2$. The additive group $\mathbb{R}^2$ plays two roles, as a surface (the plane), and as a group of isometries of the plane, acting on itself by translation. The subgroup $\mathbb{Z}^2$ defines a discrete group of isometries of the plane, whence the factor group $\mathbb{R}^2/\mathbb{Z}^2$ is also the factor surface $\mathbb{T}$.

Here we replace the plane $\mathbb{R}^2$ by the hyperbolic plane $\mathbb{H}$, and consider a discrete group G of orientation-preserving isometries of $\mathbb{H}$. We then form the factor space $G\backslash\mathbb{H}$; depending on the choice of G, the resulting surface can be, topologically, a surface of any genus (number of holes). The fact that this surface is built from $\mathbb{H}$, which is geometrically the same at every point (since there is a map in $\text{Möb}^+(\mathbb{R})$ taking any chosen point to any other), $G\backslash\mathbb{H}$ shares this property. Such spaces are called *homogeneous spaces* as they are geometrically the same essentially everywhere. (This is with the possible exception of *singular points* where there is a special symmetry

and hence a folding (called a *cone point*, where the total angle around the point is a fraction of 2π)) or a *cusp* at infinity (where the angle is 0).

Since $PSL(2,\mathbb{R})$ is the group of all orientation-preserving isometries, G will be a subgroup of $PSL(2,\mathbb{R})$. Now $\mathbb{H}$ itself is not (unlike the Euclidean plane $\mathbb{R}^2$) itself a group. Nevertheless, the factor group $G\backslash PSL(2,\mathbb{R})$ has an important role to play: as we next explain, this represents the unit tangent bundle of the surface $G\backslash\mathbb{H}$, and it is there that our geodesic flow will act.

Since the group $PSL(2,\mathbb{R})$ is not commutative, one can act on it by subgroups, on the left (as we do with G) or on the right. These have a very different character, as we shall explain.

We consider two actions of $PSL(2,\mathbb{R})$ on itself, on the right and on the left.

Now these are anti-isomorphic, via inversion; precisely, for $\mathcal{I} : PSL(2,\mathbb{R}) \to PSL(2,\mathbb{R})$ defined by $A \mapsto A^{-1}$, and writing $R_A : PSL(2,\mathbb{R}) \to PSL(2,\mathbb{R})$ for the right action map $M \mapsto MA$, $L_A : PSL(2,\mathbb{R}) \to PSL(2,\mathbb{R})$ for the left action map $M \mapsto LM$, we have the *anti*-commutative diagram (this means that it switches the order of group multiplication, as the map $\mathcal{I} : PSL(2,\mathbb{R}) \to PSL(2,\mathbb{R})$ is an anti-isomorphism of $PSL(2,\mathbb{R})$ with itself).

$$\begin{array}{ccc} SL(2,\mathbb{R}) & \xrightarrow{R_A} & SL(2,\mathbb{R}) \\ \uparrow{\scriptstyle \mathcal{I}} & & \downarrow{\scriptstyle \mathcal{I}} \\ SL(2,\mathbb{R}) & \xrightarrow{L_{A^{-1}}} & SL(2,\mathbb{R}) \end{array} \qquad \begin{array}{ccc} SL(2,\mathbb{R}) & \xrightarrow{R_{AB}} & SL(2,\mathbb{R}) \\ \uparrow{\scriptstyle \mathcal{I}} & & \downarrow{\scriptstyle \mathcal{I}} \\ SL(2,\mathbb{R}) & \xrightarrow{L_{B^{-1}A^{-1}}} & SL(2,\mathbb{R}) \end{array} \tag{98}$$

Nevertheless, the left and right actions are totally different! To explain this apparently paradoxical statement, we introduce a geometry on $PSL(2,\mathbb{R})$, as follows. One has the following general notions:

Definition 25.4. Let a group G act on a space X. This action is *transitive* iff the G-orbit of every point is all of X. Choosing some point $o \in X$, define the *stabilizer* H_o of o to be $\{h \in G : ho = o\}$, then the space of all left cosets G/H_o is called a *homogeneous space*. It corresponds bijectively to X, and can be thought of as X together with a choice of origin o. Note that the left action of G on itself induces a left action of G on G/H_o which is the same as the original action of G on X. In general, H_o is not a normal subgroup. Indeed, choosing a different point $\overline{o}$, and find a g_o such that $\overline{g}(o) = \overline{o}$. This is a change of origin, and induces a map from G/H_o to $G/H_{\overline{o}}$ given by $gH_0 \to ...$ That is, the map induced by the inner automorphism of G, ...

Note that the inner automorphism (1) does not depend on which such g is selected; it depends only on g modulo H_o.

Next we carry this out for the left action of $PSL(2,\mathbb{R})$ on $T^1(\mathbb{H})$. A convenient choice of base point o as above is i_i, the vector based at $i \in \mathbb{H}$ pointing in the direction i. Then we consider the map $M \mapsto f_M^*(i_i)$ from $PSL(2,\mathbb{R})$ to $T^1(\mathbb{H})$, so $I \mapsto i_i$.

Now $PSL(2,\mathbb{R})$ acts on itself both on the right and on the left. Via the identification of $PSL(2,\mathbb{R})$ with $T^1(\mathbb{H})$, this means we have left and right actions on $T^1(\mathbb{H})$. These must be (anti)-isomorphic, as explained above. However, from a more geometric point

of view, these actions are very different. To explain this, we indroduce a metric on $T^1(\mathbb{H})$ and hence on $PSL(2,\mathbb{R})$ which is invariant for the left action, but decidedly *not* invariant for the right action. There is moreover a more basic obstruction: $T^1(\mathbb{H})$ is not just an abstract space; it is also a fiber bundle, with a projection map from a vector $\mathbf{v}_p$ based at $p \in \mathbb{H}$ to p. And as we shall see, this projecton is not respected by the antisomorphsm.

To understand all this, we shall focus on actions, on the left and right, by the following one-parameter subgroups.

–$\{R_t : t \in \mathbb{R}\}$ where for $a = \cos(t)$ and $b = \sin(t)$, we define $R_t \in PSO(2,\mathbb{R})$ by

$$R_t = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

. We call this the *rotation subgroup.*

—$\{E_t : t \in \mathbb{R}\}$ where we set

$$E_t \equiv \begin{bmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{bmatrix}.$$

This is the *diagonal subgroup.*

– $H_t^u \equiv \begin{bmatrix} 1 & 0 \\ t & 1 \end{bmatrix}$. We call this the *lower triangular subgroup.*

– $H_t^s \equiv \begin{bmatrix} 1 & 0 \\ t & 1 \end{bmatrix}$ We call this the *lower triangular subgroup.*

Note from Fig. 63 that the flows R_t and E_t are conjugate. Unlike E_t, R_t preserves $\mathbb{H}$.

??as does the rotation flow of S^2 around an axis which...??

For this, we give the upper half space $\mathbb{H}$ the hyperbolic metric. We write $T(\mathbb{H})$ for its tangent bundle. Now the hyperbolic metric can be realized as a Riemannian metric, that is, as an inner product on $T(\mathbb{H})$ at each point. We defined the hyperbolic metric via the formula for an infinitesimal line element $\mathrm{d}s$, with $\mathrm{d}s^2 = \frac{\mathrm{d}x^2+\mathrm{d}y^2}{y^2}$; as we explained above, this can be used to define the arc length of a smooth curve. Letting $p = x + i] \in \mathbb{H}$, then the tangent space $T(\mathbb{H})|_p$ is identified with$\mathbb{C} \equiv \mathbb{R}^2$. Writing $\langle \mathbf{v}, \mathbf{w} \rangle$ for the standard inner product on $\mathbb{R}^2$, we then define $\langle \mathbf{v}, \mathbf{w} \rangle_p = \langle \mathbf{v}, \mathbf{w} \rangle / y^2$; thus for $\mathbf{v} = (x, y)$ we have $\langle \mathbf{v}, \mathbf{v} \rangle_p = \frac{x^2+y^2}{y^2}$ as desired.

Now we write $T_1(\mathbb{H})$ for the unit tangent bundle of $\mathbb{H}$. This is a circle bundle; indeed it is the product space $\mathbb{H} \times S^1$. We wish to define a Riemannian metric on this smooth manifold as well.

In fact, given two Riemannian manifolds M, N (smooth manifolds carrying a Riemannian metric) there is a canonical way to take the product $M \times N$. The tangent space at $(p, q) \in M \times N$ is $T(M)|_p \times T(N)|_q$, and the product inner product is simply $(\mathbf{v}_1, \mathbf{w}_1) \cdot (\mathbf{v}_2, \mathbf{w}_2) = \mathbf{v}_1 \cdot \mathbf{v}_2 + \mathbf{w}_1 \cdot \mathbf{w}_2$.

The tangent space of S^1 is just $\mathbb{R}$, on which the inner product is just multiplication of numbers.

The conclusion of all this is that the length along the circle is just arclength. To integrate a curve....

We have defined a Riemannian metric on $T_1(\mathbb{H})$. Next we identify $PSL(2,\mathbb{R})$ with $T_1(\mathbb{H})$, by the standard device of *choosing a base point* for the action, in this case the action on the *left* on $T_1(\mathbb{H})$ by the normalized derivative of a Möbius transformation.

First we consider how these flows act on the unit tangent bundle of the hyperbolic plane $\mathbb{H}$. Then we study these on the unit tangent bundle of the hyperbolic factor surface $G\backslash\mathbb{H}$,

We write $T_1(\mathbb{H})$ for the unit tangent bundle of $\mathbb{H}$ (in the hyperbolic metric), and f_M^* for the normalized derivative of the map f_M, which maps $T_1(\mathbb{H})$ to itself. That is, $f_M^*(\mathbf{v}_p) = c \cdot f_M'(\mathbf{v})_{f_M(p)}$ where c is $1/||f_M'(\mathbf{v})_{f_M(p)}||$. Via this definition, $PSL(2,\mathbb{R})$ acts on $T_1(\mathbb{H})$.

We consider first the rotation flow. For $a = \cos(t)$ and $b = \sin(t)$, define $R_t \in PSO(2,\mathbb{R})$ by

$$R_t = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}.$$

So $f_{R_t}(z) = (az - b)/(bz + a)$. The derivative is

$$f_{R_t}'(z) = (bz + a)^{-2}$$

so

$$f_{R_t}'(i) = (a + bi)^{-2} = (a - bi)^2.$$

Now how does this act on the tangent space at i, which is $\mathbb{C}$? By multiplication by this complex number, which is a rotation in the clockwise direction by an angle of $2t$!

That the angle is doubled agrees with the fact that we are in $PSL(2,\mathbb{R})$, so $f_{R_t} = f_{-R_t} = f_{R_{t+\pi}}$.

25.2. **Geodesic and horocycle flows on the hyperbolic plane.** Next we identify $PSL(2,\mathbb{R})$ with $T^1(\mathbb{H})$. For this we choose a *base point* for the action; this could be any element of $T^1(\mathbb{H})$, but a convenient choice is i_i, the vector based at $i \in \mathbb{H}$ pointing in the direction i. Then we consider the map $M \mapsto f_M^*(i_i)$ from $PSL(2,\mathbb{R})$ to $T^1(\mathbb{H})$, so $I \mapsto i_i$.

We shall now see that via this correspondence, the right action by R_t rotates each tangent vector, not just the base point.

Indeed, since f_{R_t} fixes the point i, while rotating the vector i_i to $\mathbf{v}_i$, then $f_M^* f_{R_t}^*$ maps i_i to $v_p = f_M^*(\mathbf{v}_i)$. This is a vector based at $p = f_M(i)$ and rotated clockwise by angle $2t$.

Thus the map $M \mapsto f_M(i)$ from $PSL(2,\mathbb{R})$ to $\mathbb{H}$ maps all matrices MR_t to the same point p in $\mathbb{H}$.

Remark 25.3. Thus the *right* action by R_t on $PSL(2,\mathbb{R})$, when this has been identified with $T^1(\mathbb{H})$, fixes each point p while rotating each tangent vector $\mathbf{v}_p$, *clockwise* by angle $2t$. But what does the *left* action do?

This defines a Möbius transformation on the plane, which is elliptic; it fixes the points i and $-i$. It preserves $\mathbb{H}$ and $\mathbb{R}$. The imaginary axis is rotated to other circles passing through $\pm i$ and perpendicular to the real axis. Note that this left action is an isometry of $\mathbb{H}$ while the right action is ...

Definition 25.5. Given a differentiable manifold M of dimension d, then the tangent space at $p \in M$, written TM_p, is isomorphic to $\mathbb{R}^d$. These fit together smoothly via the charts of M, to make up TM, the *tangent bundle* of M. This is a *fiber bundle* which means that locally it is a product of $\mathbb{R}^d$ with $\mathbb{R}^d$, with the projection from TM_p to TM, which simply sends a *tangent vector* vector $\mathbf{v}_p \in TM_p$ to p. This is actually a *vector bundle* since, while the manifold M may be curved, the tangent space is linear at each point. A *Riemannian metric* on M is a smooth choice of inner product on this bundle. One says that M together with this metric is a *Riemannian manifold*. The inner product allows one to define a smoothly varying norm at each point; if one has a norm but not necessarily an inner product then this is called a *Finsler manifold*. An example of a Finsler manifold encountered in these notes (other than a vector space with e.g. the L^p norm where $p \neq 2$) is given by the Hilbert metric on a convex set with the no-line property, which is *not* an ellipsoid. A Finsler structure allows one to define an actual metric in the sense of metric spaces of analysis. First, one defines the length of a smooth curve $\gamma : [a, b] \to M$, as above in Definition 23.3, by

$$l(\gamma) = \int_\gamma \mathrm{d}s \equiv \int_a^b ||\gamma'(t)|| \mathrm{d}t.$$

The distance between two points is then defined to be the infimum if the lengths of curves between the points. A *geodesic* is a smooth curve in M which locally minimizes length.

A basic theorem of differentiable geometry is that given the choice of a tangent vector $\mathbf{v}_p \in TM_p$, in the Finsler as well as the Riemannian case, there exists a unique geodesic tangent to this vector. We write T_1M for the *unit tangent bundle*, the tangent vectors of unit length. The *geodesic flow* of M is the flow g_t (actually on T_1M) defined by flowing at unit speed along this geodesic, and transporting the unit vector along the curve.

Remark 25.4. The existence of geodesics for the Riemannian case we encounter here is a consequence of the existence theorem of ordinary differential equations. For background see ??? Rather than getting into the (beautiful) differentiable geometry needed to treat this properly, we study concrete examples where the definitions can be given algebraically.

We recall from §23.1 that Möb($\mathbb{C}$) denotes the group of Möbius transformations on the extended complex plane $\widehat{\mathbb{C}}$; a matrix in $M \in PGL(2, \mathbb{C})$ or equivalently $PSL(2, \mathbb{C})$ (see Corollary 23.4) with $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ defines such a map by $f_M(z) = (az+b)/(cz+d)$. The *real Möbius transformations* are the subgroup $PGL(2, \mathbb{R})$ of $PGL(2, \mathbb{C})$ with real entries. Unlike $PGL(2, \mathbb{C})$ which is connected, $PGL(2, \mathbb{R})$ has two connected components. One of these is itself a group, those with determinant > 0; this defines $PGL^+(2, \mathbb{R})$ and Möb$^+(\mathbb{R})$. This subgroup of index two are the Möbius transformations which preserve the upper half plane $\mathbb{H}$, while those with negative determinant still preserve the real axis but switch the upper and lower half planes.

Now $PGL^+(2, \mathbb{R})$ is naturally isomorphic to $PSL(2, \mathbb{R})$, see Proposition 23.9 and Definition 23.3.

We recall that $\mathbb{H} = \{z = x + yi : y > 0\}$ is given the hyperbolic metric $d_{\mathbb{H}}(x, y) = c \cdot d_\gamma(x, y)$, where γ is the unique circle which meets the boundary $\mathbb{R}$ orthogonally while passing though x and y. Here, with $\eta, \xi \in \mathbb{R}$ denoting the boundary points of γ, then $d_{\mathbb{H}}(x, y) = d_\gamma(x, y) = |\log[\eta, x, y, \xi]|$. Since elements of $\text{Möb}^+(\mathbb{R})$ preserve angles, circles, the real line, and the cross ratio, they act as isometries of $\mathbb{H}$.

The map $M \mapsto f_M^*(i_i)$ from $PSL(2, \mathbb{R})$ to $T^1(\mathbb{H})$ is a bijection, and so we can identify $\mathbb{H}$ with $PSL(2, \mathbb{R})/PSO(2, \mathbb{R})$ via this choice of base point. It follows that the flow defined by E_t acting on the right on $PSL(2, \mathbb{R})$ is isomorphic to the geodesic flow on $T^1(\mathbb{H})$.

We have seen have seen above that $PSL(2, \mathbb{R})$ to the unit tangent bundle $T_1(\mathbb{H})$. We next consider three more right actions by the other one-parameter subgroups. by $\{E_t\}_{t\in\mathbb{R}}$

The unit tangent bundle of $\mathbb{H}$ can be identified with $PSL(2, \mathbb{R})$. This correspondence is easily described. Take as base point the unit vector i_i which is located at the point $i \in \mathbb{H}$ and points in the vertical direction; then, given $A \in SL(2, \mathbb{R})$, let $f_A^*(i_i)$ be the image of this vector by the derivative map of f_A, that is, it is the vector located at the point $f_A(i)$ which has been been rotated appropriately by the argument of the complex derivative. This image vector also has hyperbolic length one, as Möbius transformations are isometries for the hyperbolic metric; so this defines a map from $PSL(2, \mathbb{R})$ to the unit tangent bundle $T_1(\mathbb{H})$. The group Γ acts on $PSL(2, \mathbb{R})$ by left multiplication and one sees that $\Gamma\backslash PSL(2, \mathbb{R})$ is the unit tangent bundle of the surface $\Gamma\backslash\mathbb{H}$.

The geodesic flow on the surface is by definition the flow on this unit tangent bundle which moves a vector along its tangent geodesic at unit speed. Algebraically, this is given by right multiplication by $E_t \equiv \begin{bmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{bmatrix}$ on $\Gamma\backslash PSL(2, \mathbb{R})$. To understand this, note that this matrix is equivalent as a Möbius transformation to $\begin{bmatrix} e^t & 0 \\ 0 & 1 \end{bmatrix}$ which dilates the plane by the factor e^t, and hence moves the vector i_i up the imaginary axis at unit speed in the hyperbolic metric. The action on a general unit vector is then given by the conjugation by f_A^* which is a hyperbolic isometry, so this is indeed the geodesic flow. The unstable horocycle flow h_t^u is given by the right action of $H_t^u \equiv \begin{bmatrix} 1 & 0 \\ t & 1 \end{bmatrix}$; the stable flow acts by its transpose. As the names suggest, these preserve the unstable and stable horocycles (circles tangent to the boundary $\mathbb{R}$ of $\mathbb{H}$ which are the base points of the unstable and stable sets of the geodesic flow; for the point i_i, this "circle" being the line $y = 1$).

For the simplest example of a noncompact, finite area surface, see Fig.**??**; here (depicted in the disk model) Γ is a free group on two generators, these being two hyperbolic Möbius transformations, one which shoves the interior of the disk to the right and one which moves everything up; the curved quadrilateral in the center is a fundamental domain for this action. The left side is glued to the right, and the bottom to the top, so the resulting surface is a torus, just like for the usual gluings of a square, to get the quotient space $\mathbb{R}^2/\mathbb{Z}^2$, except that now the corner point gives

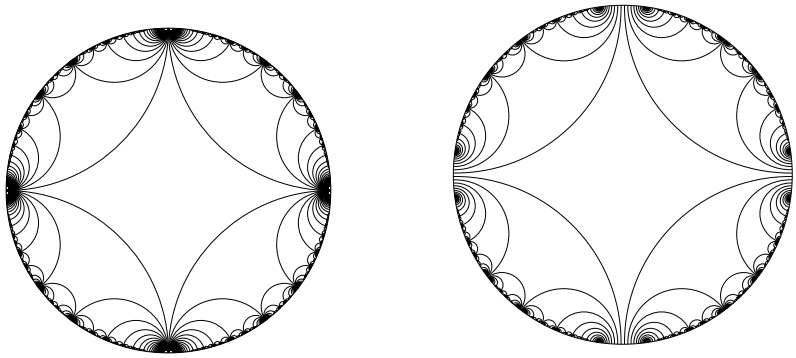


FIGURE 64. Covering space for punctured torus; after opening up the cusp.

a cusp, as it goes out to ∞ in the hyperbolic metric: this is a *punctured torus* (Fig. 64).

Classical results are:

Theorem 25.7. *The geodesic and horocycle flows g_t, h_t^u, h_t^s preserve Riemannian volume of the unit tangent bundle of the surface M. This measure is finite iff the surface area is finite. For this case, if M is compact (equivalently has no cusps) then:*
(i) g_t is ergodic, indeed is (finite entropy) Bernoulli (is measure-theoretically isomorphic to a Bernoulli flow);
(ii) h_t^u, h_t^s are uniquely ergodic, with entropy zero.

In the finitely generated, finite area case with cusps, all this is true except that h_t^u, h_t^s are only nearly uniquely ergodic; normalized Riemannian volume is the only nonatomic invariant probability measure if we disallow measures supported on horocycles tangent to cusps.

More interesting for us will be the infinite area case, where the cusp opens up to flare out in a hyperbolic trumpet, Fig. 65; we return to this below.

The flows g_t and h_t^u do not commute, but do satisfy the following commutation relation:

$$h_b g_a = g_a h_{e^{-s}b}.$$

In other words, the following diagram commutes:

FIGURE 65. Geodesic flow on punctured torus and on infinite area surface

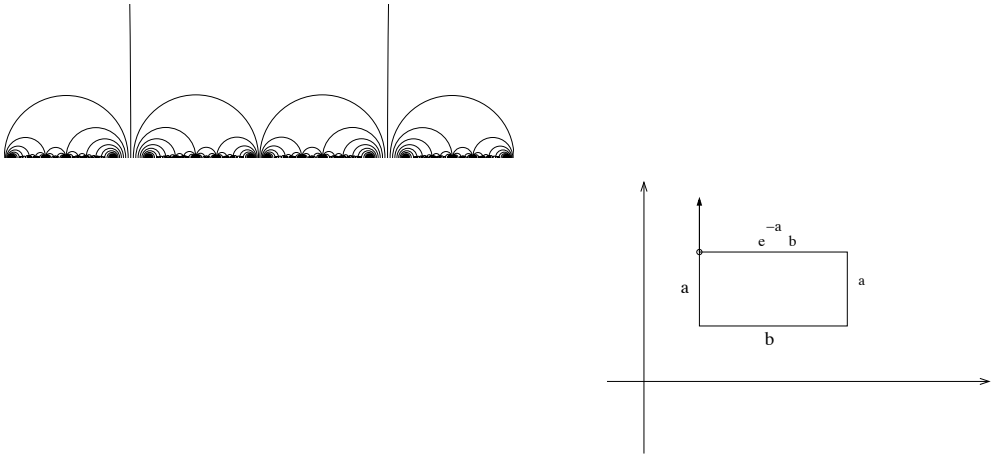


FIGURE 66. Opened cusp in upper half plane; geometric explanation of commutation relation.

$$\begin{array}{ccc} T^1(M) & \xrightarrow{h_{e^{-a}\cdot b}} & T^1(M) \\ {\scriptstyle g_a}\uparrow & & \uparrow{\scriptstyle g_a} \\ T^1(M) & \xrightarrow[h_b]{} & T^1(M) \end{array}$$

One can prove this algebraically, or see it geometrically in the upper half plane (Fig. 66).

Remark 25.5. Because of the commutation relation, the pair (geodesic flow, horocycle flow) gives an action of the $(ax + b)$-group (the real affine group) on $T^1(M)$. This

already hints that there might be a relation with fractal geometry, as fractal sets genarally exhibit symmetries with respect to both dilation and translation.

Observation: The commutation relation tells us that h_t^u is isomorphic to a speeded-up version of itself. An ergodic theorist immediately will recognise that this is very special, as the entropy of a sped-up transformation or flow multiplies by that factor, so in this case:

$$\text{entropy}((h^u)_t) = e^{-a} \cdot \text{ entropy}(h_t^u).$$

There are, thus, only two possiblities for the entropy of the flow $(h^u)_t$: 0, or ∞!

25.3. **Coding the modular flow.** Next we study the geodesic flow of a specific Riemann surface, the modular surface $PSL(2,\mathbb{Z})\backslash\mathbb{H}$.

This is a simple example from the algebraic point of view, as we simply consider the discrete subgroup of $PSL(2,\mathbb{R})$ with integer entries. In that sense it is analogous to the surface $\mathbb{R}^2/\mathbb{Z}^2$, (the square torus), where $\mathbb{R}^2$ is the group of isometeries of the plane. For a closer analogy

By the **modular flow** we mean the right action of $\{E_t\}_{t\in\mathbb{R}}$ on $\Gamma\backslash SL(2,\mathbb{R})$, where

$$E_t \equiv \begin{bmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{bmatrix}.$$

We quickly describe its other guises, as the geodesic flow on the modular suface, and as the Teichmüller flow of the torus. Once we have done this, we shall be free to designate this flow as g_t in all cases.

As noted above, $SL(2,\mathbb{Z})\backslash SL(2,\mathbb{R}) \cong PSL(2,\mathbb{Z})\backslash PSL(2,\mathbb{R})$, so the right actions of $\{E_t\}_{t\in\mathbb{R}}$ on these two spaces are isomorphic.

We quickly describe its other guises, as the geodesic flow on the modular suface, and as the Teichmüller flow of the torus. Once we have done this, we shall be free to designate this flow as g_t in all cases.

Noting that λM and M define the same map, then $M \mapsto f_M$ defines a homomorphism from $PGL(2,\mathbb{R}) \to \text{Möb}(\mathbb{R})$, and from $PSL(2,\mathbb{R}) \to \text{Möb}^+(\mathbb{R})$; one shows easily that these are isomorphisms.

More precisely, we consider one of the most important such surfaces, both historically and because of connections with number theory: the *modular surface*. We introduce this from the number theory side, first recalling some basic facts.

Every irrational $x \in (0,1)$ has a unique infinite continued fraction expansion

$$x = [n_0 n_1 \dots] = \cfrac{1}{n_0 + \cfrac{1}{n_1 + \cdots}}$$

with integers $n_i \geq 1$; rational numbers have (two) finite expansions. For example,

$$\frac{2}{3} = \frac{1}{3/2} = \frac{1}{1+1/2} = [1\,2] = [1\,1\,1].$$

Dynamics is brought into the picture by the **Gauss map** Φ, defined on the $[0,1]$ by $\Phi(0) = 0$ and $\Phi(x) = \{1/x\} \equiv 1/x - [1/x]$, the fractional part of $1/x$, for $x \neq 0$.

Writing $\Pi^+ = \Pi_0^\infty \mathbb{N}^*$for $\mathbb{N}^* = \{1, 2, \cdots\}$, with the left shift σ defined on $\underline{n}^+ = (.n_0 n_1 \dots)$ by $\sigma(\underline{n}^+) = (.n_1 n_2 \dots)$, then the map $\pi((.n_0 n_1 \dots)) = [n_0 n_1 \dots]$ conjugates σ on Π^+ to Φ on $(0,1) \setminus \mathbb{Q}$. Given $\underline{n}^+ \in \Pi^+$, we define for each $k \geq 0$ a fraction written in lowest terms $p_k/q_k = [n_0 n_1 \cdots n_k]$, so these rational numbers approximate $x = [n_0 n_1 \cdots]$.

Here are some natural (and very classical) questions about this expansion:
1/ what is the significance of a periodic point for the Gauss map, i.e. of a periodic continued fraction expansion?
2/ does a given digit $l = n_k$ occur with a definite frequency, Lebesgue-almost surely?
3/ does Lebesgue measure on $(0,1)$ converge under iteration to some Φ-invariant measure, absolutely continuous with respect to Lebesgue measure?
4/ If so, is this measure ergodic for the map?

Here, first, are some answers:
1/ As was known long ago, periodic and eventually periodic expansions correspond to quadratic irrational numbers (roots of quadratic polynomials) in $(0,1)$; for example $x = [111\cdots]$ satisfies $\{\frac{1}{x}\} = \frac{1}{x} - 1 = x$ so $x^2 + x - 1 = 0$ and $x = \frac{-1+\sqrt{5}}{2}$; thus $1 + x = 1.618\cdots$ is the golden number.
2/ This will be answered by the Birkhoff ergodic theorem, once we know 3/ and 4/!
3/ This was solved by Gauss; he showed that the probability measure with density $(\log 2(1+x))^{-1}$ on $[0,1]$ is Φ-invariant (as is easily verified, once one has the formula!) and that the iterates of Lebesgue measure converge to it.
4/ Ergodicity can be shown in several ways:
– by direct argument with distortion estimates, see Billingsley [Bil65] p. 40 ff.
– By the general Renyi-Bowen-Adler argument which works for expanding, countable branched maps of the interval with bounded distortion, see e.g. [Mañ87] p. 168.
– Via the connection between continued fractions and geodesics on the modular surface. Here ergodicity for the geodesic flow follows by the very general *Hopf argument*, see [Hop39], [Hop71], §39 avoiding all distortion estimates.

That there must be a link between continued fractions and geodesics is easy to believe (we associate a geodesic in the upper half plane to the continued fraction expansions of its two endpoints) but making this precise can be done in a variety of ways, some trickier than others; see the survey [KU07]. Our favorite approach is due to Arnoux [Arn94], and will be fundamental to other parts of these notes, as will be seen.

25.4. **Continued fractions.** First, a remark on notation: writing the continued fraction of an irrational $x \in (0,1)$ as

$$x = [n_0 n_1 \dots] = \cfrac{1}{n_0 + \cfrac{1}{n_1 + \cdots}},$$

we begin the expansion of x with n_0 rather than with the more traditional choice of n_1; this will agree with the usual shift notation of ergodic theory, where 0 indicates the coordinate of present time, and will be especially convenient below when we switch to the bilateral shift space.

The k^{th} **approximant** of $x = [n_0 n_1 \dots]$ is $p_k/q_k \equiv [n_0 \dots n_k]$, written in lowest terms; this satisfies the recurrence relations:

$$\begin{aligned} p_{k+1} &= p_{k-1} + n_{k+1} p_k \\ q_{k+1} &= q_{k-1} + n_{k+1} q_k \end{aligned} \tag{99}$$

where, to get started, we define $(p_{-2}, q_{-2}) = (1, 0)$ and $(p_{-1}, q_{-1}) = (0, 1)$, so $(p_0, q_0) = (1, n_0)$, $(p_1, q_1) = (n_1, 1 + n_0 n_1)$ and so on.

As is well known, this can be nicely expressed in matrix form. Defining for $m \geq 1$ $R_m = \begin{bmatrix} 0 & 1 \\ 1 & m \end{bmatrix}$, then we have

$$R_{n_0} R_{n_1} \cdots R_{n_k} = \begin{bmatrix} 0 & 1 \\ 1 & n_0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & n_1 \end{bmatrix} \cdots \begin{bmatrix} 0 & 1 \\ 1 & n_k \end{bmatrix} = \begin{bmatrix} p_{k-1} & p_k \\ q_{k-1} & q_k \end{bmatrix}. \tag{100}$$

We shall, in fact, use different matrices; defining

$$P = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \text{ and } Q = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix},$$

then one has, for k even,

$$P^{n_0} Q^{n_1} P^{n_2} \dots P^{n_k} = \begin{bmatrix} 1 & n_0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ n_1 & 1 \end{bmatrix} \cdots \begin{bmatrix} 1 & n_k \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} p_k & p_{k-1} \\ q_k & q_{k-1} \end{bmatrix}$$

and

$$Q^{n_0} P^{n_1} Q^{n_2} \dots Q^{n_k} = \begin{bmatrix} q_k & q_{k-1} \\ p_k & p_{k-1} \end{bmatrix},$$

while for k odd

$$P^{n_0} Q^{n_1} P^{n_2} \dots Q^{n_k} = \begin{bmatrix} p_{k-1} & p_k \\ q_{k-1} & q_k \end{bmatrix} \text{ and } Q^{n_0} P^{n_1} Q^{n_2} \dots P^{n_k} = \begin{bmatrix} q_{k-1} & q_k \\ p_{k-1} & p_k \end{bmatrix}.$$

Although at first glance this is slightly more complicated, we shall see below how naturally the matrices P, Q arise in the present context.

Recall that $SL(2, \mathbb{R})$ is the group of (2×2) real matrices with determinant one and $GL(2, \mathbb{R})$ those with determinant $\neq 0$. Restricting the entries to $\mathbb{Z}$ defines the subgroups $SL(2, \mathbb{Z})$ and $GL(2, \mathbb{Z})$; we note that $GL(2, \mathbb{Z})$ is the set of (2×2) integer matrices with determinant ± 1. The orthogonal group $SO(2, \mathbb{R}) \subseteq SL(2, \mathbb{R})$ consists of the matrices $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ with determinant one. We also shall encounter the projective linear group $PSL(2, \mathbb{R})$; here matrices A and λA (both in $SL(2, \mathbb{R})$) are identified for $\lambda \neq 0$, but since $\det \lambda A = \lambda^2 \det A = \lambda^2 = 1$, we have $PSL(2, \mathbb{R}) \cong SL(2, \mathbb{R}) / \pm I$, while $PSO(2, \mathbb{R}) \cong SO(2, \mathbb{R}) / \pm I$. We observe also that $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R}) \cong PSL(2, \mathbb{Z}) \backslash PSL(2, \mathbb{R})$, since we factor by $\pm I$ both above and below (see (**??**)). The connection with hyperbolic geometry comes from the fact that there is an identification of $PSL(2, \mathbb{R}) / PSO(2, \mathbb{R})$ with the upper half plane $\mathbb{H}$; this is explained in §25.5.

We write $\Gamma = SL(2, \mathbb{Z})$; this is the **modular group** (sometimes that name is used for $PSL(2, \mathbb{Z}) = SL(2, \mathbb{Z}) / \pm I$).

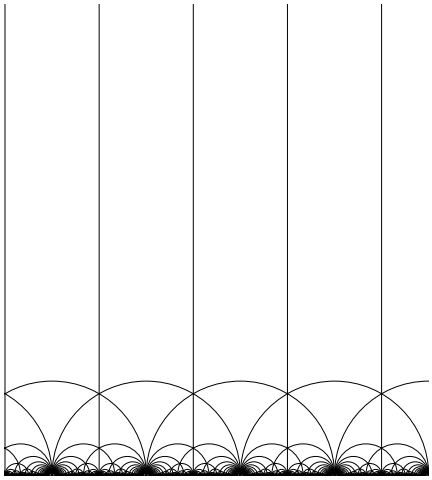


FIGURE 67. A fundamental domain for the action of the modular group Γ

25.5. **The modular flow, the geodesic flow and the Teichmüller flow.** The **modular space** is $PSL(2,\mathbb{Z})\backslash\mathbb{H}$; that is, it is the quotient space of the hyperbolic plane defined by the action on the left of the group $PSL(2,\mathbb{Z})$ (or equivalently $\Gamma = SL(2,\mathbb{Z})$) by Möbius transformations.

We claim that Γ is generated by by the matrices $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Since $\Gamma/\pm I$ is a subgroup of $PSL(2,\mathbb{R})$, which is isomorphic to the group of Möbius transformations $\text{Möb}^+(\mathbb{R})$ on $\mathbb{H}$, this subgroup $\text{Möb}^+(\mathbb{Z})$ of $\text{Möb}^+(\mathbb{R})$ is generated by the two maps $z \mapsto z+1$ and $z \mapsto -1/z$. A fundamental domain for this action is illustrated in Fig. 67.

To verify the claim and see exactly what group Γ is algebraically, recall from Lemma 19.1 that $SL(2,\mathbb{N})$ is the free semigroup on the two generators $P = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $Q = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. We have:

Proposition 25.8. *$SL(2,\mathbb{Z})$ is generated (not freely) by P, Q; $SL(2,\mathbb{Z})$ is freely generated by the two matrices $\widehat{J} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ and $K = \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix}$ which have orders 4 and 6 respectively, and so $SL(2,\mathbb{Z}) \cong \mathbb{Z}^4 * \mathbb{Z}^6$; we note that $P = \widehat{J}K^{-2}$ and $Q = \widehat{J}^{-1}K^2$*

Here $*$ indicates the free product of the two groups; one does not assume commutativity as with a usual product of groups! And "freely generated by" means there are no other relations. For a proof see [Mag74].

Below we also consider the space $\Gamma(2)\backslash\mathbb{H}$; this is a six-fold cover of the modular surface, with a fundamental domain for that action shown in Fig. 68.

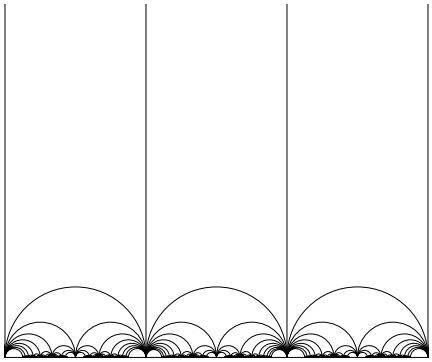


FIGURE 68. A fundamental domain for the action of the principal modular subgroup $\Gamma(2)$, of index 6 in Γ.

Next we decribe the connection with the Teichmüller flow. Given a matrix $B \in SL(2,\mathbb{R})$ with rows ${}^t\mathbf{v}_1, {}^t\mathbf{v}_2$, we send it by a map β to the parallelogram with vertex at the origin and with sides given by these vectors. This has area 1; it inherits an orientation from $\mathbb{R}^2$, which is positive since $\det(B) > 0$.

The collection of all such parallelograms defined up to translation is $T_1(\text{Teich})$, the unit tangent bundle of the Teichmüller space of the torus; for the Teichmüller space itself we consider the parallelograms up to rotations. The map β gives a one-to-one correspondence from $SL(2,\mathbb{R})$ to the positively oriented parallelograms with area 1 with a vertex at the origin. The matrix $-B$ gives the same parallelogram rotated by angle π, which is a translate of the original. Therefore we have:

Proposition 25.9. *The map $\beta : PSL(2,\mathbb{R}) \to T_1(\text{Teich})$ is a bijection.* □

The *modular space* of the torus is a factor space of the Teichmüller space; its unit tangent bundle $T^1(\text{Mod})$ is described by allowing all possible basis changes for the lattice. These are given by left actions of $PSL(2,\mathbb{Z})$. So we conclude that $T^1(\text{Mod})$ corresponds to $PSL(2,\mathbb{Z})\backslash PSL(2,\mathbb{R})$.

It remains to see what the right action of E_t does to the lattices. And as one immediately sees, acting by E_t applies a hyperbolic flow to the plane, so it expands the lattice exponentially in the direction of the $x-$axis, while contracting it in the $y-$direction.

This gives what is called the **Teichmüller flow of the torus**.

So in conclusion, these are isomorphic: the modular flow, i.e. right action of $\{E_t\}_{t\in\mathbb{R}}$ on $\Gamma(m)\backslash SL(2,\mathbb{R})$ or $PSL(2,\mathbb{Z})\backslash PSL(2,\mathbb{R})$; the geodesic flow on the modular surface; the Teichmüller flow of the torus.

25.6. **Arnoux' cross-section for the modular flow.** Now we come to the work of [Arn94] which plays a basic role below. An insight here is that rather than using the fundamental domain for the surface to find a cross-section for the geodesic flow, one should work directly in the unit tangent bundle, first finding a nice fundamental domain *there* from which the flow cross-section will be apparent.

The key idea of how to do this was inspired by Veech's analysis of the Teichmüller flow of surfaces of genus ≥ 2, see e.g. [Via06]; interpreting the modular flow as the Teichmüller flow leads to an especially transparent construction of a cross-section, with very nice combinatorial properties as we shall see.

One technical difference to Veech's work is that rather than employing Rauzy induction, Arnoux cuts down the interval alternately from the left and the right. This is a more symmetrical approach in the present setting.

We define a subset of $SL(2,\mathbb{R})$: $\mathcal{B}$ is the collection of matrices $B = \begin{bmatrix} a & c \\ -b & d \end{bmatrix}$ satisfying:

(1) $a, b, c, d \geq 0$
(2) $\det B = 1$
(3) $\mathcal{B}$ is a union of disjoint sets $\mathcal{B} = \mathcal{B}^{\mathbf{0}} \cup \mathcal{B}^{\mathbf{1}}$, where for $B \in \mathcal{B}^{\mathbf{0}}$, $0 < a < 1 \leq b$ and $d < c$, and for $B \in \mathcal{B}^{\mathbf{1}}$, $0 < b < 1 \leq a$ and $c < d$.

We say $B \in \mathcal{B}$ has **parity** $\epsilon = \mathbf{0}$ or $\epsilon = \mathbf{1}$ when it is in $\mathcal{B}^{\mathbf{0}}$ or $\mathcal{B}^{\mathbf{1}}$ respectively.

We have:

Proposition 25.10. *$\mathcal{B}$ is a fundamental domain for the left action of $SL(2,\mathbb{Z})$ on $SL(2,\mathbb{R})$. Also, $\mathcal{B}$ is a fundamental domain for the left action of $PSL(2,\mathbb{Z})$ on $PSL(2,\mathbb{R})$.*

Proof. Let $C \in SL(2,\mathbb{R})$, and let $\Lambda \subseteq \mathbb{C}$ be the lattice generated by the rows of C. We claim that there exists a unique matrix $B \in \mathcal{B}$ whose rows span the same lattice. But left multiplication of C by an element of $SL(2,\mathbb{Z})$ just changes basis in this lattice, so this will prove the proposition.

Our proof is geometrical. We construct a "Markov partition" by the following algorithm, depicted in Figure 69. (In fact these are actual Markov partitions, but for a *nonstationary* dynamical system; see [AF05]).

(i)Draw a closed horizontal line segment of length 2, centered at each point of the lattice Λ.

(ii)Extend a vertical line from each lattice point until it meets the horizontal segment.

(iii) There are three cases: the location where it meets is to the left of the lattice point, is at the point or is to the right. We consider the case where it lies to the left. Remove the rest of the segment, to the left of this hitting point. The length of this horizontal piece, to the left of the lattice point, defines the number a. Note that $a < 1$.

(iv) Extend the horizontal line to the right, until it hits the vertical segment. This length defines b. Note that $b \geq 1$.

(v) Finally, extend the vertical segment further upwards until it hits the horizontal segment. There are two cases: it hits in the interior of the segment or at an endpoint. (Geometrically, in the lattice, the two endpoints are identical.) We assume it is in

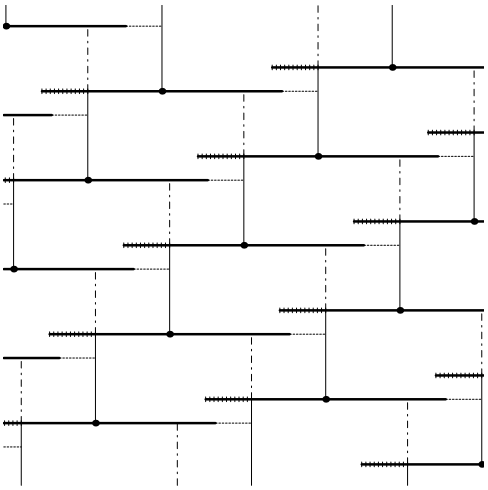


FIGURE 69. Construction of the Markov partition, given a lattice in the plane; here $a < b$ so $B \in \mathcal{B}^0$.

the interior. We call c the length of the vertical segment and d the distance along the vertical segment from the lattice point to where this segment was hit by the horizontal. Therefore $d \leq c$.

The case where the vertical segment hits a lattice point corresponds to $a = b = 1$, while the horizontal segment hitting a lattice point corresponds to $c = d$; both happen only for the square lattice $\mathbb{Z} + \mathbb{Z}$. All of these are ruled out by hypothesis.

Now since $\det(C) = 1$, the area of the parallelogram spanned by its rows is 1. This is a fundamental domain for the action of Λ on $\mathbb{C}$, so the torus (which is the quotient space) has area 1.

In this construction, we have drawn two boxes. Their union tiles the plane under translation by Λ (see Fig. 70), hence is also a fundamental domain for Λ. So their total area is 1.

We see from Figure 69 that d, c are the vertical sides of the boxes with bases a, b respectively. This completes the proof of the claim and the Proposition. □

As in [AF05] we define $\mathcal{B}_0^{\mathbf{0}}$ and $\mathcal{B}_0^{\mathbf{1}}$ to be the subsets of $\mathcal{B}$ with $b = 1$ and $a = 1$ respectively, and set $\mathcal{B}_0 = \mathcal{B}_0^{\mathbf{0}} \cup \mathcal{B}_0^{\mathbf{1}}$. Now under the projection of $SL(2, \mathbb{R})$ to $SL(2, \mathbb{Z})\backslash SL(2, \mathbb{R})$, $\mathcal{B}$ maps in a one-to-one way since it is a fundamental domain. Hence its subset $\mathcal{B}_0$ also maps injectively, thus we can naturally identify $\mathcal{B}_0$ with $SL(2, \mathbb{Z}) \cdot \mathcal{B}_0 \subseteq SL(2, \mathbb{Z})\backslash SL(2, \mathbb{R})$. We let T denote the return map of the geodesic flow on $SL(2, \mathbb{Z})\backslash SL(2, \mathbb{R})$ to $\mathcal{B}_0$ (with this identification).

As above, $\widehat{\Pi}$ with map $\widehat{\pi}$ denotes the two-point extension of the bilateral continued fraction shift map.

We now show how $(\widehat{\Pi}, \widehat{\sigma})$ factors onto $(\mathcal{B}_0, T)$.

Proposition 25.11. *$\mathcal{B}_0$ is a cross-section for the modular flow. The transformation $(\widehat{\Pi}, \widehat{\sigma})$ is semiconjugate to the return map T of the flow to $\mathcal{B}_0$ off of an invariant nowhere dense set, which has measure zero for the natural extension of Gauss measure*

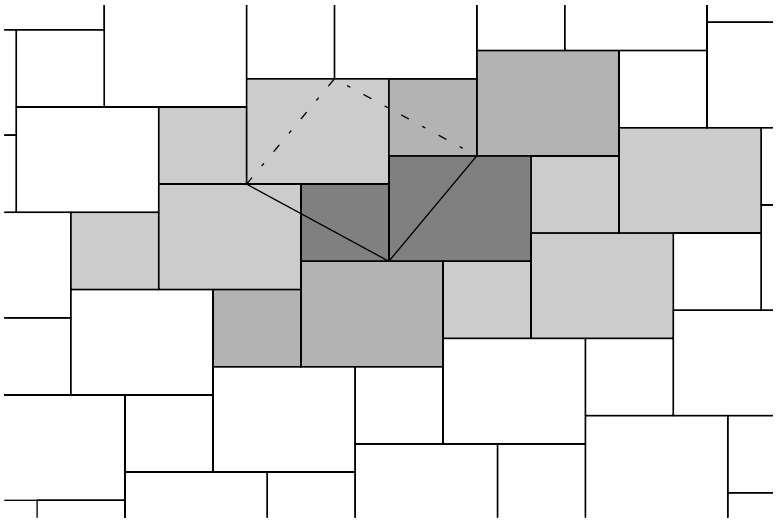


FIGURE 70. Two fundamental domains for the lattice Λ: the parallelogram and the two-box partition. In the above picture, the horizontal axis gives the expanding direction of the Teichmüller flow.

and corresponds to rational points. The successive returns of $B_0 \in \mathcal{B}_0$ *to this cross-section are* $B_1, \dots, B_i$ *with* $B_{i+1} = A_i B_i D_i$ *where* $B_i = \begin{bmatrix} a_i & c_i \\ -b_i & d_i \end{bmatrix}$, $D_i = \begin{bmatrix} \lambda_i & 0 \\ 0 & \lambda_i^{-1} \end{bmatrix}$ *and*
for parity **0**:
$a_i = [n_i n_{i+1} \dots], b_i = 1, d_i/c_i = [n_{i-1} n_{i-2} \dots]$, *and* $\lambda_i = 1/a_i$, *and* $A_i = \begin{bmatrix} 1 & 0 \\ n_i & 1 \end{bmatrix}$,
for parity **1**:
$b_i = [n_i n_{i+1} \dots], a_i = 1, c_i/d_i = [n_{i-1} n_{i-2} \dots]$, *and* $\lambda_i = 1/b_i$ *and* $A_i = \begin{bmatrix} 1 & n_i \\ 0 & 1 \end{bmatrix}$.

The matrix B_0 *is a is a unit tangent vector to the modular space, and the geodesic tangent to it goes to* $\xi = -a_0/b_0$ *at* $+\infty$ *and* $\eta = c_0/d_0$ *at* $-\infty$. *For parity* **0**, $\xi \in (-1, 0)$ *and* $\eta \in (1, +\infty)$, *and for parity* **1**, $\xi \in (-\infty, -1)$ *and* $\eta \in (0, 1)$.

Proof. First we show it is a cross-section. By Proposition 25.10, $\mathcal{B}$ is a fundamental domain for the left action of $SL(2, \mathbb{Z})$ on $SL(2, \mathbb{R})$, hence the modular flow corresponds to the flow on this fundamental domain, so it suffices to check the statement there. Now if for example $B_0 \in \mathcal{B}_0^{\mathbf{0}}$, then $0 < a < 1 \leq b$ and we apply the matrix E_t on the right, considering how this acts on the parallelograms defined by the row vectors: geometrically the direction of the $x-$axis is dilated until the smaller of the two boxes has length 1. That is, taking $t_0 = \log a_0^{-1}$, then $B_0 E_{t_0} \in \mathcal{B}_0^{\mathbf{1}}$. Hence $\mathcal{B}_0$ is a cross-section for the modular flow.

Next, let $B_0 \in \mathcal{B}_0^{\mathbf{0}}$. We define a sequence of positive integers $\underline{n} = (\dots n_0 n_1 \dots)$ from the continued fraction expansions $[n_0 n_1 \dots] = a_0$ and $[n_{-1} n_{-2} \dots] = d_0/c_0$.

From the sequence $\underline{n}$, we define matrices A_k, B_k, D_k for $k \in \mathbb{Z}$ as above, taking the sequence (A_k) to have parity **0**.

Now consider the orbit of B_0 under the geodesic flow on the covering space $\mathbb{H}$, $\{B_0E_t : t \in \mathbb{R}\}$. At time $t_0 = \log(1/a_0)$, the boxes determined by B_0 have been rescaled by $D_0 \equiv D_{t_0}$ so the new left-hand base length is equal to one. We see that the matrix $B_1 = A_0(B_0D_0)$ is equivalent to B_0D_0 modulo the left action of $SL(2,\mathbb{Z})$, and is in the set $\mathcal{B}_{0,\mathbf{1}}$. This represents the first return to our cross-section. Therefore the return sequence is: $\ldots B_0 \mapsto B_1 \mapsto B_2 \mapsto \ldots$, which determines the sequence $\underline{n}$.

The image of $\widehat{\Pi}$ is all $B \in \mathcal{B}_0$ with nonterminating continued fraction expansions, i.e. with irrational ratios of a to b, c to d; this set is a dense G_δ with full Gauss measure, and is invariant for the return map.

Finally, the asymptotics in $\mathbb{H}$ of the geodesic tangent to B_0 is seen via the map $M \mapsto F_M(I_i)$, applying the matrix E_t on the right; this finishes the proof. □

26. Nonlinearity: Shub's theorem

Before we study the smooth conjugacy of nonlinear hyperbolic maps of the interval, where a projective metric proof of Ruelle's version of the Perron Frobenius Theorem will play a key role, we consider the question of topological congugacy and the wonderful theorem on expanding maps of the circle due to Shub.

The proof given here of Shub's theorem [Shu85] is based on [Nit71], with some changes in the organization and proofs of the lemmas.

As before we write $\mathbb{T}^1$ for the circle $\mathbb{R}/\mathbb{Z}$, and $\pi : \mathbb{R} \to \mathbb{T}^1$ for the projection map $\pi : x \mapsto x(\mathrm{mod}\, 1)$. A fundamental domain for the action of $\mathbb{Z}$ on $\mathbb{R}$ is $[0,1)$, so we embed $\mathbb{T}^1$ in $\mathbb{R}$ as $[0,1) \subseteq \mathbb{R}$. So it will make sense to speak of the derivative of a map $f : \mathbb{T}^1 \to \mathbb{T}^1$ via that embedding.

Definition 26.1. Let $f : \mathbb{T}^1 \to \mathbb{T}^1$ be differentiable. We say f is **expanding** iff $\exists \lambda$ such that the derivative Df satisfies $|Df(x)| > \lambda > 1$ for all x.

Definition 26.2. Given $f : \mathbb{T}^1 \to \mathbb{T}^1$ continuous, a function $\bar{f} : \mathbb{R} \to \mathbb{R}$ is a **lift** of f iff $\bar{f}$ is continuous and $\pi \circ \bar{f} = \bar{f} \circ \pi$, i.e. the following diagram commutes:

$$\begin{array}{ccc} \mathbb{R} & \xrightarrow{\bar{f}} & \mathbb{R} \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\ \mathbb{T}^1 & \xrightarrow{f} & \mathbb{T}^1 \end{array}$$

We need the following basic result from topology, see e.g. [Arm83] p. 97:

Lemma 26.1. *Given a continuous map $f : \mathbb{T}^1 \to \mathbb{T}^1$ and writing $p = f(0)$, then for any choice of $\tilde{p} \in \pi^{-1}(p)$, there exists a unique lift $\bar{f}$ such that $\bar{f}(0) = \tilde{p}$.*

(Here we write 0 both for the point in $\mathbb{T}^1$ and in $\mathbb{R}$, by the imbedding of $[0,1)$ in $\mathbb{R}$.)

Definition 26.3. Let $f : \mathbb{T}^1 \to \mathbb{T}^1$ be continuous and let $\bar{f}$ be a lift of f. The **degree** of f is $\bar{f}(x+1) - \bar{f}(x)$.

For this to make sense we need:

Lemma 26.2. *For $f : \mathbb{T}^1 \to \mathbb{T}^1$ continuous,*

(i) For any fixed x, $\bar{f}(x+1) - \bar{f}(x) \in \mathbb{Z}$.
(ii) This number does not depend on the point x.
(iii) Furthermore, it is independent of the choice of lift.
(iv) For any $n \in \mathbb{Z}$, $\bar{f}(x+n) - \bar{f}(x) = n \cdot \deg f$.

Proof. (i): Now $\pi \circ \bar{f} = f \circ \pi$ and $\pi(x) = \pi(x+1)$. So $f \circ \pi(x) = f \circ \pi(x+1)$ but the left side is $\pi \circ \bar{f}(x)$ and the right side is $\pi \circ \bar{f}(x+1)$. Thus $\bar{f}(x)$ and $\bar{f}(x+1)$ are in the same preimage of a point by π so they differ by an integer k.
(ii): The function $g(x) = \bar{f}(x+1) - \bar{f}(x)$ is continuous and integer-valued hence is constant.
(iii): If $\tilde{f}$ is another lift, then $\pi \circ \bar{f}(x) = f \circ \pi(x) = \pi \circ \tilde{f}(x)$, so $\bar{f}(x) = \tilde{f}(x) + k$ for some $k \in \mathbb{Z}$; again this is constant by continuity, and so $\bar{f}(x+1) - \bar{f}(x) = \tilde{f}(x+1) - \tilde{f}(x)$.
(iv): From the definition this holds for $n = 1$; by induction it is true for $n \in \mathbb{Z}$. □

We shall also need:

Lemma 26.3.
(i) If $\bar{f}$ is continuous and $\bar{f}(x+1) - \bar{f}(x) = n$, then $\bar{f}$ is a lift of some $f : \mathbb{T}^1 \to \mathbb{T}^1$ of degree n.
(ii) If $\deg f = k$ and $\bar{f}$ is invertible, then $\bar{f}^{-1}(x+k) = \bar{f}^{-1}(x) + 1$.

Proof. (i) Defining for $x \in [0,1]$ $f(x) = \pi(\bar{f}(x))$, then $f(0) = f(1)$, so f is a well-defined continuous map of $\mathbb{T}^1$ and $\pi \circ \bar{f} = f \circ \pi$.
(ii) We have $\bar{f}(x+1) = \bar{f}(x) + k$. Applying $\bar{f}^{-1}$ on both sides, and writing $\bar{f}(x) = \tilde{x}$, then $\bar{f}^{-1}(\tilde{x}) + 1 = x + 1 = \bar{f}^{-1}(\bar{f}(x) + k) = \bar{f}^{-1}(\tilde{x} + k)$. On the other hand, $x = \bar{f}^{-1} \circ \bar{f}(x) = \bar{f}^{-1}(\tilde{x})$ so $x + 1 = \bar{f}^{-1}(\tilde{x}) + 1$. □

Let

$$E = \{\text{lifts } \bar{\alpha} \text{ of maps } \alpha : \mathbb{T}^1 \to \mathbb{T}^1 \text{ of degree } 1\}$$

and

$$E_0 = \{h : \mathbb{R} \to \mathbb{R} \text{ continuous and periodic with period } 1\}.$$

For $f, g \in E$, we define $d(f,g) = \sup_{x \in \mathbb{R}} \{|f(x) - g(x)|\}$. We write **id** for the identity map $x \mapsto x$ on $\mathbb{R}$.

Lemma 26.4. *(E, d) is a complete metric space.*

Proof. In fact, $E = E_0 + \mathbf{id}$, so E is an affine space which is isometric to the Banach space E_0 (with sup norm). □

Remark 26.1. The point is that, although the sup norm is infinite on E itself, this set is a perfectly nice metric space via this identification. A similar phenomenon happens in proofs of the stable manifold theorem, see [Shu87] where the relevant space of maps is a Banach manifold.

Now assume we are given f expanding on $\mathbb{T}^1$ and g with degree the same as f.

We define an operator $T : E \to E$ by $T(\bar{\alpha}) = \bar{f}^{-1} \circ \bar{\alpha} \circ \bar{g}$. This is illustrated in the diagram:

$$\begin{array}{ccc} \mathbb{R} & \xrightarrow{\bar{g}} & \mathbb{R} \\ \Big\downarrow{\scriptstyle T(\bar{\alpha})} & & \Big\downarrow{\scriptstyle \bar{\alpha}} \\ \mathbb{R} & \xrightarrow{\bar{f}} & \mathbb{R} \end{array}$$

To make sure this makes sense we need:

Lemma 26.5. *If f is expansive then $\bar{f}$ is $1-1$ and onto $\mathbb{R}$. If $\bar{\alpha} \in E$ then so is $T(\bar{\alpha})$.*

Proof. The map π is locally an isometry, so the derivatives are equal of f and $\bar{f}$; that is, $D\bar{f}(x) = Df(\pi(x)$ for any $x \in \mathbb{R}$, and so $|Df(x)| > \lambda > 1$ which immediately implies this is a bijection on $\mathbb{R}$ (much less would give this; it is enough that the derivative be nonzero).

Since $\bar{f}$ is invertible, T is well-defined. We claim that if $\bar{\alpha}$ is in E then $T(\bar{\alpha})$ is also in E. Writing $\tilde{\alpha} = T(\bar{\alpha})$, by Lemma 26.3(i), it is enough to show that $\tilde{\alpha}(x+1) = \tilde{\alpha}(x)+1$. Now $\tilde{\alpha}(x+1) = T(\bar{\alpha})(x+1) = \bar{f}^{-1} \circ \bar{\alpha} \circ \bar{g}(x+1) = \bar{f}^{-1} \circ \bar{\alpha}(\bar{g}(x) + \deg g) = \bar{f}^{-1}(\bar{\alpha}(\bar{g}(x)) + \deg g) = \bar{f}^{-1}(\bar{\alpha}(\bar{g}(x))) + 1 = \tilde{\alpha}(x) + 1$. □

Lemma 26.6. *The map T is a contraction on the complete metric space E.*

Proof. Given $\bar{\alpha}, \bar{\beta} \in E$, $\sup_{x\in\mathbb{R}} |\bar{f}^{-1}(\bar{\alpha}(\bar{g}(x))) - \bar{f}^{-1}(\bar{\beta}(\bar{g}(x)))| = \sup_{w\in\mathbb{R}} |\bar{f}^{-1}(\bar{\alpha}(w) - \bar{f}^{-1}(\bar{\beta}(w))| \le \lambda^{-1} \sup_{w\in\mathbb{R}} |\bar{\alpha}(w) - \bar{\beta}(w)| = \lambda^{-1} d(\bar{\alpha}, \bar{\beta})$. □

And now of course we will apply:

Lemma 26.7. *Let T be a contraction on a complete metric space. Then T has a unique fixed point.* □

Theorem 26.8. *(Shub* [Shu85]*) Let f be an expanding map of the circle (so with degree $\deg f \ge 2$ or ≤ -2), and let g be a continuous map of the circle with the same degree as f. Then there exists a unique continuous map h of degree 1 for which the following diagram commutes:*

$$\begin{array}{ccc} \mathbb{T}^1 & \xrightarrow{g} & \mathbb{T}^1 \\ \Big\downarrow{\scriptstyle h} & & \Big\downarrow{\scriptstyle h} \\ \mathbb{T}^1 & \xrightarrow{f} & \mathbb{T}^1 \end{array}$$

Proof. We know there is a unique fixed point $\bar{h}$ in E; from the equation $T(\bar{h}) = \bar{h}$, this diagram commutes:

$$\begin{array}{ccc} \mathbb{R} & \xrightarrow{\bar{g}} & \mathbb{R} \\ \Big\downarrow{\scriptstyle \bar{h}} & & \Big\downarrow{\scriptstyle \bar{h}} \\ \mathbb{R} & \xrightarrow{\bar{f}} & \mathbb{R} \end{array}$$

Since $\bar{h} \in E$, it is the lift of a degree one map h of the circle, and the other diagram commutes as well. □

Corollary 26.9. *Let f, g both be expanding maps of the same degree. Then there exists a unique conjugacy h as in the diagram.*

Proof. Switching the roles between g and f, we stick the two diagrams together:

$$\begin{array}{ccc} \mathbb{R} & \xrightarrow{\bar{f}} & \mathbb{R} \\ \downarrow \tilde{h} & & \downarrow \tilde{h} \\ \mathbb{R} & \xrightarrow{\bar{g}} & \mathbb{R} \\ \downarrow \bar{h} & & \downarrow \bar{h} \\ \mathbb{R} & \xrightarrow{\bar{f}} & \mathbb{R} \end{array}$$

Looking at the composition, we have

$$\begin{array}{ccc} \mathbb{R} & \xrightarrow{\bar{f}} & \mathbb{R} \\ \downarrow \bar{h}\circ\tilde{h} & & \downarrow \bar{h}\circ\tilde{h} \\ \mathbb{R} & \xrightarrow{\bar{f}} & \mathbb{R} \end{array}$$

but there is a unique fixed point for this operator T (with $g = f$ itself) and since the identity map on $\mathbb{R}$ works, we must have $\bar{h} \circ \tilde{h} = \mathbf{id}_{\mathbb{R}}$. Hence $\bar{h}$ is invertible. Since it is in E, it is the lift of a map h of degree one which is also invertible. □

26.1. **Some consequences.** We now harvest some immediate consequences, each of which gives an early taste of much more general theorems from hyperbolic dynamics. See for example [Nit71], [Shu87], [KH95] as well as [Shu85].

There exists a particularly nice choice for a representative of an expanding map in the equivalence class of topological conjugacy:

Corollary 26.10. *(Classification of expanding maps of the circle) An expanding map f of the circle with degree d is topologically conjugate to the linear map $x \mapsto dx (mod 1)$.* □

Since a map g close to f will also be expanding, and since their lifts will stay close hence the degree will also be preserved, we have:

Corollary 26.11. *(Structural stability of expanding maps of the circle) Given an expanding map f of the circle, there exists a neighborhood $\mathcal{U}$ of f in the $\mathcal{C}^1-$ topology such that every $g \in \mathcal{U}$ is topologically conjugate to f.* □

Remark 26.2. $E^d = \{$lifts of maps of degree $d\}$ is a Banach manifold, isometric to E_0. Proof: $\bar{f}(x+1) = \bar{f}(x) + d$, so $\bar{f} - \bar{g} \in E_0$.

A weaker (but important) statement is:

Corollary 26.12. *(Openness for expanding maps of the circle) In the Banach manifold of all smooth maps of the circle of degree d, the expanding maps are an open subset.* □

A further immediate corollary of the proof is:

Corollary 26.13. *Taking $f = g$ expanding, the operator T on the Banach manifold E is hyperbolic, with a unique contractive fixed point (the identity map on $\mathbb{R}$).* □

27. The Thermodynamic Formalism

We have above discussed Parry measures for subshifts of finite type, as well as related invariant measures for adic transformations. These can be understood intuitively with the help of geometric models, where the measures are Lebesgue measure (length) on intervals, for a one-sided shift space, or two-dimensional Lebesgue measure (area) on boxes, for the two-sided (bilateral) case.

Let us recall the formula: Given a $(d \times d)$ nonnegative $0-1$ primitive matrix M, with right Perron-Frobenius (column) eigenvector $\mathbf{w}$ and left row eigenvector vector $\mathbf{v}^t$, with eigenvalue $\lambda > 1$, normalized so $\mathbf{w} \in \Delta$ i.e. $||\mathbf{w}|| = \sum |w_i| = 1$ and $\mathbf{v} \cdot \mathbf{w} = 1$, then

$$\mu([x_{-n-1} \dots x_1 . x_0 \dots x_n] = \lambda^{-2n-2} \mathbf{v}_a \mathbf{w}_b$$

where $a = x_{-n-1}$, $b = x_n$.

This is the area of the box labelled by this cylinder set, and is the product of the lengths of the two sides,

$$\mu^+([.x_0 \dots x_n] = \lambda^{-n-1} \mathbf{w}_b$$

and

$$\mu^-([.x_{-n-1} \dots x_1] = \lambda^{-n-1} \mathbf{v}_a,$$

corresponding to cylinder sets in Σ_M^+ and Σ_M^- respectively; these measures are invariant not for the shift map but for respective adic transformations on these one-sided shift spaces.

In the nicest situation, an Anosov map on the $2-$torus, these boxes are actual geometrical rectangles in a surface. See Figure 69 for the torus case. Nearly as well-behaved is the special Markov partition for a pseudoAnosov map on a Riemann surface given by the Veech rectangles; here the adic transformation is simply the return map of the vertical (horizontal) flow, and is an interval exchange transformation (for the torus, it is an exchange of two intervals and hence a circle rotation); this map tells us how to glue the boxes to recover the surface.

Now a Markov measure has the property of depending only on the past digit. A general continuous (i.e. non-atomic) measure depends on more, perhaps all, past digits, in a measurable way.

The geometric models for the Markov maps of an interval are linear expansions, as in Fig. 17 or 48; a pseudoAnosov map can be thought of as the product of two such maps.

We want to extend this study to general hyperbolic $\mathcal{C}^{1+\alpha}$ diffeomeorphisms of a compact manifold, and it is precisely for this reason that the *Thermodynamic Formalism* was developed by Ruelle, Sinai and Bowen.

Realizing that the Parry measure is defined as the product of two eigenvectors, which are dual to each other (being a row and an column vector), it is natural to start with the collection of all signed measures as the dual space of the space of continuous functions on our compact metric space X, for instance $X = \Sigma_M^+$. Then

considering the duality between $\mathcal{C}(\Sigma_M^+)$ and $\mathcal{M}_X$, perhaps we can find an appropriate linear operator $\mathcal{L}$, on $\mathcal{C}(\Sigma_M^+)$, defined from our dynamics, with dual operator $\mathcal{L}^*$, with eigenvectors g (a continuous function) and ν (a measure) such that their product, $\mu = g\nu$, will be the invariant probability measure we seek.

That is the idea, and what is remarkable is that this all works out, and so nicely at that!

The first step is to prove an appropriate analogue of the Perron-Frobenius theorm. Here there is a restriction to Hölder functions; in terms of the measurable dependence on the past referred to above, this is a strong enough condition to give uniqueness of the eigenvectors, and corresponds to the $\mathcal{C}^{1+\alpha}$ condition for maps: given a map f of an interval, the corresponding *potential function* encountered below will be $\varphi = -\log|Df|$, which is indeed α-Hölder.

To greatly simplify the technical value of the Hölder condition, essentially "hyperbolicity plus Hölder gives a geometric series". Having this convergent series is crucial in the theory. As one might guess, sometimes these conditions can be weakened to some other convergent series; this is a challenging and active field of research!

27.1. **The Ruelle Perron-Frobenius theorem.** In this section we present a projective metric proof, due to Ferrero and Schmidt, of the Ruelle Perron-Frobenius theorem. As just indicated, this is a key part of the *thermodynamical formalism* of Ruelle, Sinai and Bowen; *why* this is both natural and useful will become clearer in sections to follow, when we discuss nonlinear maps.

The Ruelle operator. We begin by considering a general real-valued continuous *potential function* φ. Then, letting $\varphi \in \mathcal{C}(\Sigma_A^+)$, the **Ruelle operator** $\mathcal{L}_\varphi : \mathcal{C}(\Sigma_A^+) \to \mathcal{C}(\Sigma_A^+)$ is defined by:

$$(\mathcal{L}_\varphi f)(\underline{x}) = \sum_{w \in \sigma^{-1}(\underline{x})} e^{\varphi(w)} f(w).$$

Thus, for the n^{th} iterate of the operator, $\mathcal{L}_\varphi^n$, the value at $\underline{x}$ is collected from level n of the tree of preimages of $\underline{x}$, with weights e^φ along each branch.

The dual operator $\mathcal{L}_\varphi^*$ acts on the collection $\mathcal{M}$ of Borel measures on Σ_A^+ by the adjoint equation: writing $\langle m, f\rangle = \int f \mathrm{d}m$ for the pairing betwen measures and continuous functions, then for $f \in \mathcal{C}(\Sigma_A^+)$ and $m \in \mathcal{M}$, $\langle \mathcal{L}_\varphi^* m, f\rangle \equiv \langle m, \mathcal{L}_\varphi f\rangle$. That is,

$$(\mathcal{L}_\varphi^* m)(f) \equiv m(\mathcal{L}_\varphi f) \equiv \int_{\Sigma_A^+} (\mathcal{L}_\varphi f) \mathrm{d}m.$$

A special case is the *normalized* situation where we have a potential function ψ such that $p = e^\psi$ gives a probability weighting, i.e. so that $\sum_{\{\sigma(w)=\underline{x}\}} p(\underline{w}) = \sum_{\{\sigma(\underline{w})=\underline{x}\}} e^{\psi(\underline{w})} = 1$. Then we have

$$(\mathcal{L}_\psi f)(\underline{x}) = (\mathcal{L}_{\log p} f)(\underline{x}) = \sum_{\sigma(\underline{w})=\underline{x}} p(\underline{w}) f(\underline{w}).$$

Equivalently, $\mathcal{L}_{\log p}(1) = 1$ for the constant function 1. In this normalized case, if μ is a probability measure such that $\mathcal{L}_{\log p}^*(\mu) = \mu$, we say that μ is a *p-balanced measure*.

(The original term is *g-measure*, where the function is denoted by g rather than p; the concept is due to M. Keane; see [Kea72], [PP90]). Then the Ruelle operator and its dual give the analogue of a stochastic matrix, acting on column and row vectors respectively; see ??? below.

In the next proposition we give some ways of understanding the meaning of this mysterious formalism. Later we explain the connection to Markov shifts and the Shannon-Parry measure, which will clarify things further.

References for these results are [Kea72], [Led74], see also [PP90]).

Proposition 27.1.
(a)The dual operator $\mathcal{L}_\varphi^*$ is, equivalently, defined by its action on point masses: for $\underline{x} \in \Sigma_A^+$ with preimages $\underline{y}, \underline{w}$,

$$\mathcal{L}_\varphi^*(\delta_{\underline{x}}) = e^{\varphi(\underline{y})}\delta_{\underline{y}} + e^{\varphi(\underline{w})}\delta_{\underline{w}}.$$

(b) Parts (i) and (ii) are equivalent:
(i) ν is an eigenmeasure with eigenvalue λ, i.e. $\mathcal{L}_\varphi^*(\nu) = \lambda\nu$
(ii)

$$\frac{d\nu \circ \sigma}{d\nu} = \lambda e^{-\varphi}.$$

(c) Any invariant measure μ on Σ_A^+ is an eigenmeasure for a unique normalized measurable potential $\varphi = \log p$, with

$$\frac{1}{p} \equiv e^{-\varphi} = \frac{d\mu \circ \sigma}{d\mu}.$$

Proof. The equation in (a) follows directly from the definition; the converse holds since linear combinations of point masses are weak-$*$ dense in the space of measures. □

??Proof???

27.2. **Matrix examples.** We examine the connection with Markov shifts, §15.

27.3. **Hölder functions and the Ruelle operator.** Of special interest will be the case where our potential function φ is such that, as in $(b), (c)$ above, there exists an eigenmeasure ν with eigenvalue λ, with $\nu \circ \sigma$ locally absolutely continuous with respect to ν_φ. This will be guaranteed by a Hölder condition, as we now explain.

Given a $(d \times d)$ matrix A with enties 0 and 1, as before $\Sigma_A^+ \subseteq \Sigma^+ = \Pi_0^\infty\{0, \dots, d-1\}$ is the subshift of finite type with transitions those allowed by A, and with left shift map σ. Recall that the collection of all thin $k-$cylinder sets $[x_0 \dots x_n]$ is denoted by $\mathbb{C}_0^k$. We also define the whole space, Σ_A^+, to be the only "(-1)-cylinder set"; thus $\mathbb{C}_0^{-1} = \{\Sigma_A^+\}$. The space Σ_A^+ has been given the product topology; this is metrizable, and we shall use the metric

$$d(\underline{x}, \underline{y}) = \begin{cases} 1 & \text{if } x_0 \neq y_0 \ , \\ 2^{-n} & \text{if } x_0 = y_0 \text{ and } n = \inf\{n > 0 : \ x_n \neq y_n\}. \end{cases}$$

The Borel $\sigma-$algebra of Σ_A^+ is denoted by $\mathcal{B}$. We write $\mathcal{C}(\Sigma_A^+)$ for the set of continuous real-valued functions on Σ_A^+.

For $k \geq 0$ (recalling that we have set $\mathbb{C}_0^{-1} = \Sigma_A^+$), we define the k^{th} **variation** of $f \in \mathcal{C}(\Sigma_A^+)$,

$$\text{var}_k(f) = \sup\{|f(\underline{x}) - f(\underline{y})| \ : \ \underline{x}, \underline{y} \in C \in \mathbb{C}_0^{k-1}\}.$$

For $\alpha \in (0,1)$, we define the class of functions whose variation is exponentially small as a function of k, for base α:

$$\mathcal{H}_\alpha \equiv \{f \ : \ \exists \ c > 0 \text{ with } \text{var}_k f \leq c\alpha^k\} \text{ for all } k \geq 0\}.$$

These are the Hölder functions with exponent $-\log\alpha/\log 2$ with respect to the metric d. We write $||\cdot||_\infty$ for the sup norm on $\mathcal{C}(\Sigma_A^+)$, and define a norm on $\mathcal{H}_\alpha \subseteq \mathcal{C}(\Sigma_A^+)$ by

$$||f||_\alpha \equiv ||f||_\infty + c$$

where c is the inf of the possible Hölder constants for that exponent (or equivalently, for that base α); thus,

$$c = \sup_k\{\text{var}_k f \cdot \alpha^{-k}\}.$$

We define $\mathcal{H}_\alpha^b$ to be the subset of $\mathcal{H}_\alpha$ with Hölder constant $c \leq b$.

For a potential $\varphi \in \mathcal{C}(\Sigma_A^+)$ we define the *pressure*

$$P(\varphi) = \sup\left\{\int \varphi \ dm + H(m)\right\}$$

where the sup is taken over the collection of invariant probability measures and $H(m)$ denotes the entropy of the map T when the shift space is given that measure. (We remark that, for instance, a normalized potential φ always has pressure $P(\varphi) = 0$). If an invariant measure m is ergodic and is such that this sup is attained there, one says that m is an **equilibrium state** for that potential. A main theorem of the subject is that for Hölder potentials, there exists a unique equilibrium state (Theorem 1.22 in [Bow75]). We recall how this measure, μ, is produced.

For the case of a Hölder potential φ, it is known that one can always change to an equivalent normalized potential ψ. The equivalence relation here is that of cohomology:

Definition 27.1. One says ψ, φ are **cohomologous** iff there exists u such that

$$\psi = \varphi + u \circ T - u.$$

We call u a **transfer function**; if u belongs to some special class, e.g. it is Hölder, or in L^2, or continuous we shall then say that ψ, φ are **Hölder (L^2, continuously) cohomologous** respectively. In this case, one can find such an u Hölder (with the same exponent), and so the resulting ψ is Hölder as well [Bow75].

In the special case where ψ is cohomologous to the constant function zero, that is there exists a function u such that

$$\psi(x) = u \circ T(x) - u(x), \tag{101}$$

one says that the function ψ is a **coboundary**. Thus ψ and φ are cohomologous iff $\psi - \varphi$ is a coboundary.

For the above example of producing a normalized potential, the transfer function u is produced in an interesting way; it is $u = \log h$ where h is the unique eigenfunction for the Ruelle operator $\mathcal{L}_\varphi$.

27.4. **Cohomology, potential functions and change of basis.** We examine this mysterious *cohomology equation* more closely. First, in the matrix formulation of the Ruelle operator, we see that the cohomology equation for potentials $\log p$ and g,

$$\log p = g + \log h - \log h \circ \sigma$$

is turned into similarity of matrices:

$$\widetilde{P} = HLH^{-1}.$$

Thus the cohomology can be thought of simply as expressing the matrix in terms of a different basis. We point out now that this works in general; see also [AF02].

Proposition 27.2. *Let φ, v be continuous functions: $\Sigma_A^+ \to \mathbb{R}$. Define the Ruelle operator by*

$$(\mathcal{L}_\varphi f)(\underline{x}) = \sum_{w \in \sigma^{-1}(\underline{x})} e^{\varphi(w)} f(w);$$

define $\mathcal{V}$ to be the multiplication operator

$$\mathcal{V} : f(\underline{x}) \to v(\underline{x}) f(\underline{x}).$$

Then, with $\mathcal{L}_\psi$ denoting the Ruelle operator for the potential $\psi = \varphi + \log v - \log v \circ \sigma$, we have

$$\mathcal{L}_\psi = V^{-1} \circ \mathcal{L}_\varphi \circ V.$$

Proof.

$$\mathcal{L}_\psi : f(\underline{x}) \mapsto \sum e^{\varphi(w) + \log v(w) - \log v \circ \sigma(w)} f(w) = \frac{1}{v(\underline{x})} \sum_{\sigma w = \underline{x}} e^{\varphi(w)} \big(f(w) v(w)\big)$$

which proves the claim. □

27.5. **A projective metric proof of the Ruelle-Perron-Frobenius theorem.** For other proofs see [Bow75], [Wal75], [Led74], [PP90]; here we present the projective metric proof of Ferrero and Schmidt [FB79], [FB88], see also [Via97], [Liv96], [BG95].

Theorem 27.3. *Assume A is a primitive $0-1$ matrix. Let $\varphi \in \mathcal{H}_\alpha^b$, with $\mathcal{L}_\varphi$ the Ruelle operator acting on $\mathcal{C}(\Sigma_A^+)$. Then there exists a unique maximum eigenvalue $\lambda > 0$ for $\mathcal{L}_\varphi$; this is $\lambda = e^P$ where $P = P(\varphi)$ is the pressure. There exists a positive eigenfunction h, unique up to multiplication by a constant. For this eigenvalue, there also exists for the dual operator $\mathcal{L}_\varphi^*$ an eigenmeasure ν, unique when normalized so that $\nu(\Sigma_A^+) = 1$ and $\nu(h) = 1$.*

The potential $\psi = \varphi + \log h \circ T - \log h - P$ is normalized: it has pressure is 0; the operator $\mathcal{L}_\psi$ has eigenvalue $1 = e^0 = e^P$ with unique positive eigenfunction 1,

while the dual operator $\mathcal{L}_\psi^*$ has a unique eigenmeasure of eigen value one μ, with μ invariant and $\mu = h \cdot \nu$. For all $f \in \mathcal{C}(\Sigma_A^+)$,

$$||\mathcal{L}_\psi^t(f) - \mu(f)||_\infty \to 0$$

as $t \to \infty$. For all $f = I_P$, where $P \in \mathbb{C}_0^k$ is a k-cylinder set for some k, this convergence is exponentially fast: there exists $c > 0$, $\beta \in (0,1)$ such that for all k, for all t,

$$||\mathcal{L}_\psi^t(f) - \mu(f)||_\infty < c \cdot \beta^t.$$

We need first a series of lemmas. We fix $\alpha \in (0,1)$ and $b \geq 1$. Write

$$B_n = \exp\left(2b \sum_{n+1}^{\infty} \alpha^k\right), \tag{102}$$

and define Λ_1 to be the set of all functions $f : \Sigma_A^+ \to [0, \infty)$ such that for each $n \geq 0$,

$$f(\underline{x}) \leq B_n f(\underline{y}) \text{ if } \underline{x}, \underline{y} \in C \in \mathbb{C}_0^{n-1} \tag{103}$$

That is, for $n \geq 1$ $\underline{x}, \underline{y}$ are in the same thin cylinder $[.x_0 \dots x_{n-1}]$, while for $n = 0$ means $\underline{x}, \underline{y} \in \{\Sigma_A^+\}$.

Next we define for $p \geq 1$

$$\Lambda_p = \{f \geq 0 : \ f(\underline{x}) \leq B_n f(\underline{y}) \text{ if } \underline{x} = \underline{y} \text{ on } [0 \dots (n-1)], \text{ for each } n \geq p\}. \tag{104}$$

Remark 27.1. To motivate these definitions, we note that for the case where $f \in \Lambda_1$ is never zero, equivalently $f = e^F$ where F is $\alpha-$Hölder with constant $2b\alpha/(1-\alpha)$. Similarly, nonzero $f \in \Lambda_p$ means that restricted to any thin cylinder $C \in \mathbb{C}_0^{p-1}$, F is $\alpha-$Hölder with constant $2b\alpha/(1-\alpha)$. This is because

$$\frac{f(\underline{x})}{f(\underline{y})} \leq B_n$$

so

$$\log f(\underline{x}) - \log f(\underline{y}) = F(\underline{x}) - F(\underline{y}) \leq \log B_n = 2b \sum_{n+1}^{\infty} \alpha^k = \frac{2b}{1-\alpha} \alpha^{n+1} = \frac{2b\alpha}{1-\alpha} \alpha^n.$$

Next we set

$$\Lambda = \cup_{p=1}^{\infty} \Lambda_p.$$

Lemma 27.4. *$\Lambda_1 \subseteq \Lambda_2 \subseteq \dots \subseteq \Lambda_n \dots \subseteq \Lambda$, and*
(1) Λ_p and Λ are convex cones which satisfiy Furstenberg's condition.
(2) The difference set $\Lambda - \Lambda$ is dense in the continuous functions $\mathcal{C}$;

Proof. We prove (1). We check, writing K for Λ_p or Λ:
$K + K \subseteq K$ since

$$(f+g)(\underline{x}) \leq B_n f(\underline{x}) + B_n g(\underline{x}) = B_n (f+g)(\underline{x})$$

$aK \subseteq K$ for $a \geq 0$:

$$af(\underline{x}) \leq aBf(\underline{y}) = B(af)(\underline{y})$$

Now we check Furstenberg's condition. Indeed, the cone of nonnegative continuous functions $\mathcal{C}^+ \supseteq \Lambda$ has this property: take $f, g \in \mathcal{C}^+$ such that $f \neq g$. We claim that $f + t(g - f)$ is not in $\mathcal{C}^+$ for all $t \in \mathbb{R}$. But since $f \neq g$, there is an $\underline{x}$ with $f(\underline{x}) = a \neq b = g(\underline{x})$. And for positive $a \neq b$, $a + t(b - a)$ will be negative for some $t \in \mathbb{R}$, confirming the claim.

Next, by Lemma **??**, the defining condition for Λ_p makes no restrictions for the difference of values $F(\underline{x}) - F(\underline{y})$ for $\underline{x}, \underline{y}$ in a cylinder set in $\mathbb{C}_0^n$ for $n < p$. So clearly these sets are nested increasing. To prove (2), in particular, all nonnegative step functions which are constant on cylinders in $\mathbb{C}_0^{p-1}$ are included- with arbitrarily large differences between the steps. Now from the cone property, $\Lambda_p - \Lambda_p$ is a vector space, hence contains all the step functions; these are dense in $\mathcal{C}$, proving (1). □

We now fix $\varphi \in \mathcal{H}_\alpha^b$ as in the statement of the theorem. Given an allowed string $\underline{x} = (.x_0x_1\dots)$, we write $j\underline{x}$ for $(.jx_0x_1\dots)$ where $A_{jx_0} = 1$.

Lemma 27.5. *$\mathcal{L}_\varphi\Lambda_p \subseteq \Lambda_{p-1}$ for all $p > 1$, and $\mathcal{L}_\varphi\Lambda_1 \subseteq \Lambda_1$.*

Proof. We are given $f \in \Lambda_p$, $\varphi \in \mathcal{H}_\alpha^b$ so for all $n \geq 0$, when $\underline{x} = \underline{y}$ on some $C \in \mathbb{C}_0^{n-1}$, recalling here that $\mathbb{C}_0^{-1} = \{\Sigma_A^+\}$, then $|\varphi(\underline{x}) - \varphi(\underline{y})| \leq b\alpha^n$ whence

$$\exp(\varphi(\underline{x})) \leq \exp(b\alpha^n)\exp(\varphi(\underline{y})) \qquad \text{for all } n \geq 0, \text{ and} \tag{105}$$

$$f(\underline{x}) \leq B_n f(\underline{y}) \qquad \text{for all } n \geq p. \tag{106}$$

We shall show that for all $n \geq p - 1$, for all $\underline{x}, \underline{y}$ with $x_i = y_i$ for $0 \leq i < n - 1$,

$$(\mathcal{L}_\varphi f)(\underline{x}) < B_n \mathcal{L}_\varphi f(\underline{y}).$$

When $\underline{x} = \underline{y}$ on $[0 \dots n-1]$, then if $A_{jx_0} = 1$, $j\underline{x} = j\underline{y}$ on $[0 \dots n]$, so $\exp(\varphi(j\underline{x})) \leq \exp(b\alpha^{n+1})\exp(\varphi(j\underline{y}))$ for all $n \geq 0$ while $f(j\underline{x}) \leq B_{n+1}f(j\underline{y})$ for all $n + 1 \geq p$. We note that

$$\exp(b\alpha^{n+1})B_{n+1} = \exp(-b\alpha^{n+1})\exp(2b\alpha^{n+1})B_{n+1} = \exp(-b\alpha^{n+1})B_n < B_n.$$

Therefore for all $n \geq p - 1$,

$$(\mathcal{L}_\varphi f)(\underline{x}) \equiv \sum_{j:\, A_{j\underline{x}_0}=1} e^{\varphi(j\underline{x})} f(j\underline{x}) \leq \exp(b\alpha^{n+1})B_{n+1} \sum_{j:\, A_{jx_0}=1} e^{\varphi(j\underline{y})} f(j\underline{y}) \tag{107}$$

$$= \exp(-b\alpha^{n+1})B_n\mathcal{L}_\varphi f(\underline{y}) < B_n\mathcal{L}_\varphi f(\underline{y}). \tag{108}$$

Thus in particular, $\mathcal{L}_\varphi\Lambda_2 \subseteq \Lambda_1$, but since $\Lambda_1 \subseteq \Lambda_2$, also $\mathcal{L}_\varphi\Lambda_1 \subseteq \Lambda_1$. □

Lemma 27.6. *Writing $d_p(\cdot,\cdot)$ for the projective metric on Λ_p, then for $f, g \in \Lambda_p$, $d_p(f, g) = \log(\beta/\alpha)$ where*

$$\beta = \sup_{n \geq p} \sup_{\substack{x_i = y_i \\ 0 \leq i < n}} \frac{B_n f(\underline{y}) - f(\underline{x})}{B_n g(\underline{y}) - g(\underline{x})}$$

and

$$\frac{1}{\alpha} = \sup_{n\geq p} \sup_{\substack{x_i=y_i \\ 0\leq i<n}} \frac{B_n g(\underline{y}) - g(\underline{x})}{B_n f(\underline{y}) - f(\underline{x})}.$$

Proof. $d_p(f,g) = \log \beta/\alpha$ where α, β are the largest, respectively smallest numbers such that $\alpha g \leq f \leq \beta g$, i.e. such that $f - \alpha g, \beta g - f \in \Lambda_p$. Thus

$$(f - \alpha g)(\underline{x}) \leq B_n(f - \alpha g)(\underline{y})$$

whence

$$\alpha(B_n g(\underline{y}) - g(\underline{x})) \leq (B_n f(\underline{y}) - f(\underline{x}))$$

so

$$\alpha = \inf_{n\geq p} \inf_{\substack{x_i=y_i \\ 0\leq i<n}} \frac{B_n f(\underline{y}) - f(\underline{x})}{B_n g(\underline{y}) - g(\underline{x})}.$$

Similarly

$$\beta = \sup_{n\geq p} \sup_{\substack{x_i=y_i \\ 0\leq i<n}} \frac{B_n f(\underline{y}) - f(\underline{x})}{B_n g(\underline{y}) - g(\underline{x})}.$$

Lemma 27.7. *The projective diameter Δ of $\mathcal{L}_\varphi \Lambda_1$ in Λ_1 is finite.*

Proof. □

Write $d = d_1$ and $\mathcal{L} = \mathcal{L}_\varphi$. By the triangle inequality, $d(f,g) \leq d(f,1) + d(1,g)$ where 1 is the constant function.

From the Lemma, for $f \in \Lambda_1$, $d(\mathcal{L}f, 1) \leq \log(\beta/\alpha)$ where

$$\beta = \sup_{n\geq 1} \sup_{\substack{x_i=y_i \\ 0\leq i<n}} \frac{B_n \mathcal{L}f(\underline{y}) - \mathcal{L}f(\underline{x})}{B_n - 1}$$

and

$$\frac{1}{\alpha} = \sup_{n\geq 1} \sup_{\substack{x_i=y_i \\ 0\leq i<n}} \frac{B_n - 1}{B_n \mathcal{L}f(\underline{y}) - \mathcal{L}f(\underline{x})}.$$

Since $(\mathcal{L}_\varphi f)(\underline{y}) < B_n \mathcal{L}_\varphi f(\underline{x})$, we have for all $\underline{x}, \underline{y}$ with $x_i = y_i$ on $0 \leq i \leq n-1$,

$$\frac{B_n \mathcal{L}f(\underline{y}) - \mathcal{L}f(\underline{x})}{B_n - 1} \leq \frac{(B_n^2 - 1)\mathcal{L}f(\underline{x})}{B_n - 1} \leq (B_n + 1) \sup |\mathcal{L}f|$$

and since by (108) $0 \leq \mathcal{L}_\varphi f(\underline{x}) \leq \exp(-b\alpha^{n+1}) B_n \mathcal{L}_\varphi f(\underline{y})$, for each $n \geq 0$,

$$\frac{B_n - 1}{B_n \mathcal{L}f(\underline{y}) - \mathcal{L}f(\underline{x})} \leq \frac{B_n - 1}{B_n(1 - \exp(-b\alpha^{n+1}))\mathcal{L}f(\underline{y})} \leq \frac{B_n - 1}{B_n(1 - \exp(-b\alpha^{n+1}))} \frac{1}{\inf |\mathcal{L}f|}$$

Thus

$$\log(\beta/\alpha) \leq \sup_{n\geq 0}(B_n + 1) \sup_{n\geq 0} \frac{(B_n - 1)}{B_n(1 - \exp(-b\alpha^{n+1}))} \frac{\sup |\mathcal{L}f|}{\inf |\mathcal{L}f|}$$

Since $\leq B_0 = \exp(2b\alpha/(1-\alpha))$, and B_n decreases to 1, setting $K = B_0$ this is $\leq$??? □

28. Nonlinearity: Smooth structures

In the previous section, we have seen how expanding maps of the circle are classified up to topological congugacy (by Shub's theorem).

Here we begin to study smooth conjugacies, with degrees $\mathcal{C}^1$ or higher of differentiability. This will further divide each topological equivalence class into uncountably many smooth classes.

The first tool we shall need is **bounded distortion**. This has many variants, in different parts of dynamics. Our present setting of hyperbolic circle maps provides a good place to begin an understanding of such phenomena.

This will simultaneously help us study smooth conjugacy, and give us a key to understanding the *small scale structure* of our maps.

28.1. Bounded distortion property.

First we need some definitions.

Definition 28.1. Given two metric spaces (X, d) and $(\widehat{X}, \widehat{d})$, a function $f : X \to \widehat{X}$ is **Hölder** with **exponent** $\alpha > 0$ and **constant** $c > 0$ iff for each $x, y \in X$ then

$$\widehat{d}(f(x), f(y)) \leq c(d(x, y))^{\alpha}.$$

We denote by $\mathcal{H}^{\alpha}(X, \widehat{X})$ the collection of all α- Hölder functions.

Note that $\mathcal{H}^1 = \text{Lip}$, the Lipschitz functions. A first interesting fact is:

Lemma 28.1. *For $X = I = [0, 1]$ with the Euclidean metric, then if $\alpha > 1$, $\mathcal{H}^{\alpha} \equiv \mathcal{H}^{\alpha}(I, I)$ is the set of all constant functions. The same is true if we repace I by a path-connected differentiable manifold with a metric which changes smoothly along differentiable curves.*

Proof. If $f : I \to I$ is Hölder for $\alpha > 1$, then

$$|f(x) - f(x_0)| \leq c|x - x_0|^{\alpha},$$

so

$$\left| \frac{f(x) - f(x_0)}{x - x_0} \right| < \frac{|x - {x_0}^{\alpha}|}{|x - x_0|} \to 0$$

as $x \to x_0$, so by the Mean Value Theorem f is constant. (The same proof works along paths.) □

Remark 28.1. For $X = C$ the Cantor set, which is totally disconnected, $\mathcal{H}^{\alpha}$ is much larger. The derivative (one- or two-sided) still exists and is zero at every point, by the above argument, but the Mean Value Theorem can no longer be applied, and indeed, f need not be constant.

??? the same since 2006:

two contraction mappings $\varphi_0, \varphi_1 : I \longrightarrow I$. We consider first the case where these maps are orientation-preserving, and are *strict* contractions in the sense that the derivatives satisfy $0 < \alpha < D\varphi_i < \beta < 1$. We also require that

$$0 = \varphi_0(0) < \varphi_0(1) < \varphi_1(0) < \varphi_1(1) = 1.$$

This implies that the intervals $I_0 \equiv \varphi_0(I), I_1 \equiv \varphi_1(I)$ are disjoint. We assume that φ_0, φ_1 are $\mathcal{C}^{1+\gamma}$ maps for some $\gamma \in (0, 1]$.

Remark 28.2. Here $\mathcal{C}^{k+\gamma}$ means the k^{th} derivative $D^k\varphi_i$ is Hölder continuous with exponent γ; note that $\mathcal{C}^{1+1}$ means $D\varphi_i$ is Lipschitz, so $\mathcal{C}^2$ implies $\mathcal{C}^{1+1}$ (by compactness) but not conversely. (Exponent $\gamma > 1$ is excluded because in that case, since the domain I is connected, φ_i is identically constant hence immediately of order $\mathcal{C}^\infty$ - while the whole purpose of Hölder conditions is to have *intermediate* grades of smoothness).

We define $f : I_0 \cup I_1 \to I$ to be the map with inverse branches φ_0, φ_1. Note that since $D\varphi_i$ are bounded away from 0 and ∞, it follows that f is $\mathcal{C}^{1+\gamma}$ with same Hölder exponent, but with different Hölder constant.

Inductively, form

$$I_{x_0 \dots x_n} = \varphi_{x_0}(\varphi_{x_1} \dots (\varphi_{x_n}(I)))$$

where $x_k \in \{0, 1\}$; $\bigcup I_{x_0 \dots x_n}$ (union over all choices, with n fixed) is the n^{th} level approximation to the Cantor set, C, defined as

$$C = \bigcap_{n=0}^{\infty} \bigcup I_{x_0 \dots x_n}.$$

The restriction of the map f to C maps C to itself and is (just as for the middle-third set) conjugate to the Bernoulli shift (Σ^+, σ), via the map $\pi : (x_0 x_1 \dots) \mapsto x$, where x is the unique element of $\cap_{n=0}^{\infty} I_{x_0 \dots x_n}$.

A set C together with map $f : I_0 \cup I_1 \to I$, defined in this way from strict contractions φ_0, φ_1, will be called a **strictly hyperbolic $\mathcal{C}^{1+\gamma}$ Cantor set (with map)**.

We learned this version of bounded distortion from M. Urbanski. See also e.g. [SS85]; pp. 169-170 of [Mañ87] has some interesting historical remarks; Bowen and Sullivan are among the many authors who have used related tools masterfully; for some sophisticated variants see e.g. [KH95] and [dMvS93], §7 of [MU03].

Lemma 28.2. *With f as above, $\exists K > 0$ such that for all n, for any $\delta > 0$, if J is an interval such that $f^m|_J$ is 1-1 and the image $f^m(J)$ has diameter less that δ, then for all $x, y \in J$,*

$$e^{-K\delta^\gamma} < |\frac{Df^m x}{Df^m y}| < e^{K\delta^\gamma}.$$

We mention that one sees from the proof that if c is the Hölder constant for $\log |Df|$, then the constant K is given by $K = c\beta^\gamma/(1 - \beta^\gamma)$.

28.2. **Scaling functions and g-measures.**

Definition 28.2. R is called the **scaling function** of C.

Proof. We will first show that for each y, $R_n(y)$ $n = 1, 2, \dots$ is a Cauchy sequence. Since

$$S^m(I_{y_{-(n+m)} \dots y_{-1}}) = I_{y_{-n} \dots y_{-1}}$$

and similarly for the subintervals, applying the Mean Value Theorem and Bounded Distortion Property (Corollary 2.2) we have for all $m \geq 0$

$$R_n(y) = R_{n+m}(y) e^{\pm K\beta^{n\gamma}}$$

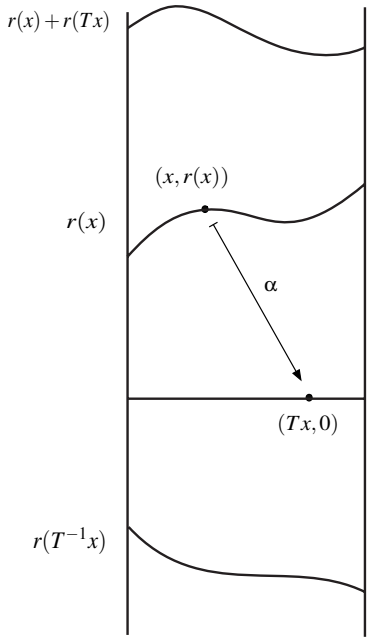


FIGURE 71. Fundamental domains for the special flow

Therefore $R_n(y)$ is Cauchy sequence (i.e. each of its three coordinates is) hence it converges; call the limit $R(y)$. Next, if $y, w \in \underset{-\infty}{\overset{-1}{\to}} \prod\{0,1\}$ agree on the coordinates $-n, \ldots, -1$ then since $R_n(y) = R(y)e^{\pm K\beta^{n\gamma}}$ and $R_n(y) = R_n(w)$, we have

$$R(y) = R(w)e^{\pm 2K\beta^{n\gamma}}.$$

Writing $\|\cdot\|$ for sup norm in $\mathbb{R}^3$, this implies that, with the log taken by components,

$$\| \log R(y) - \log R(w) \| \leq 2K(d_\beta(y,w))^\gamma,$$

i.e. $\log R$ is Hölder continuous with exponent γ; therefore so is R. □ □

29. Examples of cohomology in dynamics

We have encountered above the rather mysterious cohomology equation; here we shall give several examples which help explain its meaning.

Next:

–change of special flow cross-section and flow isomorphism

–circle skew product and change of origin

–smooth conjugacy of doubling maps and chain rule

–Sinai's Lemma

29.1. **Special flows: nonpositive "return times" and change of cross-section. (Invertible case)** Let (X, T, μ) be an invertible measure-preserving transformation, and $r : X \to \mathbb{R}$ a measurable function. In §10.6, for the case where r is positive, we defined the *special flow with return time* r and *cross-section map* T in the standard

way; here we give a more general definition using equivalence relations, which makes sense of "return times" which are not necessarily positive.

We begin with the *vertical flow* on $X \times \mathbb{R}$, by simply moving upward at unit speed, $\tau_t(x,s) = (x, s+t)$. We define an equivalence relation on $X \times \mathbb{R}$, whose equivalence classes are the orbits of a map $\alpha : X \times \mathbb{R} \to X \times \mathbb{R}$, with

$$\alpha(x,s) = (Tx, s - r(x))$$

We write $\sim_r$ for this equivalence relation. We write $\Omega_r = X \times \mathbb{R}/\sim_r$.

Note that $(x, r(x)) \sim_r (Tx, 0)$ (as in the usual definition of special flow) and also for example $(x,0) \sim_r (T^{-1}x, r(T^{-1}x))$.

Proposition 29.1. *The vertical flow τ_t on $X \times \mathbb{R}$ induces a flow on Ω_r.*

Proof. We need only check that $\alpha(\tau_t(x,s)) = \tau_t(\alpha(x,s))$, which is immediate. □

Example 25. If $r(x) > 0$ for all x, then this is isomorphic to the usual special flow with return-time function r. In this case,

$$\{(x,s) : 0 \le s \le r(x)\}$$

is a **fundamental domain** for the $\mathbb{Z}$- action: there is exactly one point from each equivalence class, except for the upper and lower boundary, where we have the identification of $(x, r(x))$ with $(Tx, 0)$. Other copies of the fundamental domain are indicated in Figure 72. (This shows a semiflow, i.e. not a flow but an action of the semigroup $\mathbb{R}^+$, since the base map is the doubling map on the circle $T(x) = 2x(\mathrm{mod}1)$ which is not invertible). For this particular example, $r(x) = \sin(4\pi x) + 2$ so the curves are all periodic and are trigonometric polynomials, $r(x) + r(2x) + r(4x)$ and so on.

Definition 29.1. We say r and $\widehat{r}$ are **cohomologous** iff there exists a real-valued function u such that

$$\widehat{r}(x) = r(x) - u(x) + u(Tx). \tag{109}$$

This defines another flow $\widehat{\tau}$ on $\widehat{\Omega} = X \times \mathbb{R}/\widehat{\sim}$.

Proposition 29.2. *The map $\Phi : \Omega \to \widehat{\Omega}$ defined by*

$$\Phi : (x,s) \mapsto (x, s - u(x))$$

is a flow isomorphism.

Proof. It is enough to show that the following diagram commutes, as is easily checked:

$$\begin{array}{ccc} X \times \mathbb{R} & \xrightarrow{\Phi} & X \times \mathbb{R} \\ \downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle \widehat{\alpha}} \\ X \times \mathbb{R} & \xrightarrow{\Phi} & X \times \mathbb{R} \end{array}$$

□

Example 26. Again in the special case of $0 < r(x)$, suppose now that $0 \le u(x) \le r(x)$. Then $\widehat{r}$ is just the return time to the new cross-section $\{(x, u(x))\}$ of the flow, see Fig. **??**

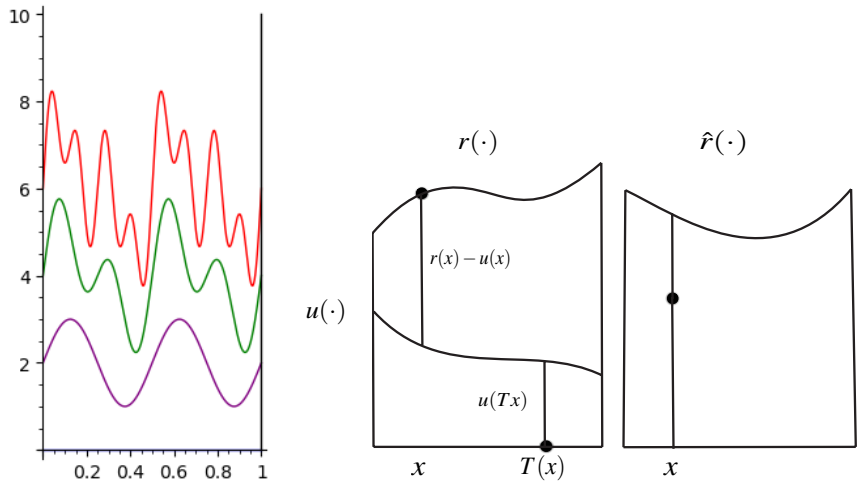


FIGURE 72. On left, fundamental domains for the map α, partitioning $X \times \mathbb{R}^+$, giving a special semiflow over the doubling map of the circle. For the flow case, this extends to a partition of $X \times \mathbb{R}$. On right, return to a new cross-section, the graph of $u(x)$.

In the special case where r is the constant function $r = 1 > 0$, thus there exists a function u such that

$$r(x) = 1 + u \circ T(x) - u(x), \tag{110}$$

then the special flow with return time 1 is called the *suspension flow* over the map T. If r is cohomologous to the constant 1, then the special flow with return time r is isomorphic to this suspension flow.

Special semiflows: the noninvertible case)

Definition 29.2. By a *semiflow* on a space X we mean an action of $(\mathbb{R}^+, +)$ on X. This is often called a *semigroup* as we sometimes do in these notes but is more properly called a *monoid* as it has an identity element. (A semigroup is a set S with an associative binary operation on it; a group is a moonid with inverses). See [Aki13]. For a development of the ergodic theory of semiflows see [LR04].

We let (X, T, μ) be a not-necessarily invertible measure-preserving transformation, and $r : X \to \mathbb{R}$ a measurable function. We make all the same definitions as for the invertible case, except now we have the *vertical semiflow* on $X \times \mathbb{R}^+$, by simply moving upward at unit speed, $\tau_t(x, s) = (x, s + t)$, now for $t \geq 0$. See Fig. 72.

29.2. **Smooth changes of coordinates.**

Example 27. Let X be a one-dimensional space, S^1 or $\mathbb{R}$ or I, and let $T, S, h : X \to X$ be $\mathbb{C}^{1+\alpha}$ maps for some $\alpha \in (0, 1]$ with derivative > 0 everywhere. We assume that h is a diffeomorphism conjugating the two smooth maps T and S, so $h^{-1} \circ S \circ h = T$. Then the two functions $\varphi(x) = \log DT(x)$ and $\psi(x) = \log DS(h(x))$ are Hölder cohomologous, by the fuction $u(x) = -\log Dh(x)$.

Proof. By the Chain Rule, $DT(x) = Dh^{-1}(S(h(x))) \cdot DS(h(x))S \cdot Dh(x)$. Now $Dh^{-1}(S(h(x))) = 1/Dh(h^{-1}(S(h(x)))) = 1/Dh(T(x))$. Taking logs, $-(\log Dh)(Tx) + (\log DS) \circ h)(x) + (\log Dh)(x) = (\log DT)(x)$, so for φ, ψ and u as defined above, we have $\varphi(x) = \psi(x) - u(Tx) + u(x)$ as claimed. □

29.3. **Smooth change of metric.** Let (X, T) be as in the last example, for $X = S^1, \mathbb{R}$ or I, equipped with order and origin 0 and with the standard Euclidean metric $d(\cdot, \cdot)$. Now let $\rho(\cdot, \cdot)$ be a second metric which is differentiably equivalent, in the sense that the function $f(x) = \rho(0, x)$ for $x > 0$, $f(x) = -\rho(x, 0)$ for $x < 0$ is differentiable and > 0. Then $\log DT|_d$ and $\log DT|_\rho$ are cohomologous

Proof. In fact this is equivalent to the previous example: define a map from X to X by $x \mapsto h(x) \equiv f(x)/b$ where b is the total ρ-length in the case of I or S^1 and is just 1 for $X = \mathbb{R}$. Then define $S : X \to X$ by $S = h \circ T \circ H^{-1}$. We are now in the situation of the previous example. □

Remark 29.1. In fact in much more generality, e.g. for a smooth map f of a manifold M "changing the (Riemannian) metric" is equivalent to conjugation of the map. And in the special case of one-dimensional manifolds, where the derivative is real-valued, cohomology again corresponds to a change in the reference coordinates of some type, in this case, the metric.

To extend the cohomology ideas to the higher dimensional case, we would need to move beyond real or circle-valued cocycles to those with values in e.g. matrix groups. See [Liv71], [Liv72], [PP97] for some beginnings in the vast related literature.

29.4. **Time-shifts and time averages.** We begin with a small observation: if we are given φ and $\widetilde{\varphi}$ and are looking for a function u such that

$$\widetilde{\varphi}(x) = \varphi(x) + u \circ T(x) - u(x)$$

then if we choose a value for $u(x)$ for some x, this choice determines u on the rest of the orbit of x, from the cohomology equation, for:

$$u \circ T(x) = \widetilde{\varphi}(x) - \varphi(x) + u(x).$$

Now of course a nonmeasurable solution u always exists (just choose one point in each orbit by the Axiom of Choice, define, say, $u = 0$ there and use the above remark!) but if u is measurable, this observation indicates that there won't be so much freedom in defining u. We see this in a much stronger form in Prop. 30.3.

??Comments: leads to positive function

Proposition 29.3. *Let (X, T) be a dynamical system and let $\varphi : X \to \mathbb{R}$. Then the functions $\varphi \circ T^n$ for $n \in \mathbb{Z}$ and $\frac{1}{n} S_n(\varphi) = (\varphi + \varphi \circ T + \ldots \varphi \circ T^n)/n$ for n fixed are cohomologous to φ.*

Proof. For $\psi = \varphi \circ T$ we take $u = \varphi$, and we have:

$$\varphi \circ T = \varphi + \varphi \circ T - \varphi.$$

Note: we can see this in the special flow example, Fig. 71; if we take as our new cross-section $(x, r(x)$ then the new return-time is indeed $r \circ T$.

Using that same picture, we can already guess the function u for $\varphi \circ T^2$: it is $u = \varphi + \varphi \circ T$, for then,

$$\varphi \circ T^2 = \varphi + (\varphi + \varphi \circ T) \circ T - (\varphi + \varphi \circ T) = \varphi + \varphi \circ T + \varphi \circ T^2 - \varphi - \varphi \circ T.$$

Next we note that if φ_1 is cohomologous to φ by u_1 and φ_2 is cohomologous to φ by u_2, then $(\varphi_1 + \varphi_2)/2$ is cohomologous to φ by $(u_1 + u_2)/2$. This gives the proof for $S_n\varphi/n$. And in fact, the function u will be

$$u = \sum_{k=0} np^n(S_k\varphi).$$

□

??check above. Note: any two fns are nonmeas cohom!

29.5. **Skew products.** Let (X, T) be a (possibly noninvertible) transformation (this could be in the differentiable, topological, or measure category) and let Y be a second measure space, with $\mathcal{T}(Y)$ some collection of maps of Y. Now given some function $\varphi : X \to \mathcal{T}$, we then define a transformation $\mathbb{T}_\varphi : X \times Y \to X \times Y$ by

$$T_\varphi(x, y) = (Tx, \varphi_x(y).$$

Here we have written $\varphi_x \equiv \varphi(x)$. We call T_φ the **skew product transformation** over the **base** X with **skewing function** φ.

Example 28. For a first example, let $f : M \to M$ be a smooth map (not necessarily invertible) on a d- dimensional differentiable manifold M, with tangent bundle TM. We assume that there is a single chart $\Phi : M \to \mathcal{U} \subseteq \mathbb{R}^d$. This gives us a way of uniquely representing the derivative $Df(x)$ as a $d \times d$ matrix $D_f(x)$ Now the derivative map

$$Df(x, \mathbf{v}) \mapsto (f(x), D_{f(x)}\mathbf{v})$$

is a skew product transformation on $M \times \mathbb{R}^d$.

Example 29. A second example is the famous T/T^{-1} transformation studied by Kalikow.

Here the base is the Bernoulli shift space $\Sigma = \Pi_{-\infty}^{+\infty}\{0, 1\}$ with left shift σ and with independent product measure (p, q) for $p, q > 0$ and $p + q = 1$, and

$$\varphi(\underline{x}) = T \text{ if } x_0 = 0, T^{-1} \text{ if } x_0 = 1.$$

We write σ_φ for the skew product.

Now defining $S_0 = 0$, $S_n = \sum_{i=0}^{n-1}(2x_i - 1)$ for $n \in \mathbb{Z}$ we have a random walk on the integers; and $\sigma_\varphi^n(x, y) = (\sigma^n x, T^{S_n(x)}y)$, so one can think of this for fixed $\underline{x}$ as a random sequence of transformations.

Remark 29.2. Indeed, in some sense *every* skew product is a *random dynamical system*; see [Wal89] for a nice introduction.

Kalikow in a famous paper [Kal82], see Example 29, answered a conjecture of Ornstein, the second part of this theorem; the first part being not hard and the second a major achievement:

Theorem 29.4.

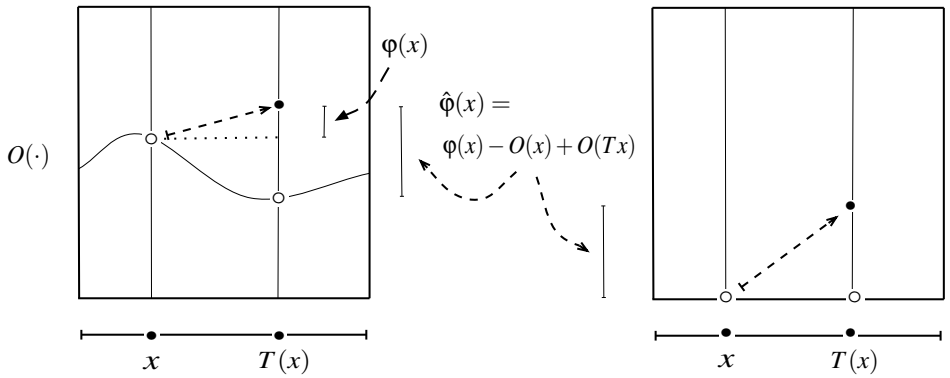


FIGURE 73. Change of origin in circle fibers (warning: diagram needs to be changed: signs are switched, as $\widehat{\varphi}(x) = \varphi(x) + \mathcal{O}(x) - \mathcal{O}(Tx)$!)

(i)If $p \neq q$, so the random walk drifts to ∞, then σ_φ is Bernoulli, i.e. is measure-theoretically isomorphic to a Bernoulli shift;
(ii) If $p = q$ then this map is K (Kolmogorov, i.e. every factor has positive entropy) but not Bernoulli.

29.6. **Circle-valued skew products; change of origin in circle fibers.** Another important class of examples are **circle-valued skew products**, where Y is the circle, $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, and the maps $\varphi(x)$ are circle rotations $R_\theta(x) = x + \theta$ on $\mathbb{T}$, also written as $R_\theta(x) = x + \theta(\mathrm{mod}1)$

For a circle-valued skew product given by a (perhaps noninvertible) map (X, T) and skewing function $\varphi : X \to \mathbb{T}$, so

$$T_\varphi(x, s) = (Tx, s + \varphi(x)), \tag{111}$$

we consider the effect of changing coordinates on $X \times \mathbb{T}$ by choosing a new origin for each fiber.

The new origin is given by a function $\mathcal{O} : X \to \mathbb{T}$, so the new coordinates of a point (x, s) are $(x, s - \mathcal{O}(x))$, see Fig. 73. The point above x of height $s = \mathcal{O}(x)$ is mapped by T_φ to the point above Tx of height $\mathcal{O}(x) + \varphi(x)$. This is distance $\widehat{\varphi}(x)$ above $\mathcal{O}(Tx)$, for the new skewing function $\widehat{\varphi}$. That is,

$$\mathcal{O}(x) + \varphi(x) = \mathcal{O}(Tx) + \widehat{\varphi}(x)$$

so

$$\widehat{\varphi}(x) = \varphi(x) + \mathcal{O}(x) - \mathcal{O}(Tx).$$

If we replace $\mathcal{O}$ by $u = -\mathcal{O}$, this gives the cohomology equation, with *transfer function* u, and we have:

Proposition 29.5. *Given a function u, defining*

$$\widehat{\varphi}(x) = \varphi(x) + u(Tx) - u(x),$$

then the map

$$\Phi : (x, s) \mapsto (x, s + u(x)) = (x, s - \mathcal{O}(x))$$

defines an isomorphism from T_φ to $T_{\widehat{\varphi}}$, corresponding to choice of a new origin $\mathcal{O}(x) = -u(x)$ on each fiber.

Proof. We have: $\Phi \circ T_\varphi \circ \Phi^{-1}(x, r) = \Phi(T_\varphi(x, r - u(x))) = \Phi(Tx, r - u(x) + \varphi(x)) = (Tx, r - u(x) + \varphi(x) + u(Tx)) = (Tx, r + \widehat{\varphi}(x))$, so the diagram commutes, verifying the claim.

$$\begin{array}{ccc} X \times \mathbb{T} & \xrightarrow{T_\varphi} & X \times \mathbb{T} \\ \downarrow \Phi & & \downarrow \Phi \\ X \times \mathbb{T} & \xrightarrow{T_{\widehat{\varphi}}} & X \times \mathbb{T} \end{array}$$

□

Corollary 29.6. *Given two circle-valued skew products, then if the skewing functions are cohomologous (mod* 1*), the maps are isomorphic, via an isomorphism preserving the base maps.*

Modifying to this context Definition 27.1, one says that φ is **cohomologous to zero (mod** 1**)**, or that the function φ is a **coboundary (mod** 1**)**, if there exists a function u such that

$$\varphi(x) = u \circ T(x) - u(x) \ (\mathrm{mod}\, 1). \tag{112}$$

Thus again, two functions $\varphi, \widehat{\varphi}$ are cohomologous iff they differ by a coboundary (mod 1). If φ itself is a coboundary, we have these geometric and dynamical consequences:

Proposition 29.7. *These are equivalent for the skew product $T_\varphi(x, s) = (Tx, s + \varphi(x))$:*
*(i) φ is a coboundary (mod*1*), that is, there exists $u : X \to \mathbb{T}$ such that*

$$\varphi = u \circ T - u \, (mod\, 1);$$

(ii) there exists a function $\mathcal{O} : X \to \mathbb{T}$ such that the graph of $\mathcal{O}$ is an invariant subset for T_φ;
(iii) there exists a fiber-preserving isomorphism from T_φ to $T \times Id$.

Proof. As we have seen in Prop. 29.5, $\mathcal{O} = -u$ represents a new choice of origin, and φ is cohomologous to 0 iff for the new skewing function $\widehat{\varphi} = 0$ we have (mod 1)

$$\widehat{\varphi}(x) = 0 = \varphi(x) + u(Tx) - u(x), \tag{113}$$

that is,

$$0 = \varphi(x) - \mathcal{O}(Tx) + \mathcal{O}(x).$$

The graph of $\mathcal{O}$ is invariant iff

$$T_\varphi(x, \mathcal{O}(x)) = (Tx, \mathcal{O}(x) + \varphi(x)) = (Tx, \mathcal{O}(Tx))$$

iff $\mathcal{O}(x) + \varphi(x) = \mathcal{O}(Tx)$ equivalently (113). The cohomology corresponds to isomorphism to a new skew product $T_{\widehat{\varphi}}$ by Corollary 29.6, and this is $T \times Id$ iff $\widehat{\varphi} = 0$. Given a fiber-preserving isomorphism Φ, we define $\mathcal{O}$ by defining its graph to be the inverse image by Φ of $X \times \{0\}$. That is, $\mathcal{O}$ is such that $\Phi^{-1}(x, 0) = (x, \mathcal{O}(x))$. □

29.7. **Nonergodicity and coboundaries modulo 1.** All the previous discussion is valid for the topological, measure or smooth categories. Now we move to the context of measurable dynamics.

Remark 29.3. Furstenberg in [Fur61] investigated the question of which skewing functions give uniquely ergodic skew product transformations, assuming the base map is uniquely ergodic. His motivation was to generalize *Weyl's Theorem*, the simplest form of which states that (see Exercise 4.1) a circle rotation R_θ is minimal iff it is uniquely ergodic, iff θ is an irrational number. The unique ergodicity can be proved in a variety of ways; see [Pet89] pp.156-158 regarding this and related results. The circle rotation is an example of a *Kroneker system*, [Fur81]. Weyl proved further that for any real polynomial $p(x)$ with at least one of the nonconstant coefficients irrational, then the sequence $p(n) : n \geq 0$ is uniformly distributed (mod1). On p. 68 of [Fur81] Furstenberg gives a beautiful dynamical proof of this fact. (It is at first surprising that this has anything to do with dynamics: what is the map for degree of p greater than one?)

We next examine when circle-valued skew products are *relatively* uniquely ergodic, defined as follows:

Definition 29.3. We let $(X, \mathcal{A}, \mu)$ be a measure space with T a (not necessarily invertible) ergodic m.p.t.. We let λ denote Lebesgue measure on the circle $\mathbb{T}$ and $\varphi : X \to \mathbb{T}$ a measurable function. Defining $\widehat{X} = X \times \mathbb{T}$ and $\widehat{\mu} = \mu \times \lambda$, T_φ defines a measure-preserving map of $(\widehat{X}, \widehat{\mu})$.

Let $(X, \mathcal{A}, \mu)$ be a measure space with $T : X \to X$ measurable and μ an ergodic invariant probability measure. Assume we have $\varphi : X \to \mathbb{T}$ measurable, with $T_\varphi : X \times \mathbb{T} \to X \times \mathbb{T}$ denoting the skew product. This is a measurable map for the product σ-algebra. We write $\pi : X \times \mathbb{T} \to X$ with $\pi(x, t) = x$ for the projection and define the *fiber above* x to be $\pi^{-1}(x)$. We let $\mathcal{M}_\mu$ denote the collection of all T_φ-invariant probability measures with marginal μ, i.e. which project to μ, and make the following definition. We say T_φ is **uniquely ergodic relative to** μ (or simply μ-uniquely ergodic) iff $\mathcal{M}_\mu$ is the singleton $\widehat{\mu} = \mu \times \lambda$.

Now if φ is cohomologous to zero (mod 1), then as we have seen in Proposition 29.7 the skewing function φ is a coboundary if and only if the graph of $\mathcal{O}$ is a T_φ-invariant subset of $X \times \mathbb{T}$. This is valid in general, but now we are in the measurable context, and this implies the map is not ergodic: to find an invariant measure which is not $\widehat{\mu}$, just lift μ to a measure supported on that graph. Or, more generally, add parallel bands of mass along the graph of $\mathcal{O}$.

An even more transparent view of this nonergodicity is that, also from Proposition 29.7, the skew product is then isomorphic to $T \times \mathrm{Id}$ and so is certainly not ergodic.

This is the simplest way the skew product can be nonergodic: if it can be "straightened out" by a change of origin to $T \times \mathrm{Id}$. More generally, it could be nonergodic if it straightens out to $T \times R_\theta$ for θ rational, i.e. if there is an integer k so that $k\theta = 0 (\mathrm{mod}\, 1)$. For this to happen, the equation (112) is replaced by:

$\varphi(x) = j/k + u \circ T(x) - u(x)$ for some $1 \le j \le k-1$, or equivalently:

$$k\varphi(x) = u \circ T(x) - u(x) \pmod 1 \tag{114}$$

for some nonzero $k \in \mathbb{Z}$. In this case the straightened map, instead of fixing the circle, permutes k equal intervals. In fact conversely, as we will now see, this is all that can happen.

The next theorem represents a generalization of Lemma 2.1 from [Fur70], regarding unique ergodicity. Of course if the base transformation itself is uniquely ergodic, then relative unique ergodicity implies unique ergodicity of the skew product.

Proposition 29.8. *Let $(X, \mathcal{A}, \mu)$ be a probability space with T a measurable, measure-preserving map. Let $\varphi : X \to \mathbb{T} = \mathbb{R}/\mathbb{Z}$ be measurable. The following are equivalent for the circle-valued skew product T_φ:*
(a) $\mu \times \lambda$ is not ergodic;
(b) T_φ is not μ-uniquely ergodic;
(c) there exist $k \in \mathbb{Z}$ such that $k\varphi$ is a coboundary (mod 1), i.e. there exists $u : \Omega \to S^1$ measurable such that

$$k\varphi = u \circ T - u (mod\, 1). \tag{115}$$

Proof. Two proofs of (c) $\Longrightarrow$ (a) have been described above; for the first, by $(i) \Longrightarrow (iii)$ in Prop. 29.7, the graph of $\mathcal{O} \equiv -u$ is invariant; we lift the measure μ to the graph, giving a different invariant measure which hence projects to μ on the base. For the second, as in $(i) \Longrightarrow (iii)$ of Prop. 29.7 plus (114), T_φ is fiber-preserving isomorphic to $T \times R_{j/k}$ for some $0 \le j \le k-1$.

(b) $\Longrightarrow$ (a): We will show that if $\widehat{\mu} \equiv \mu \times \lambda$ is ergodic, then $\mathcal{M}_\mu = \{\widehat{\mu}\}$. We learned this argument from Eli Glasner; another nice argument can be given using generic points, following [Fur61].

Write $\mathcal{M}_\mu$ for the collection of T_φ-invariant measures on the σ-algebra $\widehat{\mathcal{A}} = \mathcal{A} \times \mathcal{B}$ where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{T}$, with marginal μ. The $\widehat{T}$-invariant measures form a convex set, with the ergodic measures as the extreme points, see (ii) of Prop.7.1. Let $\tilde{\mu} \in \mathcal{M}_\mu$. Rotating this measure by the same angle simultaneously in each fiber produces another invariant measure, $R_\theta \tilde{\mu}$. But if we average these rotated measures along each circle fiber, integrating by Lebesgue measure λ on $\mathbb{T}$, we get the measure $\widehat{\mu}$. Therefore, $\widehat{\mu}$ is a convex combination of the rotated measures. Hence if $\widehat{\mu}$ is ergodic, then $\tilde{\mu} = \widehat{\mu}$.

(a) $\Longrightarrow$ (b). Supposing $\widehat{\mu}$ is not ergodic, then (by Prop.7.1) it can be written as a convex combination of two invariant measures, $\widehat{\mu}_1$ and $\widehat{\mu}_2$. The only thing to check (to contradict μ-unique ergodicity) is that they project to μ. But whatever measure they project to must be absolutely continuous with respect to the projection of $\widehat{\mu}$, i.e. μ; hence by ergodicity of μ this equals μ.

(a) $\Longrightarrow$ (c). We cannot improve on Furstenberg's beautiful little argument. For convenience we switch to the multiplicative notation used by Furstenberg, writing the circle now as S^1, the set of complex numbers with modulus one.

Assume that $\widehat{\mu}$ is not ergodic for T_φ. Then (by [Bil65] p. 13) there exists an invariant non-constant measurable (real or complex)-valued function F in $L^2(X \times$

$\mathbb{T}, \widehat{\mu})$. By Fubini's theorem, for μ-a.e. circle fiber, F is in L^2 of that fiber. So there are Fourier coefficients $a_n(w)$ for the fiber over w, with $F(x,\theta) = \sum_{-\infty}^{\infty} a_n(x)\, e^{in\theta}$. Now calculating $F \circ T_\varphi$, uniqueness of the Fourier coefficients implies that $a_n(w) = a_n(Tw)\, e^{in\varphi(w)}$ for each n. By ergodicity of T, the modulus of each $a_n(w)$ is μ-a.s. constant. Since F is assumed non-constant, for some $k \neq 0$, $|a_k| \neq 0$ (μ-almost surely). So for that k, we can normalize a_k in the equation above. Then, changing back to additive notation, we define $u(w)$ by $e^{-iu(w)} = a_k(w)/|a_k|$. The equation then becomes $k\varphi = u \circ T - u$, proving (c). □

Remark 29.4. The set of k such that (c) holds is an ideal in $\mathbb{Z}$. If, say, ℓ generates this principal ideal, we can describe the collection of all $\widehat{T}$-invariant functions: they can be expressed as some combination of the G_k just defined, for all multiples k of ℓ.

Remark 29.5. The key argument of Furstenberg can be viewed as follows. The equation for cohomology (mod 1), $\varphi = u \circ T - u$, can if u has modulus one be written multiplicatively as $z(w) = a(w)/a(Tw)$ (this is *multiplicative* cohomology, and is the notation Furstenberg uses), giving equivalently $a(w) = a(Tw)z(w) = a(Tw)e^{i\varphi(w)}$. To produce an invariant graph we only need this to happen for one Fourier coefficient $a = a_k$, and for that it is enough for one to be nonzero and then normalize.

Corollary 29.9. *Let $(X, \mathcal{A}, \mu)$ be a probability space with T an ergodic measure preserving map from X to X and let $\varphi : X \to \mathbb{R}$ measurable. Write $S_n\varphi$ for the partial sums. Suppose there does not exist $k \in \mathbb{Z}$ such that $k\varphi$ is a coboundary (mod 1). Then for a.e. x, $S_n f(x)$ is uniformly distributed (mod 1).*

Proof. We build the circle-valued skew product T_φ on $\widehat{X} = X \times \mathbb{T}$ and apply $(c) \implies (a)$ of Proposition 29.8. Thus we know that if there does not exist u measurable such that $k\varphi = u \circ T - u (\mathrm{mod}\, 1)$, then the map T_φ is μ- uniquely ergodic hence ergodic.

Now let $F \in \mathcal{C}(\mathbb{T})$. We extend this to $\widehat{F} : \widehat{X} \to \mathbb{T}$ by $\widehat{F}(x,t) = F(t)$. This is a measurable function which is continuous on the fiber over x, for each x. Note that $\widehat{F} \in L^1(\widehat{X})$.

The Birkhoff ergodic theorem implies that for every such $\widehat{F}$, hence for each continuous function $F \in C(\mathbb{T})$, there is a full measure set E_F such that for every $(x,t) \in E_F \subseteq \widehat{X}$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} F \circ {T_\varphi}^n(x,t) = \int_{\widehat{X}} \widehat{F} d\widehat{\mu} = \int_{\mathbb{T}} F \mathrm{d}\lambda \tag{116}$$

Since $\mathbb{T}$ is compact, $C(\mathbb{T})$ is separable. Let $\{F_i\}_{i\in\mathbb{N}}$ be a dense subset of $C(\mathbb{T})$. Then (116) holds for every (x,t) in $\cap_{i\in\mathbb{N}} E_{F_i}$. This passes over to every $F \in C(\mathbb{T})$: we approximate F ε-uniformly by functions F_i, and both sides of (116) are within ε hence equal.

In conclusion, $\widehat{\mu}$-almost every (x,t) is a *fiber generic point*, that is, (116) holds for every $F \in C(\mathbb{T})$.

Suppose (x,t) is a fiber generic point, then $(x,\{t+s\})$ is also a fiber generic point for any $s\in\mathbb{T}$. This is because, for every $F\in C(\mathbb{T})$ defining $\widehat{F}_s(x,t)=\widehat{F}(x,\{t+s\})$,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\widehat{F}\circ T_f^n(x,t+s)=\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\widehat{F}_s\circ T_f^n(x,t)=\int_{\widehat{X}}\widehat{F}_s d\widehat{\mu}$$
$$=\iint_{\widehat{X}}\widehat{F}_s d\lambda d\mu=\iint_{\widehat{X}}\widehat{F}d\lambda d\mu=\int_{\widehat{X}}\widehat{F}d\widehat{\mu}=\int_{\mathbb{T}}F\mathrm{d}\lambda.$$

So the set of all fiber generic points is $E\times\mathbb{T}$ for some $E\subset X$. Since this set has full λ-measure, also $\mu(E)=1$. Now for every $x\in E$ $(x,0)$ is a fiber generic point. This means that for every $F\in\mathcal{C}(\mathbb{T})$,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\widehat{F}(T^nx,\{S_nf(x)\})=\int_{\mathbb{T}}F\mathrm{d}t.$$

Using continuous functions F to approximate $\mathbf{1}_{[a,b)}$ for $0\le a<b<1$, one then gets that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\mathbf{1}_{[a,b)}(\{S_nf(x)\})=b-a,$$

where $\{t\}$ denotes the fractonal part of $t\in\mathbb{R}$, so we conclude that $S_nf(x)$ is u.d. mod 1 for every $x\in E$.

Equivalently, there exists a set of measure 0, $N\subseteq X$, such that for every $w\in X\setminus N$, then for every $p\in\mathbb{T}$, the point (w,p) is a generic point for T_φ. Thus choosing any interval $J\subseteq\mathbb{T}$, the frequency of times spent in $X\times J$ is $\lambda(J)$, so the frequency of $S_n\varphi(p)$ in J is $\lambda(J)$ as claimed.

□

The proof of Cor.29.9 was worked out together with Xuan Zhang.

29.8. **Julia set scenery flow.** In the above case we took our skewing function to be $\varphi:X\to\mathbb{T}$. This can also be written multiplicatively, using $S^1\subseteq\mathbb{C}$ the complex numbers of modulus one.

In fact The material in the previous section comes from pp. 485-487 of [BFU02] where that is the case, as we explain.

The *model scenery flow* constructed there combines (28) and (**??**). There we consider as base map a rational function f on the complex plane, acting on its Julia set $\mathcal{J}\subseteq\mathbb{C}$. Now $\mathbb{C}$ is a one-dimensional complex manifold, with one chart (the identity), and with tangent space $\mathbb{C}$ itself, so $Df(z,w)=(f(z),(f'(z))w)$ defines a skew product on $\mathbb{C}\times\mathbb{C}$. Next, for $z\neq 0$, setting $\arg(z)=z/|z|)\in S^1$, then for $w\in S^1$,

$$(z,w)\mapsto(f(z),\arg(f'(z)\cdot w)$$

defines an S^1- valued skew product.

29.9. **Change of velocity in flows.**

29.10. **Orbit equivalence.** Consider an invertible map $T : X \to X$. The T-orbit of a point $x \in X$ is $\mathcal{O}_T(x) = \{T^n(x) : n \in \mathbb{Z}\}$; suppose there is another map S on the same space, with the same orbits. That means there exists $\varphi : X \to \mathbb{Z}$ such that $T(x) = S^{\varphi(x)}(x)$; more generally, there is a $\Phi : X \times \mathbb{Z} \to \mathbb{Z}$ such that

$$T^n(x) = S^{\Phi(x,n)}(x)$$

.

We conclude:

Proposition 29.10. *Invertible transformations (X, T) and (X, S) are orbit equivalent iff there exists $\varphi : X \to \mathbb{Z}$ such that $T(x) = S^{\varphi(x)}(x)$, iff there exists a cocycle $\Phi : X \times \mathbb{Z} \to \mathbb{Z}$ generated by φ, satisfying $T^n(x) = S^{\Phi(x,n)}(x)$.*

Proof. We must have:

$$S^{\Phi(x,n+m)}(x) = T^{m+n}(x) = T^m(T^n(x)) = S^{\Phi(T^n(x),m)}(x)$$

whence Φ is a cocycle and is the cocycle generated by φ.

Now suppose the maps are orbit equivalent; that is, there exists some bijection $\Psi : X \to X$ such that for every x, $\Psi(\mathcal{O}_T(x)) = \mathcal{O}_T(x)$. Then we can define a transformation S and a cocycle as follows.....

Conversely, given a bijection

□

measurability....

Kechris example??? Nonmeas cocycle? Nonmeas OE?

30. Cocycles in the thermodynamical formalism

We have already encountered the Ruelle Perron-Frobenius Theorem, one of the key tools in the *thermodynamical formalism* developed by Ruelle, Sinai and Bowen. In the next sections, we study this further, in particular the role of cocycles.

30.1. **Dependence on the future: Bowen's proof of Sinai's lemma.**

Proposition 30.1. *Let $\varphi : \Sigma_A \to \mathbb{R}$ be Hölder continuous. Then there exists $\widetilde{\varphi}$ which depends only on the future coordinates Σ_A^+ which is Hölder cohomologous to φ.*

First we give an idea of where the proof could have come from, and then enter the details (this is Lemma 1.6 of [Bow75]).

By the **stable segment** of a point $\underline{x} \in \Sigma_A$ we shall mean the set of all w such that $w_i = x_i$ for all $i \geq 0$. Pictorially this is represented in Fig. 74. We wish to find a function u such that for

$$\widetilde{\varphi} = \varphi + u \circ \sigma - u, \tag{117}$$

we have $\widetilde{\varphi}(w) = \widetilde{\varphi}(x)$ for each such stable segment. Now from this equality plus (117),

$$\widetilde{\varphi}(\underline{x}) = \varphi(\underline{x}) + u \circ \sigma(\underline{x}) - u(\underline{x}) = \widetilde{\varphi}(w) = \varphi(w) + u \circ \sigma(w) - u(w)$$

so subtracting,

$$\varphi(w) - \varphi(\underline{x}) = u \circ \sigma(\underline{x}) - u \circ \sigma(w) + u(w) - u(\underline{x}). \tag{118}$$

FIGURE 74. The stable segment of a point for the coding by a Markov partition for a toral automorphism

Adding up along an orbit of length k, we have a telescoping sum and get

$$S_k\varphi(w) - S_k\varphi(\underline{x}) = u \circ \sigma^k(\underline{x}) - u \circ \sigma^k(w) + u(w) - u(\underline{x}). \tag{119}$$

We shall define u in such a way as to be continuous. Then $u\circ\sigma^k(\underline{x}) - u\circ\sigma^k(w) \to 0$ as $k \to \infty$ since $\underline{x}, w$ are in the same stable segment (and hence are forward asymptotic), and hence (119) gives us

$$u(w) - u(\underline{x}) = \lim_{k\to\infty} S_k\varphi(w) - S_k\varphi(\underline{x}). \tag{120}$$

Now suppose we choose $u(w)$ to be equal to 0, for one particular string in each stable segment. A convenient way to do that is to choose this w to only depend on x_0, the 0^{th} coordinate of $\underline{x}$. Call this choice $w = \gamma(\underline{x})$. Then equation (120) yields the following definition of u:

$$u(\underline{x}) = \lim_{k\to\infty} S_k\varphi(\underline{x}) - S_k\varphi(\gamma(\underline{x})). \tag{121}$$

So all we have to do is *start* with this definition, prove that u so defined is indeed continuous, and then, running the logic backwards, we will be done.

Here are the details.

Proof. For each symbol j in the alphabet $\mathcal{A}$, we choose an infinite past string $a^-(j) = (\dots a_{-2}a_{-1}a_0 = j)$. We then define $\gamma : \Sigma_A \to \Sigma_A$ by $\gamma(\dots x_{-1}x_0x_1x_2\dots)) = (a^-(x_0)x_1x_2\dots)$. Next we define u as in (121), so

$$u(\underline{x}) = \sum_{j=0}^{\infty} \varphi(\sigma^j\underline{x}) - \varphi(\sigma^j\gamma(\underline{x})).$$

Note that indeed $u(\gamma(\underline{x})) = 0$, since γ is a projection i.e. $\gamma(\gamma(\underline{x})) = \gamma(\underline{x})$. Next, since $\sigma^j\underline{x}$ and $\sigma^j\gamma(\underline{x})$ are equal on coordinates in $[-j, +\infty)$,

$$|\varphi(\sigma^j\underline{x}) - \varphi(\sigma^j\gamma(\underline{x}))| \le \mathrm{var}_j\varphi \le b\alpha^j.$$

Hence the limit exists, and u is a continuous function.

We then define φ from this by the cohomology equation (117). And by the argument before the proof, $\widetilde{\varphi}$ is, indeed, constant on each stable segment, as desired. It remains to show u is in fact α-Hölder, which will imply that $\widetilde{\varphi}$ is α-Hölder as well.

We could now conclude the proof by calling on Prop. **??**.

Or one can argue directly; we copy the argument from [Bow75]:

Now, for any $\underline{y}$ such that $x_i = y_i$ for $-n \leq i \leq n$, then

$$|u(\underline{x}) - u(\underline{y})| \leq \sum_{j=0}^{\left[\frac{n}{2}\right]} |\varphi(\sigma^j \underline{x}) - \varphi(\sigma^j \underline{y}) + \varphi(\sigma^j \gamma(\underline{y})) - \varphi(\sigma^j \gamma(\underline{x}))| + 2 \sum_{j > \left[\frac{n}{2}\right]} b\alpha^j \quad (122)$$

$$\leq 2b\Big(\sum_{j=0}^{\left[\frac{n}{2}\right]} \alpha^{n-j} + \sum_{j > \left[\frac{n}{2}\right]} \alpha^j\Big) \leq \frac{2b^{\left[\frac{n}{2}\right]}}{1-\alpha}. \quad (123)$$

Therefore u is in fact α-Hölder. Hence $\widetilde{\varphi}$ is α-Hölder as well. □

30.2. **Sinai's lemma; the proof of Sinai-Ratner.**

30.3. **Conditions which guarantee coboundaries: Livsic theory.** Suppose there exists a measurable function $u : \Omega \to S^1$ such that

$$(*) \qquad \varphi(\underline{x}) = u \circ T(\underline{x}) - u(\underline{x}) \qquad (\text{for } \mu\text{-a.e. } \underline{x}).$$

And suppose now we know moreover that $\varphi : \Omega \to S^1$ is not just measurable but is continuous. We will say φ is **cohomologous to zero in the class of** e.g. measurable, or continuous, functions if there exists u in that class with $\varphi(\underline{x}) = u \circ T(\underline{x}) - u(\underline{x})$.

Livsic proved in [Liv71], [Liv72] the equivalence in various situations of the following:
(1) φ is cohomologous to zero in the class of continuous functions
(2) φ is cohomologous to zero in the class of measurable functions
(3) $S^n\varphi(\underline{x}) = 0$ for each periodic point (where n is the period of $\underline{x}$).

Roughly speaking, the proofs depend on these hypotheses:
– T is hyperbolic (e.g. an Anosov diffeomorphism, or a shift map)
– φ is Hölder
(and then (1) states that there exists u Hölder).

In [Liv71] a proof in the order (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (1) is given. Of course (1) $\Longrightarrow$ (2) is trivial. See Theorem 1 in [Liv71] (and below) for (3) $\Longrightarrow$ (1). In the proof of (2) $\Longrightarrow$ (3) (Remark 2 of [Liv71]) it is assumed that u is essentially bounded and that the invariant measure μ (with respect to which the coboundary equation holds μ-a.s.) is positive on open sets. In [Liv72] a completely different, direct proof of (2) $\Longrightarrow$ (1) is given. Livsic proves this for real-valued cocycles over an Anosov diffeomorphism with μ smooth measure. This proof has the advantage that one no longer needs to assume that u is bounded (measurable is enough). However now more is required of the measure: the essential property that gets used is what is known in hyperbolic dynamics as *absolute continuity*. This is the equivalence of the conditional measures on e.g. two different stable leaves, mapped by the holonomy homeomorphisms given by sliding along the unstable leaves.

This method was extended (again for real-valued cocycles) in [PUZ89] to our situation, a mixing conformal repellor. There, the absolute continuity condition is replaced by a weaker condition (see below).

For the case we are interested in at present, circle-valued cocycles, of course boundedness is not an issue. However now Livsic' proof of (2) $\implies$ (3) fails (a simple counterexample is given in [**?**]) ???. Fortunately the second method still works - so our proof of equivalence will be (2) $\implies$ (1), (3) $\implies$ (1) (and of course the easy reverse implications). Our proof of (2) $\implies$ (1) is based on [PUZ89]. We give full details here for completeness and clarity.

From now on Ω is the natural extension of a mixing conformal repellor $T : J \to J$, with shift map σ on $\Omega \subseteq \Pi_{-\infty}^{\infty} J$. Let μ be an ergodic invariant measure on Ω. We will state the conditions needed on μ for the proof; this will be satisfied for the case we are actually interested in (the Gibbs state of $d \log |DT|$). To describe this hypothesis, we need to define conditional measures of μ with respect to the "past".

We write $\Pi_0^{\geq k}$ for the subset of $\Pi_k^{\infty} J$ consisting of allowed strings. We define $\pi_k : \Omega = \Pi_0 \to \Pi_0^{\geq k}$ to be the natural projection, i.e. $\pi_k(\underline{z}) = (z_k, z_{k+1}, \dots)$, and will also write π for π_0. Since $\Pi_0^{\geq 0}$ is naturally identified with J, we will also think of π as a map from Ω to J.

Lemma 30.2. *Let μ be an ergodic and invariant probability measure on Ω. Let $\varphi : \Omega \to S^1$ be a measurable function such that $\mu-$ almost surely φ only depends on J; that is, there exists an invariant set G of full measure such that for $\underline{w}$, $\underline{z}$ in G with $\pi_0(\underline{w}) = \pi_0(\underline{z})$, then $\varphi(\underline{w}) = \varphi(\underline{z})$. Assume that there exists $u : \Omega \to S^1$ measurable with $\varphi = u \circ \sigma - u$. Then u also depends only on J (μ-a.s.).*

Proof. We will show that there exists an invariant set $\widetilde{G} \subseteq \Omega$ such that for $\underline{x}$, $\underline{y} \in \widetilde{G}$ with $\pi(\underline{x}) = \pi(\underline{y})$, then $u(\underline{x}) = u(\underline{y})$. We recall Lusin's Theorem, which states that a measurable function is almost uniformly continuous. Thus, given $\varepsilon > 0$ (we will fix some $\varepsilon < \frac{1}{2}$) there exists $C \subseteq \Omega$ with measure $\geq 1 - \varepsilon$ such that u is uniformly continuous on C. By the Birkhoff ergodic theorem, there is a set $B \subseteq \Omega$ of full measure such that C is sampled well by every $x \in B$, i.e. the density of time x spends in C is equal to $\mu(C)$, so in particular is greater than $\frac{1}{2}$. Now let $\widetilde{G} = B \cap G$; this is an invariant set of measure one and we can assume also that the coboundary equation holds on all of $\widetilde{G}$. Let $\underline{x}$, $\underline{y}$ in $\widetilde{G}$ be in the same π- fiber. We have:

$$u(\underline{x}) - u(\underline{y}) = S^n \varphi(\underline{x}) - S^n \varphi(\underline{y}) + u \circ \sigma^n(\underline{x}) - u \circ \sigma^n(\underline{y})$$

which equals $u \circ \sigma^n(\underline{x}) - u \circ \sigma^n(\underline{y})$ since φ depends on J. Since two subsets of the integers with density greater than $\frac{1}{2}$ meet infinitely often, there is a subsequence of times such that u^{n_j} of $\underline{x}$ and $\underline{y}$ are simultaneously in C. Now $\underline{x}$ and $\underline{y}$ are in the same stable set, i.e. $\sigma^n(\underline{x}) - \sigma^n(\underline{y}) \to 0$ (in the circle) as $n \to \infty$. By uniform continuity therefore, the right-hand side goes to 0 as $j \to \infty$, hence $u(\underline{x}) = u(\underline{y})$. □ □

Proposition 30.3. *Let $\varphi : \Omega \to S^1$ be Hölder continuous and such that φ only depends on J. Assume there exists $u : \Omega \to S^1$ measurable such that $\varphi = u \circ \sigma - u$ (μ-almost surely). Then u is also Hölder continuous (after redefinition on a μ-null set).*

Proof. By the Lemma, there is an invariant set G of full measure such that the coboundary equation is satisfied and such that for $\underline{x}$, $\underline{y} \in G$ with $\pi(\underline{x}) = \pi(\underline{y})$, we know that $u(\underline{x}) = u(\underline{y})$.

Again, let C be a set of continuity for u from Lusin's theorem and B a good set from the Birkhoff theorem, but this time for time averages going toward $-\infty$.

Now let $\underline{x},\ \underline{y} \in B \cap G$ with $\text{dist}(\underline{x},\ \underline{y}) < \delta$. We will show that there exists $c > 0,\ \gamma \in (0,1)$ such that $|u(\underline{x}) - u(\underline{y})| \leq c \cdot \text{dist}(\underline{x},\ \underline{y})^\gamma$. This will finish the proof, since if u is uniformly Hölder on a dense set, it extends to a Hölder continuous function on Ω. By continuity therefore, the equation $\varphi = u \circ \sigma - u$ will be valid on all of Ω. And finally, since $B \cap G$ has full measure, u will have been changed on a set of measure zero, as claimed.

We will show $|u(\underline{x}) - u(\underline{y})| \leq c|x_0 - y_0|^\gamma$, which implies the previous inequality. Here $x_0,\ y_0 \in J$ are the projections of $\underline{x},\ \underline{y}$ i.e. the 0^{th} coordinate in $\Omega \subseteq \Pi_{-\infty}^{\infty} J$, and $|\cdot|$ denotes distance in S^1 and in $\mathbb{C}$ on the left and right sides respectively.

Let U be an open set in J containing x_0 and y_0, with diameter less than δ. Since $B \cap G$ has full measure in Ω, by the Lemma ??, there exists some past fiber F over U such that $F \cap (B \cap G)$ has full measure with respect to μ_F. Let $\tilde{\underline{x}},\ \tilde{\underline{y}}$ in F project to $x_0,\ y_0$ in J. Since F is an unstable set, for some $c > 0$ and some $\lambda \in (0,1)$, we have $|\pi(\sigma^{-n}(\tilde{\underline{x}})) - \pi(\sigma^{-n}(\tilde{\underline{y}}))| \leq c\lambda^n$ for all $n \geq 0$. Since φ depends on $\mathcal{J}$ and is Hölder continuous, with some exponent γ, we can take

$$|S^{-n}\varphi(\tilde{\underline{x}}) - S^{-n}\varphi(\tilde{\underline{y}})| \leq \sum^{n} |\varphi(\sigma^{-k}\tilde{\underline{x}}) - \varphi(\sigma^{-k}\tilde{\underline{y}})| \leq ???$$

Now since $\underline{x},\ \underline{y}$ and $\tilde{\underline{x}},\ \tilde{\underline{y}}$ are in G, we have

$$u(\underline{x}) - u(\underline{y}) = u(\tilde{\underline{x}}) - u(\tilde{\underline{y}}) = S^{-n}\varphi(\tilde{\underline{y}}) - S^{-n}\varphi(\tilde{\underline{x}}) + u \circ \sigma^{-n}(\tilde{\underline{x}}) - u \circ \sigma^{-n}(\tilde{\underline{y}}).$$

By the same argument as in Lemma ???? above, since $\tilde{\underline{x}},\ \tilde{\underline{y}}$ are in B,

$$|u \circ \sigma^{-n}(\tilde{\underline{x}}) - u \circ \sigma^{-n}(\tilde{\underline{y}})| \to 0$$

along a subsequence. Therefore for all $\underline{x},\ \underline{y}$ in $B \cap G$,

$$u(\underline{x}) - u(\underline{y}) \leq \lim_{n \to \infty} |S^{-n}\varphi(\tilde{\underline{y}}) - S^{-n}\varphi(\tilde{\underline{x}})| \leq c|x_0 - y_0|^\gamma,$$

as claimed. □ □

$\underline{\varphi}(x) = u \circ T(x) - u(x)$ for μ-a.e. x
$\varphi :\ \Omega \to S^1$ measurable: (a) there exists u measurable with $\varphi(x) = u \circ T(x) - u(x)$
???
$u(\underline{x}) - u(\underline{y}) = S^n\varphi(\underline{x}) - S^n\varphi(\underline{y}) + u \circ \sum^n(\underline{x}) - u \circ \sum^n(\underline{y})$
with $\pi(\underline{x}) = \pi(\underline{y})$, then $u(\underline{x}) = u(\underline{y})$.

30.4. **Some consequences of the Ruelle Perron-Frobenius Theorem.** From the Ruelle-Perron-Frobenius theorem one then harvests a number of important results. First, using the uniform convergence, it follows quite easily (Lemma 1.14 in [Bow75]) that:

Lemma 30.4. *There exists $c > 0$ such that for $P, Q \in \mathbb{C}_0^k$,*

$$|\mu(P \cap \sigma^{-t}Q) - \mu P \mu Q| \leq c \cdot \mu P \mu Q \beta^t.$$

We recall that the transformation σ is *mixing* if for all cylinder sets $P, Q \in \mathbb{C}_0^k$, for each k,

$$\mu(P \cap \sigma^{-t}Q) \to \mu P \mu Q$$

as $t \to \infty$. The transformation is **weak Bernoulli** (with respect to the standard partition $\{[0], [1]\}$) if this much stronger fact holds: $\forall k$,

$$\sum_{P,Q} |\mu(P \cap \sigma^{-t}Q) - \mu P \mu Q| \to 0$$

as $t \to \infty$, where the sum is taken over all $P, Q \in \mathbb{C}_0^k$. By a theorem of Ornstein [Orn73], [Shi73], an invertible transformation which has a weak Bernoulli generating partition is Bernoulli, i.e. is measurably isomorphic to a Bernoulli shift.

Hence as an immediate corollary of Lemma 30.4 one has (Theorems 1.14 and 1.25 in [Bow75]):

Theorem 30.5. *The shift map σ with measure μ is mixing.*

and

Theorem 30.6. *The transformation (σ, μ) is weak Bernoulli, hence (by Ornstein's theorem) its natural extension is Bernoulli.*

With slightly more work it follows (1.26 in [Bow75]) that:

Theorem 30.7. *For $\alpha \in (0, 1)$ there exists $D > 0$ and $\eta \in (0, 1)$ such that for all $f, g \in \mathcal{H}_\alpha$,*

$$\left| \int f \cdot g \circ \sigma^t \, d\mu - \int f \, d\mu \int g \, d\mu \right| \leq D||f||_\alpha ||g||_\alpha \eta^t.$$

This is called the **Exponential Cluster Property** by Bowen; it is also known as having an **exponential decay of correlations**.

Hölder functions appear in two ways in the theory: as potential functions land as an observable, making a measurement on the system and for which one wants to study time averages, correlations and so on. Now we encounter this second situation.

This is the next theorem in [Bow75], the Central Limit Theorem, for Hölder observables; no proof is given there, rather Ratner's paper [Rat73] is cited; here we write $S_n f$ for the function $S_n f(\underline{x}) \equiv \sum_{j=0}^{n-1} f(\sigma^j(\underline{x}))$:

Theorem 30.8. *(Central Limit Theorem) For f Hölder and with $\int f \, d\mu = 0$ there exists $\sigma \in [0, \infty)$ such that for all real r,*

$$\mu\{S_n f / n^{\frac{1}{2}} > r\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{r} e^{-x^2/2\sigma^2} dx.$$

We give a proof in §30.5 below.

Note: If the variance $\sigma^2 = 0$, it is understood that the Gaussian distribution is replaced here by the limiting distribution as $\sigma^2 \to 0$, that is, by point mass at 0.

30.5. **Gordin's proof of the CLT.**

31. More on Cocycles

31.1. **Cohomology and nonsingular transformations.** There are two main ways to generalize the standard ergodic theory situation of a measure-preserving transformation on a probability space: to *infinite* invariant measures, and to **measure-class preserving** transformations, i.e. maps which preserve the collection of null sets but not the measure itself. In this last case the measure is called **quasi-invariant**, and such a map $(X, \mathcal{A}, T, \mu)$ is called a **nonsingular** transformation.

To introduce the subject, we begin with the simplest example, derived from a measure-preserving transformation $(X, \mathcal{A}, T, \nu)$. Now let $f(x) \in L^1(X, \nu)$ and assume also that $f(x) > 0$ almost surely. We define a measure μ by $d\mu = f(x)d\nu$, by which we mean

$$\mu(A) = \int_A d\mu = \int_A \frac{d\mu}{d\nu} d\nu = \int_A f(x) d\nu$$

for any $A \in \mathcal{A}$, so in other words, $f = d\mu/d\nu$ is the Radon-Nikodym derivative of μ with respect to ν.

Then μ is invariant if and only if the function f is. So for non-invariant f, the map (X, T, μ) is a nonsingular transformation. One case where this could be useful is when ν is infinite, for then $f \in L^1$ produces an equivalent finite quasi-invariant measure μ. But in general, any questions about the nonsingular transformation can be answered by considering the invariant measure ν, with which it is much easier to work.

The situation is quite different if there exists *no* equivalent (finite or infinite) invariant measure. There the nonsingularity is intrinsic, so we have to develop a truly new theory. (These three types of transformations-finite or infinite measure-preserving, or truly nonsingular- were termed **Types I, II, and III** by von Neumann; they play a basic role in C^*-algebra theory and in the study of the equivalence relation generated by orbits of the map).

Of course the first problem will be to identify in some way which nonsingular transformations are *truly* nonsingular in this sense. For that a basic tool is the Radon-Nikodym cocycle.

For simplicity, we assume that T is invertible, so $\mu \circ T$ makes sense, and we define a measurable function $R_\mu(x) : X \to (0, +\infty)$ by

$$R_\mu(x) = \frac{d\mu \circ T}{d\mu}.$$

This defines a **multiplicative cocycle** i.e. taking values in the multiplicative group $(\mathbb{R}^{>0}, \cdot)$. It will also be useful for us to consider the additive version of this:

$$\varphi_\mu \equiv \log R_\mu.$$

As before, these functions are called cocycles because they extend to a function on $X \times \mathbb{Z}$ by multiplying, or summing, along an orbit.

Proposition 31.1. *Given a quasi-invariant measure μ for $(X, \mathcal{A}, T)$, then there exists an equivalent invariant measure iff ϕ_μ is a coboundary, R_μ is a multiplicative coboundary.*

Proof. The condition is that there exist a measurable function u such that $\varphi_\mu(x) = u \circ T(x) - u(x)$, i.e. that φ be cohomologous to 0, and equivalently that R_μ be multiplicatively cohomologous to 1. But

$$\mu \circ T(A) = \int_A d(\mu \circ T) = \int_A \frac{d\mu \circ T}{d\mu}$$

while

$$\mu(A) = \int_A d\mu.$$

Now R_μ is cohomologous to 1 means there exists $w(x)$ such that
so these are equal for all A iff ... □

Example 30. Here is an interesting (and illustrative) example of a nonsingular transformation. Let Σ_A be a subshift of finite type with A a primitive $0-1$ matrix, and let $\mathcal{O}$ be an order on the corresponding Bratteli diagram, with $T = T_\mathcal{O}$ the adic transformation on Σ_A^+ defined fron this. Now since A is primitive, we know by Theorem ?? that (Σ_A^+, T) has a unique invariant probability measure. To produce a quasi-invariant measure, let $\varphi : \Sigma_A \to \mathbb{R}$ be continuous, and supppose the Ruelle operator $\mathcal{L}_\varphi$ has eigenvalue $\lambda > 0$. We know (again by the primitivity of A that there exists a unique normalized eigenmeasure ν with eigenvalue λ, for the dual operator $\mathcal{L}_\varphi^*$. By Prop. 27.1, equivalently we have

$$\frac{d\nu \circ \sigma}{d\nu} = \lambda e^{-\varphi}.$$

This .....

31.2. **Maharam's skew product and the flow of weights.** As an aid to understanding a nonsingular transformation $(X, \mathcal{A}, T, \mu)$, Maharam introduced an associated measure-preserving transformation, a skew product with real fibers.

To describe this we first build a skew product taking values in the group of multiplicative reals $(\mathbb{R}^{>0}, \cdot)$, this being given Lebesgue measure λ. Since λ is not invariant by dilation, the idea is that a dilation in the fiber can compensate for the nonsingularity of μ, as follows. Defining as before $R_\mu = d\mu \circ T / d\mu$, we take as our skewing function R_μ^{-1}, and so our multiplicative skew product is:

$$T_\Phi(x, a) = (Tx, R_\mu^{-1}(x) \cdot a).$$

This has Radon-Nikodym derivative

$$\frac{d(\mu \times \lambda) \circ T_\Phi}{d(\mu \times \lambda)} = \frac{d\mu \circ T}{d\mu} \cdot \frac{d(R_\mu^{-1}(x) \cdot \lambda)}{d\lambda} = R_\mu(x) \cdot R_\mu^{-1}(x) = 1$$

so the measure is preserved.

It will be convenient to write this is additve form, with fibers $(\mathbb{R}, +)$, so the skewing function is $\varphi \equiv \log \Phi = \log R_\mu^{-1}(x) = -\log R_\mu(x)$, and the measure λ then becomes m with $\mathrm{d}m = e^x \mathrm{d}x$ on $\mathbb{R}$. This last fact can be seen as follows: $m([0, s]) = \lambda([e^0, e^s]) = e^s - 1 = \int_0^s e^x \mathrm{d}x$.

32. Ergodic theory and the boundary at infinity of a group

Furstenberg's idea of the boundary of a group can best be motivated by two apparently quite different examples: Brownian motion on the disk and the free group (and semigroup) on two generators.

If we start a Brownian path $B(t)$ at the center of the unit disk D, it will a.s. at some time t_0 encounter the boundary circle. At that point we stop the path. Now it is a fact that Brownian motion is invariant under conformal coordinate changes, so if we move this picture to the hyperbolic disk Δ (the interior of the unit disk with the Poincaré metric), then this same path is also a Brownian path in the new metric, up to a time change. By this time change t_0 has become $+\infty$, so we have proved that Brownian motion in the Poincaré disk a.s. converges to a point on the boundary, defined to be simply the boundary circle in this case.

As usual, one would like to study random walk approximations to this Brownian motion. Now the group $G = PSL(2, \mathbb{R})$, the real Möbius transformations, are the orientation-preserving isometries of the Poincaré disk (see §**??**). So, just as we consider random walks in the integer lattice $\mathbb{Z} \oplus \mathbb{Z}$ as approximations to Brownian motion in the plane, we could study random walks in a discrete subgroup $\Gamma \subset G$. A classical example is the **modular group**, the subgroup with integer entries. This is one of Furstenberg's starting points, and also will be of the main interest in this paper. But first we consider a simpler discrete model, the free group.

It is in fact best to take first not a group but a *semi*group, the free semigroup on two generators a, b, written FS_2. We consider the simplest one-sided infinite paths, those with increments $x_0, x_1, \cdots$ in the set of generators $\{a, b\}$. We picture the semigroup as the nodes of an infinite binary tree; the base node is the identity $e \in FS_2$. The collection of all infinite paths corresponds to $\Sigma^+ = \Pi_0^\infty \{a, b\}$, the Bernoulli shift space; we think of the boundary points as the "ends" of such a path, which can be identified with the path itself since there is a unique way to get there. Thus $\partial FS_2 = \Sigma^+$; we give $\{a, b\}$ the discrete topology and Σ^+ the corresponding product topology, which makes Σ^+ homeomorphic to the Cantor set. We then topologise $FS_2 \cup \partial FS_2$ so as to give the following notion of convergence: for $g_n \in FS_2$ and $x \in \partial FS_2$, $g_n \to x$ as $n \to \infty$ iff for some k_n increasing to infinity, $g_n = x_0 \dots x_{k_n}$.

In this example, there is no cancellation, so you can only proceed directly out to the boundary. Note that FS_2 acts on $FS_2 \cup \partial FS_2$ continuously, by left multiplication.

There is a natural measure on ∂FS_2, the Bernoulli $(\frac{1}{2}, \frac{1}{2})$ measure $\times_{i=0}^\infty (\frac{1}{2}\delta_a + \frac{1}{2}\delta_b)$. Here is another way to define convergence, which is clearly equivalent: $g_n \to x$ iff $g_n(\mu) \to \delta_x$ (where $g_n(\mu)$ denotes the push-forward of the measure by the left multiplication). This is the definition taken by Furstenberg, as it will work well in the case of a general group. We note that μ is the hitting measure of the independent $(\frac{1}{2}, \frac{1}{2})$ random walk on FS_2.

The next case to consider is the free group, F_2. Here steps along a path can cancel, so following the semigroup case, we define the boundary to be the collection of one-sided infinite cancelled sequences in the symbols $\{a, b, a^{-1}, b^{-1}\}$. This space is a subshift of finite type Σ_A^+, with a (4×4) $0-1$ matrix A expressing the simple rule

that a^{-1} cannot follow a and so on. Again there is a natural measure, which we now describe.

The one-sided random walk paths correspond to $\Pi^+ = \Pi_0^{+\infty}$, which gives the sequence of increments x_i. We write ν for the $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ Bernoulli measure on this shift space.

Given a sequence $\underline{x} = (x_0, x_1, \dots) \in \Pi^+$, we define $l_n(\underline{x})$ to be the length k of the collapsed word $w_0 w_1 \cdots w_k = x_0 x_1 \cdots x_n$, so e.g. for $x = (.aa^{-1}baa^{-1}ba^{-1})$ then $w_0 w_1 w_2 = bba^{-1}$ and $l_5(\underline{x}) = 2$. We have:

Lemma 32.1. *The collapsed length $l_n(\underline{x}) \to +\infty$ as $n \to \infty$ for ν-a.e. $x \in \Pi^+$.*

Proof. Note that the length of a word increases with probability $3/4$ unless $l_0 = 0$ (in which case the length surely increases). Thus l_n is the path of a reflected random walk on the integers with drift, and so by the law of the iterated logarithm, $l_n(\underline{x}) \to +\infty$ a.s. □

Proposition 32.2. *The map $\pi : \Pi^+ \to \Sigma^+$ given by cancellation is well- defined for ν-a.s. $\underline{x} = (x_0, x_1, \dots) \in \Pi^+$. The image μ of the measure ν is the measure of maximal entropy on Σ_A^+, the Parry-Shannon measure; for a.e. one-sided random walk path (g_n), $g_n \to w$ for the point $w \in \partial F_2 = \Sigma_A^+$ which is just the infinite cancelled string equal to $\pi(\underline{x})$. Define a left action of the group F_2 on the topological space $F_2 \cup \Sigma_A^+$; on F_2 this is defined by left multiplication and on Σ_A^+ is left multiplication followed by cancellation. This action is continuous. The measure μ is equal to the hitting measure of the independent $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ random walk on F_2.*

Proof. The map is well-defined by the lemma. The action is clearly continuous. The Parry-Shannon measure (which is the measure of maximal entropy) is the Markov measure defined by the invariant row vector $\pi = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ and the matrix $\frac{1}{3} \cdot A$, and this is the push-forward of ν by the map π since all three choices which do not cancel with the last letter are equally likely. This implies that μ is the hitting measure. □

Let us write $\Pi^+ = \Pi_0^\infty F_2$ for the one-sided paths $(g_n)_{n \geq 0}$ in F_2. Then the cancellation map π extends to Π^+; those paths in Π^+ which have increments in the set of generators are determined by the full shift space Σ^+ of those increments. We note that for $\underline{g} \in \Pi^+$ the map π gives a gathering of $\underline{g}$; one has trimmed off all the excess branches without changing the limit point $\underline{x} \in \partial F_2 = \Sigma^+$.

Remark 32.1. Let us consider a left action of a group G on a set S, with $x \mapsto g(x)$ for $x \in S$. In the definition of paths in a group, a key point is that one does not get from $g_n(x)$ to $g_{n+1}(x)$ by application of the single element x_{n+1}. Rather, $g_{n+1}(x) = g_{n+1} g_n^{-1} g_n(x)$. This is exactly like what happens for Kleinian groups, whose limit set is the boundary at infinity (see **??**, or hyperbolic Cantor sets (see e.g. **??**). The Cantor set is naturally identified with the boundary at ∞ of the free semigroup on two generators, so there, the descent to one more generation requires us to "go all the way back up the tree" literally, since its Cayley graph is the binary tree and the boundary corresponds to the limiting ends.

We remark that many familiar objects can be modelled as group or semigroup boundaries, or as a quotient (identification space) of those. Examples already give

are the Cantor set (corresponding to ∂FS_2), and the limit set of a Kleinian group; others are any limit set of an iterated function system (IFS), and any Julia sets (consider the tree of inverse images of a point as it converges to the Julia set). For the general theory regarding the boundary at infinity of a group or homogeneous space, see [Fur71], [Fur80], [Kai97].

33. Extension of measures

33.1. **Extension of measures; from finite to countable additivity.** Here we present fundamental results on the extension of finitely and countably additive measures, due to Caratheodory, Alexandroff, Kolmogorov, and Doob- Kakutani-Nelson-Dudley.

References for this section are [Bar66], [Hal76], [DS57], [Roy68], [Loè77]).

We now go more deeply into ideas introduced in §3:

Definition 33.1. Given a set X, a *set function* is a function defined on a collection of subsets, taking values in some linear space (i.e. vector space): the real numbers $\mathbb{R}$, the complex numbers $\mathbb{C}$, or a Banach space B; in the case of the reals we also allow the the extended reals $\overline{\mathbb{R}} = [-\infty, +\infty]$. If values are taken in $\mathbb{R}^+ = [0, +\infty]$ we say this is a *positive set function.*

Given a set X, an **algebra** $\mathcal{A}$ is a collection of subsets of A such that:

- $X \in \mathcal{A}$;
- $A \in \mathcal{A} \implies A^c \in \mathcal{A}$;
- $A, B \in \mathcal{A} \implies A \cup B \in \mathcal{A}$.

  It is a $\sigma-$**algebra** if in addition
- $A_i \in \mathcal{A}$ for $i = 1, 2, \cdots \implies \cup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Given an algebra $\mathcal{A}$, a **finitely additive measure** on $\mathcal{A}$ is a positive set function $\mu : \mathcal{A} \to [0, +\infty]$ such that $\mu(\emptyset) = 0$ and for A, B disjoint, $\mu(A \cup B) = \mu(A) + \mu(B)$. For a **countably additive measure**, also called a $\sigma-$ **additive measure**, or simply a **measure** we require that $\mathcal{A}$ be a $\sigma-$algebra, satisfying that for $\{A_i\}_{i=1}^{\infty}$ disjoint, then $\mu(\cup_{i=1}^{\infty} A_i) = \Sigma_{i=1}^{\infty} \mu(A_i)$. Thus if we have only finite additivity this will always be stated explicitly as countable additivity is the usual situation.

In the case where the set function μ takes values in $\overline{\mathbb{R}}$ we call this a **signed measure** or a **charge**. If the values are in $\mathbb{C}$ or B we call this a **complex-** or $B-$**valued** measure.

There is a concept in between finite and countable additivity: μ is a **countably additive measure on an algebra** $\mathcal{A}$ iff countable additivity holds whenever it makes sense; that is, if $\{A_i\}_{i=1}^{\infty}$ are disjoint sets in $\mathcal{A}$, and if their union happens to also be in $\mathcal{A}$, then $\mu(\cup_{i=1}^{\infty} A_i) = \Sigma_{i=1}^{\infty} \mu(A_i)$.

Countable additivity is equivalent to finite additivity plus a certain continuity property. To make this explicit (see also e.g. p.85 of [Loè77]), a set function on a $\sigma-$ algebra is said to be **continuous from below** or $\sigma-$**continuous** iff given an increasing sequence of sets $A_i \in \mathcal{A}$ with $A = \cup_{i=1}^{\infty} A_i \in \mathcal{A}$ then $\lim \mu(A_i) = \mu(A)$. It is **continuous from above** or $\delta-$**continuous** iff given a decreasing sequence of sets $A_i \in \mathcal{A}$ such that some A_n has finite measure, if $A = \cap_{i=1}^{\infty}$ and $\mu(A_1) < \infty$, then $\lim \mu(A_i) = \mu(A)$.

To pass from countable additivity to $\sigma-$continuity we *disjointify* the sequence of sets, a useful technique illustrated in the proof to follow.

Lemma 33.1. *If a set function μ on an algebra $\mathcal{A}$ is countably additive, then it is $\sigma-$continuous.*

Conversely, if it is $\sigma-$continuous and is finitely additive, then it is countably additive.

Proof. Given $(A_i)_{i=1}^{\infty}$ increasing, we define $\widehat{A}_1 = A_1, \widehat{A}_2 = A_2 \setminus A_1, \widehat{A}_3 = A_3 \setminus (A_1 \cup A_2)$ and so on, so $(\widehat{A}_i)_{i=1}^{\infty}$ are disjoint and $A_n = \cup_{i=1}^{n} \widehat{A}_i$. Then the $\sigma-$additivity immediately gives $\sigma-$continuity.

Conversely, given A_i disjoint then $\widehat{A}_n = \cup_{i=1}^{n} A_i$ increase to $A = \cup_{i=1}^{\infty} A_i$ whence $\sigma-$continuity implies $\sigma-$ additivity. □

Lemma 33.2. *If $\mu(A) < \infty$, then finite additivity of the measure implies $\mu(A \setminus B) = \mu(A) - \mu(B)$.*

(If $\mu(A) = \mu(B) = \infty$, then we can draw no conclusion, since $\infty - \infty$ is not defined, and indeed from easy examples any real value for $\mu(A \setminus B)$ is possible).

Lemma 33.3. *Assume a set function μ is finitely additive. Then $\sigma-$ continuity implies $\delta-$continuity. If $\mu(X) < \infty$, then $\delta-$continuity implies $\sigma-$ continuity.*

Proof. We assume $\sigma-$continuity. Given $(A_i)_{i=1}^{\infty}$ decreasing, then for $A = \cap_{i=1}^{\infty} A_i$ we have that $(A_1 \setminus A_k)$ increases to $A_1 \setminus A$, so by $\sigma-$continuity $\mu(A_1 \setminus A_k) \to \mu(A_1 \setminus A)$. Say without loss of generality that $\mu(A_1) < \infty$. Then by Lemma 33.2, $\mu(A_1 \setminus A_k) = \mu(A_1) - \mu(A_k)$ and $\mu(A_1 \setminus A) = \mu(A_1) - \mu(A)$. Thus $\mu(A_1) - \mu(A_k) \to \mu(A_1) - \mu(A)$ whence $\mu(A_k) \to \mu(A)$ since all of these numbers are finite.

Conversely, we assume $\delta-$continuity. Given A_k increasing to A, then $A \setminus A_k$ decreases to the empty set so $\mu(A \setminus A_k) \to 0$. Now suppose that $\mu(A) < \infty$. Then by Lemma 33.2, $\mu(A \setminus A_k) = \mu(A) - \mu(A_k) \to 0$ whence $\mu(A_k) \to \mu(A)$. Hence in particular if $\mu(X) < \infty$ we are done.

□

Lemma 33.4. *If a set function μ on an algebra $\mathcal{A}$ is countably additive, then it is $\delta-$ continuous.*

If it is $\delta-$continuous and is finitely additive, and if $\mu(X) < \infty$, then it is countably additive.

Proof. We showed in Lemma 33.1 that it is $\sigma-$continuous. By Lemma 33.3 this implies $\delta-$ continuity.

Conversely, we showed in Lemma 33.3 that if it is it is $\delta-$continuous and $\mu(X) < \infty$ it is $\sigma-$continuous. By Lemma 33.1, this implies countable additivity. □

In summary, we have proved (see Theorem 3.A of [Loè77]):

Proposition 33.5.
(i) A set function μ on an algebra $\mathcal{A}$ is countably additive iff it is finitely additive and $\sigma-$continuous.
(ii) If a set function μ on an algebra $\mathcal{A}$ is countably additive then it is finitely additive and $\delta-$continuous, with the converse holding if $\mu(X) < \infty$. .

Definition 33.2. Recall (see Definition 33.10) that power set $\mathfrak{P}(X)$ of a set X is the collection of all its subsets.

A function $\lambda : \mathfrak{P}(X) \to [0, +\infty]$ is an **outer measure** on X iff:
(i) $\lambda(\emptyset) = 0$;
(ii) $A \subseteq B \implies \lambda(A) \leq \lambda(B)$;
(iii) for any sequence of sets $\{A_i\}_{i=1}^{\infty}$, then $\lambda(\cup_{i=1}^{\infty} A_i) \leq \Sigma_{i=1}^{\infty} \lambda(A_i)$.

The second property is called **monotonicity** while the third is **countable subadditivity**.

The main examples of outer measures arise as follows.

See [DS57], Lemma 5. p 134.

-Halmos and Royden.

Proposition 33.6. *Let μ be a countably additive measure on an algebra $\mathcal{A}$ of subsets of the space X, as in Definition 33.1. Then μ^* defined on $\mathfrak{P}(X)$ by*

$$\mu^*(A) = \inf \Sigma_{i=1}^{\infty} \mu(A_i)$$

where the inf is taken over all countable covers $\{A_i\}_{i=1}^{\infty}$ of A by elements of $\mathcal{A}$, is an outer measure. Moreover, $\mu^* = \mu$ on $\mathcal{A}$.

Proof. Parts (i) and (ii) of the definition of outer measure are immediate. To prove (iii) we cover each A_i by sets $(A_i^j)_{j=1}^{\infty}$ each in $\mathcal{A}$ to within $\varepsilon/2^{-n}$, thus

$$\Sigma_{j=1}^{\infty} \mu(A_i^j) - \mu^*(A_i) < \varepsilon/2^{-n}$$

Then $A \subseteq \cup_{i,j=1}^{\infty} A_i^j$ and so $\mu^*(A) \leq \sum_{i,j=1}^{\infty} \mu(\mathcal{A}_i^j) \leq \sum_{i=1}^{\infty} \mu^*(A_i) + \varepsilon$.

Now let $A \in \mathcal{A}$. We are to show that $\mu^*(A) = \mu(A)$. We cover A by itself, so $\mu^*(A) \leq \mu(A)$. We claim it is equal.

Suppose not. Then there exist $A_i \in A$ with $A \subseteq \cup A_i$ such that $\sum \mu(A_i) < \mu(A)$. We "disjointify" them, by replacing them with the sequence of sets $\widetilde{A}_1 = A_1, \widetilde{A}_2 = A_2 \setminus A_1, \widetilde{A}_3 = A_3 \setminus (A_1 \cup A_2), \ldots$ and then define $\widehat{A}_i = A \cap \widetilde{A}_i$. Thus $\widehat{A}_i \in \mathcal{A}$, they are disjoint and $A = \cup \widehat{A}_i$. But by the definition of countable additivity on an algebra,

$$\mu(A) = \sum \mu(\widehat{A}_i) \leq \sum \mu(A_i) < \mu(A),$$

a contradiction. □

Note. It will be convenient to use (for this section only!) the algebra notation AB for $A \cap B$.

Definition 33.3. (Caratheodory) Given an outer measure λ on X, a set $A \subseteq X$ is **$\lambda-$measurable** iff for each $E \subseteq X$, $\lambda(E) = \lambda(EA) + \lambda(EA^c)$. We denote by $\mathcal{A}^*$ the collection of $\lambda-$measurable sets.

Given an outer measure λ and some $E \subseteq X$ with $0 < \lambda(E) < \infty$, we define the **outer measure relative to** E by the same formula as for measures, see §3.1:

$$\lambda_E(A) = \lambda(AE)/\lambda(E).$$

We note that with this notation, Caratheodory's definition can be restated as follows: A is $\lambda-$measurable iff $\lambda_E(E) = \lambda_E(A) + \lambda_E(A^c)$.

We denote by $\mathcal{A}_E$ the restricted $\sigma-$algebra, $\mathcal{A}_E = \{A \cap E : A \in \mathcal{A}\}$. Note that λ_E is an outer measure, on the space $(E, \mathcal{A}_E)$. Its own measurable sets are denoted $(\mathcal{A}_E)^*$. As we show below, in Theorem 33.12, this equals the restriction of the $\lambda-$measurable sets, i.e. $(\mathcal{A}_E)^* = (\mathcal{A}^*)_E$.

We next begin to explore the implications of this remarkable definition of Caratheodory. We first show that $\mathcal{A}^*$ is an algebra, then a $\sigma-$algebra, finally that the outer measure is is $\sigma-$additive hence a measure, when restricted to the Caratheodory measurable sets $\mathcal{A}^*$ (Proposition 33.11). Then we show there is a relative version of all this, for any (perhaps nonmeasurable) subset E.

Our first step is to note that Caratheodory definition has a partition version.

Definition 33.4. Let us say a partition $\mathcal{P} = \{P_i\}_{i=1}^n$ of X is $\lambda-$measurable iff for each $E \subseteq X$, $\lambda(E) = \sum_{i=1}^n \lambda(EP_i)$.

Lemma 33.7.
(i)Let $\mathcal{P}$ be a finite partition each of whose elements is $\lambda-$measurable. Then $\mathcal{P}$ is $\lambda-$measurable.
(ii)For $\{A_i\}_{i=1}^n$ disjoint and $\lambda-$ measurable, setting $A = \cup_{i=1}^n A_i$, then for any $E \subseteq X$,

$$\lambda(EA) = \sum_{i=1}^n \lambda(EA_i).$$

(iii) In particular, the restriction of the outer measure λ to $\mathcal{A}^*$ is finitely additive.

(Note that in *(iii)* we cannot yet say "is a finitely additive measure" as we have yet to prove that $\mathcal{A}^*$ is an algebra! We do that in Proposition 33.9.)

Proof. The proof of *(i)* is by induction on the number of elements n. For $n = 2$ this is true by the definition of $\lambda-$measurable set. Now suppose we know it for n, for any subset E.

Then given a partition $\mathcal{P} = \{P_i\}_{i=1}^{n+1}$, let $A = \cup_{i=1}^n P_i$, so $A^c = P_{n+1}$. But we have by the induction hypothesis, applied to the set $\widetilde{E} = AE$,

$$\lambda(EA) = \lambda(\widetilde{E}) = \sum_{i=1}^n \lambda(\widetilde{E}P_i).$$

Now since $A^c = P_{n+1}$ is $\lambda-$measurable,

$$\lambda(E) = \lambda(EA) + \lambda(EA^c) = \sum_{i=1}^n \lambda(\widetilde{E}P_i) + \lambda(EP_{n+1}) = \sum_{i=1}^{n+1} \lambda(EP_i).$$

For *(ii)*, we consider the partition $\{A, A_i\}_{i=1}^n$ and apply the proposition to the subset $\widetilde{E} = EA$. For *(iii)*, we take $E = X$.

□

We have:

Lemma 33.8. *Let $A, B \in \mathcal{A}^*$.*
Then the partition generated by A, B, that is, $\{AB, AB^c, A^cB, (A \cup B)^c\}$ is $\lambda-$measurable.

Proof. For $E \subseteq X$, since $A \in \mathcal{A}^*$,

$$\lambda(E) = \lambda(EA) + \lambda(EA^c).$$

Since $B \in \mathcal{A}^*$,

$$\lambda(EA) = \lambda(EAB) + \lambda(EAB^c).$$

and

$$\lambda(EA^c) = \lambda(EA^cB) + \lambda(EA^cB^c).$$

Therefore, since $A^cB^c = (A \cup B)^c$,

$$\lambda(E) = \lambda(EAB) + \lambda(EAB^c) + \lambda(EA^cB) + \lambda(E(A \cup B)^c). \tag{124}$$

□

This brings us to the main result we need. We follow the proof in §11 of [Hal76].

Proposition 33.9. *Given an outer measure λ on X, the collection $\mathcal{A}^*$ of $\lambda-$measurable sets is an algebra.*

Proof. It is enough to show that given $A, B \in \mathcal{A}^*$, then $A \cup B \in \mathcal{A}^*$.

Now if in equation (124) we replace E by $\widetilde{E} = E(A \cup B)$, we have

$$\lambda(E(A \cup B)) =$$
$$\lambda(E(A \cup B)AB) + \lambda(E(A \cup B)AB^c) + \lambda(E(A \cup B)A^cB) + \lambda(E(A \cup B)(A \cup B)^c).$$

We note that the last set is empty, while the first three are unchanged with $\widetilde{E}$ replaced by E. Thus,

$$\lambda(E(A \cup B)) = \lambda(EAB) + \lambda(EAB^c) + \lambda(EA^cB). \tag{125}$$

Now from (124) we had:

$$\lambda(E) = \lambda(EAB) + \lambda(EAB^c) + \lambda(EA^cB) + \lambda(E(A \cup B)^c),$$

so substituting from (125) for the first three terms on the right yields

$$\lambda(E) = \lambda(E(A \cup B)) + \lambda(E(A \cup B)^c),$$

and we are done. □

Here is a second proof, following §III.5.1 of [DS57].

Proof. This time to prove $\mathcal{A}^*$ is an algebra, we show that given $A, B \in \mathcal{A}^*$, then $AB \in \mathcal{A}^*$.

That is, we wish to show that for any $E \subseteq X$, $\lambda(E) = \lambda(E(AB)) + \lambda(E(AB)^c)$.

Now since $B \in \mathcal{A}^*$,

$$\lambda(E(AB)^c) = \lambda(E(AB)^cB) + \lambda(E(AB)^cB^c).$$

But since $(AB)^cB = BA^c$, we have that $E(AB)^cB = EBA^c$. Also, since $(AB)^cB^c = B^c$, we have that $E(AB)^cB^c = EB^c$. Thus

$$\lambda(E(AB)^c) = \lambda(EBA^c) + \lambda(EB^c). \tag{126}$$

Again since $B \in \mathcal{A}^*$,

$$\lambda(E) = \lambda(EB) + \lambda(EB^c)$$

and since $A \in \mathcal{A}^*$,

$$\lambda(EB) = \lambda(EBA) + \lambda(EBA^c).$$

It follows that

$$\lambda(E) = \lambda(EBA) + \lambda(EBA^c) + \lambda(EB^c).$$

The last two sets occur on the right-hand side of (126). Therefore, putting these together yields

$\lambda(E) = \lambda(E(AB)) + \lambda(E(AB)^c)$ as desired.

□

Next we show:

Proposition 33.10. *$\mathcal{A}^*$ is a $\sigma-$algebra.*

Proof. Since we know $\mathcal{A}^*$ is an algebra, it will be enough to show that given $\{A_i\}_{i=1}^{\infty}$ disjoint and $\lambda-$ measurable, then $A \equiv \cup_{i=1}^{\infty} A_i$ is $\lambda-$ measurable, i.e. that:

$$\lambda(E) = \lambda(EA) + \lambda(EA^c).$$

We write $A^{(m)} = \cup_{i=0}^{m} A_i$. Then $\{A_1, \ldots, A_m, (A^{(m)})^c\}$ is a partition, so by part (i) of Lemma 33.7, for any m we have:

$$\lambda(E) = \lambda(\cup_{i=0}^{m} EA_i)) + \lambda(E(A^{(m)})^c).$$

Then:

$$\lambda(E) = \lambda(EA^{(m)}) + \lambda(E(A^{(m)})^c) \geq \sum_{i=1}^{m} \lambda(EA_i) + \lambda(EA^c), \text{as } A^c \subseteq (A^{(m)})^c.$$

Since this holds for every m,

$$\lambda(E) \geq \sum_{i=1}^{\infty} \lambda(EA_i) + \lambda(EA^c) \geq \lambda(EA) + \lambda(EA^c) \geq \lambda(E),$$

using subadditivity in the last two inequalities, and we are done. □

We now extend Definition 33.4 to countably infinite partitions.

Definition 33.5. A partition $\mathcal{P} = \{P_i\}_{i=1}^{\infty}$ of X is $\lambda-$measurable iff for each $E \subseteq X$, $\lambda(E) = \sum_{i=1}^{\infty} \lambda(EP_i)$.

Proposition 33.11. *Lemma 33.7 also holds for countably infinite partitions. That is:*
(i)Let $\mathcal{P} = \{P_i\}_{i=1}^{\infty}$ be a partition each of whose elements is $\lambda-$measurable. Then $\mathcal{P}$ is $\lambda-$measurable.
(ii)For $\{A_i\}_{i=1}^{\infty}$ disjoint and $\lambda-$ measurable, setting $A = \cup_{i=1}^{\infty} A_i$, then for any $E \subseteq X$,

$$\lambda(EA) = \sum_{i=1}^{\infty} \lambda(EA_i).$$

(iii) In particular, the restriction of the outer measure λ to $\mathcal{A}^*$ is countably additive.

Proof. Defining $P_n^{\infty} = \cup_{i=n}^{\infty} P_i$, then $\{P_n^{\infty}, P_i\}_{i=1}^{n-1}$ is a finite partition. By Lemma 33.10, P_n^{∞} is $\lambda-$measurable.

Hence by the above,

$$\lambda(E) = \lambda(E(\cup_{i=1}^{\infty} P_i)) = \sum_{i=1}^{n-1} \lambda(EP_i) + \lambda(EP_n^{\infty})$$

whence $\lambda(E) \geq \sum_{i=1}^{n-1} \lambda(EP_i)$ for all n, hence $\lambda(E) \geq \sum_{i=1}^{\infty} \lambda(EP_i)$. But from subadditivity, $\lambda(E) \leq \sum_{i=1}^{\infty} \lambda(EP_i)$, completing the proof.

To show (ii), we consider the partition $\{(A_i)_{i=1}^{\infty}, (A^{\infty})^c\}$, and apply (i) to the subset $\widetilde{E} = E \cap A^{\infty}$. For (iii), apply (ii) with $E = X$. □

We summarize:

Theorem 33.12.
(i)(Caratheodory extension theorem) *If λ is an outer measure on X then the collection $\mathcal{A}^*$ of $\lambda-$measurable sets form a $\sigma-$algebra on which λ is a countably additive measure.*
(ii) We have, moreover, a relative version of this, for any (perhaps nonmeasurable) subspace E:

Let $E \subseteq X$ with outer measure $0 < \lambda(E) < \infty$. Then $(\mathcal{A}_E)^* = (\mathcal{A}^*)_E$. That is, the Caratheodory measurable sets for the relative outer measure λ_E of Definition 33.3 are the restrictions of the $\lambda-$ measurable sets.

Furthermore, λ_E is a countably additive measure on the relative $\sigma-$algebra $(\mathcal{A}^*)_E$.

Proof. Part (i) was proved in Proposition 33.10 and (iii) of Proposition 33.11.

For part (ii), the first statement follows from the definitions, and for the second, we divide the equality in (ii) of Proposition 33.11 by $\lambda(E)$. □

Example 31. The relative version in part (ii) has practical consequences.

For an example, consider the space of all functions $\mathbb{R}^{\mathbb{R}}$ with the product topology, and the subset of continuous functions $\mathcal{C}(\mathbb{R})$ with the topology of uniform convergence on compact subsets of $\mathbb{R}$. Now to define the measure ν for Brownian motion one can proceed as follows. The measure is initially defined on the algebra on on $\mathbb{R}^{\mathbb{R}}$ generated by finite cylinder sets (from the basic properties we want for Brownian motion: the Gaussian distribution at each time, the scaling property, and independent increments) and this is extended via Alexandroff's theorem below (Theorem 33.15) to the $\sigma-$algebra generated by that. We then restrict to the continuous functions, $\mathcal{C}(\mathbb{R})$ which is a nonmeasurable subset of $\mathbb{R}^{\mathbb{R}}$ with this $\sigma-$algebra, of full outer measure. The result is $(\mathcal{C}(\mathbb{R}), \mathcal{B}, \nu)$, which is a Polish space (Def. 5.2) hence a Lebesgue space.

Definition 33.6. A finitely additive measure μ on an algebra $\mathcal{A}$ of subsets of X is **$\sigma-$finite** iff there exist $A_i \in \mathcal{A}$ such that $\mu(A_i) < \infty$ and $X = \cup_{i=1}^{\infty} A_i$.

Theorem 33.13. *(Hahn extension,* [DS57]*) A countably additive measure μ on an algebra $\mathcal{A}$ of subsets of X has a countably additive extension μ^* to the $\sigma-$ algebra $\widehat{\mathcal{A}}$ generated by $\mathcal{A}$, the Caratheodory extension defined on the (a priori larger) $\sigma-$ algebra $\mathcal{A}^*$ of μ-measurable sets. This is the largest extension to $\widehat{\mathcal{A}}$. If μ is $\sigma-$finite then this extension to $\widehat{\mathcal{A}}$ is unique.*

Proof. Let μ^* denote the outer measure on $\mathfrak{P}(X)$ of Definition 33.3, and let $\mathcal{A}^*$ denote the $\sigma-$algebra of Caratheodory μ^*-measurable sets of Definition 33.3. Then by Caratheodory's extension theorem Theorem 33.12, this is countably additive on $\mathcal{A}^*$, which contains $\widehat{\mathcal{A}}$.

Suppose that $\widetilde{\mu}$ is another extension to $\widehat{\mathcal{A}}$. By the definition of μ^* in Proposition 33.6, $\mu^*(E) = \inf \sum \mu(A_i)$ where the infimum is taken over all countable covers of E by sets in $\mathcal{A}$. By monotonicity, $\widetilde{\mu}(\cup A_i) \leq \sum \mu(A_i)$. Therefore $\widetilde{\mu}(E) \leq \mu^*(E)$ and μ^* is the largest.

Next we assume μ is $\sigma-$finite: thus $X = \cup_{i=1}^{\infty} A_i$ with $A_i \in \mathcal{A}$ and $\mu(A_i) < \infty$. Let $\widetilde{A}_n = \cup_{i=1}^{n} A_i$. Thus $\widetilde{A}_n$ increase to X and each has finite measure.

Let $E \in \mathcal{A}^*$. Since $E \cap \widetilde{A}_n$ increase to E, by $\sigma-$continuity (Proposition33.5) $\widetilde{\mu}(E \cap \widetilde{A}_n)$ increases to $\widetilde{\mu}(E)$, and similarly for μ^*.

Hence if we can show that $\mu^* = \widetilde{\mu}$ for $\mu(E \cap \widetilde{A}_n)$ we will be done.

Thus supposing that $E \in \mathcal{A}^*$ is a subset of some set $F \in \mathcal{A}$ of finite measure, we are to show that $\widetilde{\mu}(E) = \mu^*(E)$.

We have seen that $\widetilde{\mu}(E) \leq \mu^*(E)$. We set $\widetilde{E} = A \setminus E$, and for the same reason have $\widetilde{\mu}(\widetilde{E}) \leq \mu^*(\widetilde{E})$.

Now $\widetilde{\mu}(\widetilde{E}) = \widetilde{\mu}(A \setminus E) = \widetilde{\mu}(A) - \widetilde{\mu}(E)$ since all sets here have finite measure (Lemma 33.2), and similarly for μ^*. Thus $\widetilde{\mu}(A) - \widetilde{\mu}(E) = \widetilde{\mu}(\widetilde{E}) \leq \mu^*(\widetilde{E}) = \mu^*(A) - \mu^*(E)$ whence $\mu^*(E) \leq \widetilde{\mu}(E)$ so we are done.

□

Remark 33.1. In practice one often tries to avoid addressing non-$\sigma-$finite spaces as they can exhibit exceptional behavior; for example this nonuniqueness can occur. Non-$\sigma-$finite spaces can be constructed by simply taking uncountably many disjoint copies of a finte measure space, such as a continuum of point masses (i.e. counting measure on the unit interval), or a continuum of line segments (such as a measure on $I \times I$ where the first interval has Lebesgue measure and the second counting measure, and we take the $\sigma-$ algebra of rectangles). For this last example, the Hahn extension is not unique, since for the Caratheodory extension the measure of the diagonal is ∞, but for another extension this is zero. Indeed, a rectangle R has the form $[a, b] \times F$ where F is a finite subset of J, and$m \times \mu(R) = (b - a) \cdot \#F$. We cover the diagonal D by a cover $\cup_{i \in S} R_i$. Note that the index set S must be uncountable since each F is discrete. But any uncountable sum $\sum_{i \in S} r_i \equiv \sup_{\text{finite subsets G of } S} \sum_{i \in G} r_i$ of positive numbers is infinite since if there are only finitely many $r_i > 1/n$ for each n, then there are only countably many points in S. Thus $\mu^*(D) = \infty$. For a second extension, (Wikipedia, *Caratheodory Extension Theorem*), we define the measure$\tilde{\mu}(A)$ of a subset to be the sum of the measures of its horizontal sections. Then $\widetilde{\mu}(D) = 0$. ???

For our next steps we need:

Definition 33.7. A finitely additive measure μ on a topological space X with topology $\mathcal{T}$ and an algebra of sets $\mathcal{A}_0$ is **regular** on the algebra iff for $A \in \mathcal{A}_0$, for every $\varepsilon > 0$ there exists a closed subset F and an open subset U such that $F \subseteq A \subseteq U$

with $\mu(U \setminus F) < \varepsilon$. We make the similar definition for μ countably additive, defined on a $\sigma-$algebra $\mathcal{A}$.

The next statement is easy to verify using Venn diagrams, but as we have not found it in the literature, here is a rigorous proof:

Lemma 33.14.

$$(A \setminus B) \cup (C \setminus D) \supset (A \cup C) \setminus (B \cup D).$$

Proof.

$$(A \setminus B) \cup (C \setminus D) = (A \cap B^c) \cup (C \cap D^c) =$$
$$((A \cap B^c) \cup C) \cap ((A \cap B^c) \cup D^c) =$$
$$((A \cup C) \cap (B^c \cup C)) \cap ((A \cup D^c) \cap (B^c \cup D^c)) =$$
$$(A \cup C) \cap (B^c \cup C) \cap (A \cup D^c) \cap (B^c \cup D^c)$$

On the other hand,

$$(A \cup C) \setminus (B \cup D) = (A \cup C) \cap (B \cup D)^c =$$
$$(A \cup C) \cap (B^c \cap D^c) =$$
$$(A \cap (B^c \cap D^c)) \cup (C \cap (B^c \cap D^c)) =$$
$$((A \cap B^c \cap D^c) \cup C) \cap ((A \cap B^c \cap D^c) \cup B^c) \cap ((A \cap B^c \cap D^c) \cup D^c) =$$
$$(A \cup C) \cap (B^c \cup C) \cap (D^c \cup C) \cap$$
$$(A \cup B^c) \cap (B^c \cup B^c) \cap (D^c \cup B^c) \cap$$
$$(A \cup D^c) \cap (B^c \cup D^c) \cap (D^c \cup D^c) =$$
$$(A \cup C) \cap (B^c \cup C) \cap (A \cup D^c) \cap (B^c \cup D^c) \cap$$
$$(D^c \cup C) \cap (A \cup B^c) \cap B^c \cap (D^c \cup B^c) \cap D^c =$$
$$((A \setminus B) \cup (C \setminus D)) \cap$$
$$((D^c \cup C) \cap (A \cup B^c) \cap B^c \cap (D^c \cup B^c) \cap D^c)$$

hence the first set is larger.

□

The next theorem helps explain the essential difference between finite and countable additivity is whether or not the space is compact (for regular measures, which condition serves to link the measure to the topology).

Theorem 33.15. (Alexandroff; Theorem III.5.13, .14 of [DS57]) *If $(X, \mathcal{T})$ is a compact topological space and μ is a regular finitely additive measure on an algebra $\mathcal{A}_0$, then μ is countably additive on the algebra, and it has a unique countably additive extension to $\mathcal{A}$, the $\sigma-$algebra generated by $\mathcal{A}_0$, where it is still regular.*

Proof. Let $(A_i)_{i \geq 1}$ be disjoint elements of $\mathcal{A}_0$, and write $A = \cup_{i=1}^{\infty} A_i$. We suppose $A \in \mathcal{A}_0$ as well. We shall show that $\mu(A) \geq \sum_{i=1}^{\infty} \mu(A_i)$ (this is **superadditivity**) and the reverse. The first is easy and is true just from additivity and the assumption that $A \in \mathcal{A}_0$ so $\mu(A)$ is defined: $\mu(A) = \mu(\cup_{i=0}^{\infty} A_i) \geq \mu(\cup_{i=0}^{m} A_i) = \sum_{i=0}^{m} \mu(A_i)$, hence $\mu(A) \geq \sum_{i=1}^{\infty} \mu(A_i)$.

The second inequality (subadditivity) is not automatic, as it uses compactness plus regularity. By regularity (using the fact that $A \in \mathcal{A}_0$) there exists a closed subset F of A with $\mu(A \setminus F) < \varepsilon$; since X is compact so is F. Also by regularity, there exists for each i an open set $U_i \supseteq A_i$ with $\mu(U_i \setminus A_i) < \varepsilon/2^i$.

Now since $\mu(A_i) \geq \mu(U_i) - \varepsilon/2^i$, we have

$$\sum_{i=1}^{\infty} \mu(A_i) \geq \sum_{i=1}^{\infty} \mu(U_i) - \varepsilon.$$

Since F is compact, there exits $N \geq 1$ such that $\cup_{i=0}^{N} U_i \supseteq F$. Therefore by finite subadditivity,

$$\sum_{i=1}^{\infty} \mu(U_i) - \varepsilon \geq \sum_{i=0}^{N} \mu(U_i) - \varepsilon \geq \mu(\cup_{i=0}^{N} U_i) - \varepsilon \geq \mu(F) - \varepsilon \geq \mu(A) - 2\varepsilon.$$

Combining these two, we are done with the proof of countable additivity on $\mathcal{A}_0$.

This proof of countable additivity on the algebra (which was divined from the key topological hypotheses of compactness plus regularity) is all that is needed to prove Alexandroff's theorem, as that allows us to apply Caratheodory's ideas, via the Hahn extension (Theorem 33.13). To recall the argument, writing μ^* for the outer measure defined from μ as in Proposition 33.6, then by Proposition 33.10 the collection $\mathcal{A}_0^*$ of of μ^*-Caratheodory measurable sets is a $\sigma-$algebra, hence contains $\mathcal{A}$, and by (i) of Theorem 33.12 the outer measure μ^* is countably additive when restricted to $\mathcal{A}_0^*$. We write μ for the restriction to $\mathcal{A}$.

Lastly we show that μ is still regular on $\mathcal{A}$. Let us consider the collection $\widehat{\mathcal{A}}$ of sets in $\mathcal{A}$ satisfying the regularity condition; this contains $\mathcal{A}_0$, so if we show this is a $\sigma-$algebra, then it equals $\mathcal{A}$. Given $(A_i)_{i\geq 1}$ a sequence in $\mathcal{A}$, it will be enough to show the regularity condition for $A = \cup_{i=1}^{\infty} A_i$. We have F_i, U_i compact and open with $F_i \subseteq A_i \subseteq U_i$ and $\mu(U_i \setminus F_i) < \varepsilon/2^i$. Then by the Lemma (which remains valid for arbitrary unions)

$$\mathcal{U} \setminus F \equiv \cup(\mathcal{U}_i) \setminus \cup(\mathcal{F}_i) \supset \cup(\mathcal{U}_i \setminus \mathcal{F}_i)$$

whence

$$\lambda(\mathcal{U} \setminus F) \leq \sum \varepsilon/2^i = \varepsilon.$$

Now a countable union of closed sets may not be closed, but a finite union is, so we can replace F by $\cup_{i=1}^{n} F_i$ to finish the proof.

□

Remark 33.2. For an important example $\mathbb{R}^{\mathbb{R}}$, this is *not* yet the extension to the Borel $\sigma-$algebra for the product topology, which is larger. For that we need a theorem of Nelson. See [Nel59] and see Theorem 33.28 below.

Example 32. (Density, part (i)) We mention an interesting case where for a number of reasons the main ideas in this section cannot work. So it will be good to check each step against this "counterexample" to see what goes wrong. We return to this several times below.

Recall that the **density** of a Lebesgue measurable subset $A \subseteq \mathbb{R}$ is:

$$\lambda(A) = \lim_{T\to\infty} \frac{1}{T} m(A \cap [0,T])$$

when the limit exists. Writing $\mathcal{A}$ for the collection of sets where this is defined, this is an algebra, and λ is a finitely additive, translation-invariant probability measure on $\mathcal{A}$.

For every compact set A, $\lambda(A) = 0$. Take $A = [0,1)$ and $A_n = n + A$.Then Thus $\mathbb{R} = \cup_{n\in\mathbb{Z}} A_n$ whence $\sum_{n\in\mathbb{Z}} \lambda((A_n) = 0 < 1 = \lambda(\mathbb{R})$and λ is not countably additive.

We might nevertheless try to define an outer measure, by $\overline{\lambda}(A) = \limsup_{T\to\infty} \frac{1}{T}\int_0^T \chi_A(x)\mathrm{d}x$. Since $\sup(f+g) \leq \sup f + \sup g$, this is subadditive. However, again $\overline{\lambda}$ cannot be countably subadditive, for the above reason.

Next, we can try to extend the notion of density further via Caratheodory's definition of outer measure.

However, the hypothesis of Proposition 33.6 is not satisfied, since as noted above, although a finitely-additive measure on the algebra $\mathcal{A}$, λ is *not* countably additive on $\mathcal{A}$. Indeed, we can see directly that the definition there fails, as for any n, $n + [0,1) \in \mathcal{A}$, and then for any set $A \subseteq \mathbb{R}$, $\lambda^*(A) = \inf_{\text{covers}} \Sigma_{i=1}^{\infty}\lambda(A_i) = \Sigma_{i=1}^{\infty}\lambda(i + [0,1)) = 0$ is an outer measure (trivially), yet $\lambda(\mathbb{R}) = 1 \neq \lambda^*(\mathbb{R}) = 0$, and the last line of the statement in the Proposition does not hold.

Example 33. (Density, part (*iii*)) We look once more at Example 32. The key hypothesis in Alexandroff's extension theorem is compactness, which allows us to prove countable additivity on the algebra and hence apply Caratheodory's extension theorem. An interesting counterexample is given by the construction of a translation-invariant mean on $\mathbb{R}$: there all of these fail: the space is noncompact, countable additivity on the algebra fails, and the finitely additive measure on the algebra does not produce an outer measure.

Thus, via our definition of density, we have a finitely additive measure λ on an algebra of sets. So the hypothesis of Alexandroff's theorem would guarantee a countably additive extension, if not for one crucial problem: the noncompactness of the space $\mathbb{R}$.

Of course, this lack of compactness is not a problem for the definition of Lebesgue measure, as we define it first on the compact interval $[0,1]$ and then extend via translation.

In summary, we essentially have a choice for a translation-invariant measure on $\mathbb{R}$: either we allow an infinite measure (Lebesgue measure, or some constant multiple of it) or, if we insist on having a probability measure, then the best we can do is finite additivity.

33.2. **Measures as linear functionals: the Riesz representation theorem.** This central result is due to different authors in a variety of settings, starting with Riesz and moving through Markov, Kolmogorov, and Kakutani. See [DS57] pp. 373, 380 ff. for some history.

Definition 33.8. ([Kel75]; [DS57] §I.5.1) Consider the following possible properties of a topological space $(X, \mathcal{T})$:
(a) Singletons are closed;

(b) Any two distinct points are separated by disjoint neighborhoods;
(c) Every closed set is separated from any point not in the set by separated by disjoint neighborhoods;
(d) Any two disjoint closed sets are separated by disjoint neighborhoods.

The space is called T_1 iff it satisfies (a). It is **Hausdorff** iff it satisfies (a), (b). It is **regular** iff it satisfies (a), (c), and is **normal** iff it satisfies (a), (d).

We note that

Proposition 33.16. *(see* [DS57] *§I.5.9) A compact Hausdorff space is normal; a metric space is normal.*

Definition 33.9. An **additive set function** defined on a collection $\mathcal{S}$ of subsets of a space X, containing $\emptyset$, is a function $\mu : \mathcal{S} \to Y$ where Y is generally either $\mathbb{R}$, $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$, $\mathbb{C}$ but could also be a Banach space, satisfying:
(i) $\mu(\emptyset) = 0$
(ii) For A, B disjoint with union in $\mathcal{S}$, $\mu(A \cup B) = \mu(A) + \mu(B)$. It is **positive** iff it assumes values in $\mathbb{R}^+ = [0, +\infty]$.

If $\mathcal{S}$ is an algebra of sets then a positive additive set function μ is called a **finitely additive measure**. If Y is $\overline{\mathbb{R}}$ or a Banach space it is a **finitely additive signed** or **vector-valued measure**, respectively.

It is a **measure** if in addition, $\mathcal{A}$ is a $\sigma-$algebra and μ is **countably** additive, i.e. for $\{A_i\}_{i=1}^{\infty}$ disjoint, then $\mu(\cup_{i=1}^{\infty} A_i) = \Sigma_{i=1}^{\infty} \mu(A_i)$.

We write rba = rba(X) for the vector space of regular bounded finitely additive signed measures on X. We write rca(X) for the regular bounded countably additive signed measures on X.

We take the sup norm on $\mathcal{C}$, and define a norm on rba(X), rca(X) to be

Theorem 33.17. *(Riesz Representation Theorem: finitely additive case* [DS57]*, Theorem IV.6.2; countably additive case Theorem IV.6.3).*
(i) Let $(X, \mathcal{T})$ be a normal topological space and $\mathcal{C}$ the space of bounded (real or complex-valued) functions on X. Then the dual space $\mathcal{C}^*$ is isometrically isomorphic to rba(X).
(ii) If X is a compact Hausdorff space, then $\mathcal{C}^*$ is isometrically isomorphic to rca(X).

Proof. □

33.3. **The Stone-Weierstrass Theorem.**

33.4. **The existence of invariant measures.**

33.5. **Generic points and the Krylov-Bougliobov-Fomin theorem.**

33.6. **Building the Borel $\sigma-$algebra.** We recall:

Definition 33.10. Given a set X, we denote by $\mathfrak{P}(X)$ the collection of all subsets of X. This is the **power set** of X (so called– see [Hal74]– because it can be identified with 2^X, the set of all functions from X to $2 \equiv \{0, 1\}$ via the map $A \mapsto \chi_A$; furthermore if the cardinality of X is α, then the cardinality of $\mathfrak{P}(X)$ is 2^α.)

This is clearly a $\sigma-$algebra, so each space X has a largest and smallest $\sigma-$algebra, the power set and the **trivial $\sigma-$algebra** $\{\emptyset, X\}$.

We note that:

Lemma 33.18. *Given a set X, an (arbitrary) intersection of $\sigma-$algebras in X is a $\sigma-$algebra.* □

Definition 33.11. Given a set X and a collection of subsets $S \subseteq \mathfrak{P}(X)$, we write $\sigma(S)$ for the smallest $\sigma-$algebra containing S.

Lemma 33.19. *This makes sense. That is, there exists such a $\sigma-$algebra, smallest in the sense of containment.*

Proof. We write C for the collection of all $\sigma-$algebras which contain S, and define $\sigma(S) \equiv \cap C$. If $\mathcal{A}$ is such a $\sigma-$algebra, then $A \in C$, whence $\sigma(S) \subseteq \mathcal{A}$, so it is the smallest.

□

Remark 33.3. It is important here to note that C has at least one element: $\mathfrak{P}(X)$. This is because by definition, see [Hal50]),

$$\cap C = \{A : \forall \mathcal{A} \in C, A \in \mathcal{A}\}$$

. But if C is empty, this becomes

$$\cap C = \{A : (\mathcal{A} \in C) \implies (A \in \mathcal{A})\}$$

that is,

$$\{A : \text{it is true that } (\mathcal{A} \in C) \implies A \in \mathcal{A})\}.$$

But the statement "if $\mathcal{A} \in C$, then $A \in \mathcal{A}$" has a false premise, hence the implication is always true. That is, every set A satisfies $A \in \cap\emptyset$, so this is the set of all sets (the Universe), which is not a set; this is related to Russell's paradox. See [Hal74] p. 18.

Definition 33.12. Given a set X and a collection A of subsets of X, we denote by A_σ the collection of all countable unions of sets in A and by A_δ denote the collection of countable intersections. If we interate this procedure, we indicate it by $A_{\delta\sigma} \equiv (A_\delta)_\sigma$ and so on.

In particular, given a topological space $(X, \mathcal{T})$, we write G for $\mathcal{T}$ the collection of open sets, and F for the closed sets. So then G_δ denotes the collection of sets which are countable intersections of closed sets, and F_σ the collection countable unions of closed sets, and so on.

See [Doo12] p. 16, [Nel59], and Theorem 33.28.

By definition $\mathcal{B}$, the collection of **Borel sets**, is the $\sigma-$algebra generated by $\mathcal{T}$.

By the above, this exists.

We recall that **transfinite induction** generalizes the usual induction procedure to the first uncountable ordinal. See [Hal74] regarding ordinal numbers. A good example of how this works is given in the next proof.

Proposition 33.20. *If X is a separable metric space with topology $\mathcal{T}$, then the Borel $\sigma-$algebra can be built constructively using transfinite induction up to the first uncountable ordinal.*

Proof. We explain the statement fully in this proof. Define $G_0 = G = \mathcal{T}, G_1 = G_\delta, G_2 = G_{\delta\sigma}, G_3 = G_{\delta\sigma\delta}$ and so on for $n \in \omega = \mathbb{N} = \{0, 1, 2, 3 \dots\}$. Recall that $n \in \omega \iff n < \omega$, the first infinite ordinal. Define $G_\omega = \cup_{n<\omega} G_n$, next set $G_{\omega+1} = (G_\omega)_\delta$ and so on. At each limit ordinal we take the union of those before it, just as we did at the first limit ordinal ω. We continue in this manner up to the first uncountable ordinal ω_1. We claim that G_{ω_1} is a $\sigma-$algebra.

First, note that since X is a separable metric space, each open set is a countable union of closed balls, and each closed set is therefore a G_δ. Thus, G_{ω_1} already includes all of $F_{\delta\sigma\delta}$ and so on, so by DeMorgan's laws includes all complements.

Next, we claim a countable union of sets in G_{ω_1} is in G_{ω_1}. Write this as $\cup_I A_i$ for some index set of ordinal numbers, each countable. But then the union of these ordinal numbers $\alpha \equiv \cup I = \sup I$ is itself an ordinal number $\alpha < \omega_1$ hence countable. Thus $\cup_I A_i \in G_\alpha \subseteq G_{\omega_1}$.

Clearly any $\sigma-$algebra containing the topology must contain G_{ω_1}. Therefore this is the smallest such $\sigma-$algebra, so by definition, it is the Borel $\sigma-$algebra $\mathcal{B}$. □

33.7. **The Baire $\sigma-$algebra.** $\mathcal{B}\dashv$, the collection of **Baire sets**, is the $\sigma-$algebra generated by the collection of all dense G_δ sets.

33.8. **Joint distributions and continuous-time stochastic processes.** A major application of all the above will be to a study of continuous-time stochastic processes. For this we first define the joint distributions of random variables.

In §3.1 we have discussed independent random variables, and their relationship to product measure. This generalizes as follows.

Definition 33.13. Given a probability space $(\Omega, \mathcal{A}, P)$, and random variables $X, Y : \Omega \to \mathbb{R}$, consider the map $\Phi : \Omega \to \mathbb{R}^2$ defined by $\Phi(\omega) = (X(\omega), Y(\omega))$. We call the measure $\Phi_*(P)$ on $\mathbb{R}^2$ the **joint distribution** of the ordered pair (X, Y).

Lemma 33.21. *The joint distribution is determined by the values*

$$\{P[(X \in A) \wedge (Y \in B)]\}_{A,B \in \mathcal{B}(\mathbb{R})}.$$

Proof. $P[(X \in A) \wedge (Y \in B)] = (\Phi_*(P))(A \times B)$, and this collection of rectangles generates the product $\sigma-$algebra of $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$, and so those values determine the measure. □

Definition 33.14. Now we consider random variables $X_1, \dots X_n$. We give $\mathbb{R}^n$ the product topology $\mathcal{T}_n$. A base for $\mathcal{T}_n$ is the collection of **open cylinder sets**, that is, sets of the form $\mathcal{U}_1 \times \cdots \times \mathcal{U}_n$ with $\mathcal{U}_i$ open. We define $\mathcal{F}_n$ to be the algebra of Borel subsets. This is generated also generated by the open cylinder sets, and just as above for $n = 2$, any measure on $\mathbb{R}^n$ is determined by its values on these sets, as they generate the algebra.

So given random variables $X_1, \dots X_n$ on $(\Omega, \mathcal{A}, P)$, we define a map $\Phi : \Omega \to \mathbb{R}^n$ by $\Phi(\omega) = (X_1(\omega), \dots X_n(\omega))$. Extending the previous definition, the **joint distribution of** $(X_1, \dots X_n)$ is the measure $\Phi_*(P)$ on $\mathbb{R}^n$.

Lemma 33.22. *Given measurable spaces $(\Omega_i, \mathcal{A}_i)$ for $i = 1, \dots n$, their joint distribution is determined by the values on the cylinder sets, i.e. by $\Phi_*(P)(\mathcal{U}_1 \times \cdots \times \mathcal{U}_n) = P(X_i \in \mathcal{U}_1, \dots X_n \in \mathcal{U}_n)$, for each choice of $\mathcal{U}_i$ open in $\mathbb{R}$.*

Definition 33.15. Now suppose we are given any index set T, possibly infinite, and an indexed collection of random variables $(X_t)_{t\in T}$ with $(X_t : \Omega \to \mathbb{R})$.

We recall first the definitions of product topology and product $\sigma-$algebra. The product topology $\mathcal{T}_p$ of $\mathbb{R}^T = \Pi_{t\in\mathbb{T}}\mathbb{R}$ is generated by the open cylinder sets: sets of the form $(\otimes\mathcal{U}_t)_{t\in F}$, for some finite $F \subseteq T$.

We note that the space $\mathbb{R}^F$ is a factor of $\mathbb{R}^T$ via the projection map $\pi_F(x) = (x(t))_{t\in F}$. Thus the product topology $\mathcal{T}_F$ on $\mathbb{R}^F$ is the quotient topology.

The Borel algebra $\mathcal{F}_F$ for $\mathbb{R}^F$ is generated by the topology $\mathcal{T}_F$ and hence again by the collection of open cylinder sets. A cylinder set for this algebra is more general, by definition a set of the form $(\otimes A_t)_{t\in F}$ where $A \in \mathcal{B}(\mathbb{R})$.

We next consider the whole space $\mathbb{R}^T$. We define a **finite cylinder set** to be a subset of the form $(\otimes A_t)_{t\in F}$ for some finite $F \subseteq \mathbb{R}$. The algebra generated by all the finite cylinder sets defines the **finite cylinder algebra** $\mathcal{F}$ of $\mathbb{R}^T$. Note the relationship with the projection maps: for each F we have the projection $\pi_F : \mathbb{R}^T \to \mathbb{R}^F$; this takes $\mathcal{F}_F$ to an isomorphic algebra $\widehat{\mathcal{F}}_F$ of $\mathbb{R}^T$, via $A \mapsto \pi^{-1}(A)$. That is, it is generated by the collection $\cup_F\widehat{\mathcal{F}}_F$.

The **finite cylinder $\sigma-$algebra** $\mathcal{F}_0$ of $\mathbb{R}^T$ is the $\sigma-$algebra generated by $\mathcal{F}$.

Returning to the random variables, we define the map $\Phi_F : \Omega \to \mathbb{R}^F$ where $\Phi_F(\omega) = (X(t))_{t\in F}$. The probability measure $\mu_F = (\Phi_F)_*(P)$ on $\mathbb{R}^F$ with the algebra just defined is the $F-$ **joint distribution** of the random variables $(X_t)_{t\in T}$.

The importance of this $\sigma-$algebra comes from the Kolmogorov extension theorem, see Theorem 33.24. Supposing that we are given joint distributions for each finite subset $F \subseteq \mathbb{R}$, a priori defined on different probability spaces Ω_F, this tells us that if they satisfy the natural consistency conditions, then they fit together in such a way that they can be defined simultaneously on a single space Ω, giving a measure on the finite cylinder $\sigma-$algebra $\mathcal{F}_0$ of $\dot{\mathbb{R}}^{\mathbb{R}}$, where $\dot{\mathbb{R}}$ is the one-point compactification.

Now naturally we should like to make use of the product topology on $\dot{\mathbb{R}}^{\mathbb{R}}$, which by Tychonoff's product theorem is a nice space (a compact Hausdorff space) and hence of its Borel $\sigma-$ algebra $\mathcal{B}$.

However technical difficulties arise here because the Borel $\sigma-$algebra is much larger than the finite cylinder $\sigma-$algebra $\mathcal{F}_0$. This is in marked contrast to the case of countable time, where for $T = \mathbb{Z}$ these two $\sigma-$algebras are equal. Now it is true that the product $\sigma-$ algebra for the infinite product space $\mathbb{R}^T$ and the product topology are both generated by the finite open cylinder sets. However to form open sets we can take arbitrary unions of these open cylinders, whereas only countable unions are allowed in forming $\mathcal{F}_0$. In the case of $\mathbb{R}^{\mathbb{Z}}$ the space is metrizable, so open sets are a countable union of open cylinders; for $\mathbb{R}^T$ this is not true, as the space is *not* metrizable.

We mention that often in practice one can recover a nicer space by restricting to a subspace, for example the continuous paths $\mathcal{C} \subseteq \mathbb{R}^{\mathbb{R}}$, , as we do for Brownian motion.

Corollary 33.23. *Each set in $\mathcal{F}_0$ depends on a countable subset of the index set $\mathbb{R}$.*

33.9. **The Kolmogorov and Kakutani-Nelson embeddings.**

Note that the joint distribution μ_F defines a measure on the algebra $\mathcal{F}_F$. The collection of such measures satisfies two obvious consistency properties:

(i) μ_F only depends on the subset F, not on an ordering of F.
(ii) If $F' \subseteq F$, then μ_F projects onto $\mu_{F'}$.

So to have an extension we certainly need that the joint distributions satisfy these two simple consistency conditions. Kolmogorov's famous result shows this is enough:

Theorem 33.24. (Kolmogorov embedding theorem) *Let T be any index set. For each $t \in T$ let $(\Omega_t, \mathcal{A}_t, P_t)$ be a probability space and $X_t : \Omega_t \to \mathbb{R}$ measurable. Assume that in addition to the random variables X_t we have specified all the finite joint distributions. That is, given $(t_1, t_2, \dots t_n)$ and measurable sets $E_i \subseteq \mathbb{R}$, we are given the probability that $X_{t_i} \in E_i$ for $i = 1, \dots, n$. Asumme that these joint distributions satisfy the two consistency conditions:*
(i) the joint probabilties only depend on the collection $\{t_1, t_2 \dots, t_n\}$ and not on their order;
(ii) they are consistent in that if one more index t_{n+1} is added, the probabilities add. That is, the probability that $\big((X_{t_i} \in E_i \text{ for } i = 1, \dots, n) \text{ and } (X_{t_{n+1}} \in \mathbb{R})\big)$ equals the probability that $(X_{t_i} \in E_i \text{ for } i = 1, \dots, n)$.

Then there exists a probability space $(\Omega, \mathcal{A}, P)$ which simultaneously represents all $(X_t)_{t\in T}$. That is, there exist random variables $\widehat{X}_t : \Omega \to \mathbb{R}$ such that all the finite joint distributions equal those for X_t.

In fact we can take Ω to be the path space $\mathbb{R}^{\mathbb{R}}$ with the $\sigma-$algebra $\mathcal{F}_0$, generated by the algebra $\mathcal{F}$ of finite cylinder sets.

Proof. The consistency conditions are just a complicated way to say that (i) the joint distribution for $F = \{t_1, t_2 \dots, t_n\}$ define a measure on $\mathbb{R}^F$. Condition (ii) merely states that that this extends to a finitely additive measure on the finite cylinder algebra, that is, the algebra $\mathcal{F}$ of subsets of $\mathbb{R}^T$ generated by all the $\mathbb{R}^F$ for $F \subseteq T$ finite.

Now by Alexandroff's theorem, because of the key fact of compactness, this finitely additive measure is in fact *countably* additive on $\mathcal{F}$, hence as explained in the proof of that theorem, extends via the Caratheodory outer measure construction to a unique countably additive measure on the $\sigma-$algebra $\mathcal{F}_0$ it generates, finishing the proof. □

The finite cylinder $\sigma-$algebra $\mathcal{F}_0$ is not the only important $\sigma-$algebra on $\mathbb{R}^T$. If we give $\mathbb{R}^T$ the product topology $\mathcal{T}_p$, then by the Tychonoff product theorem, this is a compact Hausdorff space. We then define $\mathcal{F}_p$ to be the $\sigma-$algebra of Borel subsets, the $\sigma-$algebra generated by the topology. Note that a base for the $\mathcal{T}_p$ is the collection of open sets of the form $(\otimes \mathcal{U}_t)_{t\in F}$ such that $\mathcal{U}_t$ is open in $\mathbb{R}$. This is therefore also a base for the $\sigma-$algebra $\mathcal{F}_p$.

Sometimes this gives nothing new; indeed:

Proposition 33.25. *For a finite or countable index set T, the product and finite cylinder $\sigma-$algebras $\mathcal{F}_p$ and $\mathcal{F}_0$ are equal. If T is uncountable, then $\mathcal{F}_p$ is strictly larger.*

This rather surprising fact will be immediate from:

Lemma 33.26. *If T is uncountable,then for any $f \in \mathbb{R}^T$, the singleton $\{f\}$ is in $\mathcal{F}_p$ but is not in $\mathcal{F}_0$.*

Proof. We show that $\{f\}$ is a closed set for the product topology. Indeed let $g \neq f$. Thus there exists $t \in \mathbb{R}$ such that $g(t) \neq f(t)$. Let $\mathcal{U} \subseteq \mathbb{R}$ be an open neighborhood of $g(t)$ which misses $f(t)$. Then $\{h : h(t) \in \mathcal{U}\}$ is such a neighborhood.

□

To understand this interesting (and consequential) result we have to dive deeper into the nature of the Borel sets.

For uncountable index set T, there are disadvantages to this statement. The reason is because we would like to use the product topology $\mathcal{T}_p$ on $\dot{\mathbb{R}}^T$, where $\dot{\mathbb{R}}$ is the one-point compactification, as by the Tychonoff product theorem, this is a fairly nice topological space (a compact Hausdorff space), together with the corresponding Borel $\sigma-$ algebra $\mathcal{B}_p$. However:

Proposition 33.27. *For T finite or countable, $\dot{\mathbb{R}}^T$ with the product topology $\mathcal{T}_p$ is metrizable, and the Borel $\sigma-$ algebra $\mathcal{B}_p$ equals the finite cylinder $\sigma-$ algebra $\mathcal{F}_0$. Moreover, the Baire and Borel $\sigma-$ algebras of $\mathcal{T}_p$ are equal. When T is uncountable, $\dot{\mathbb{R}}^T$ with the product topology $\mathcal{T}_p$ is a nonmetrizable compact Hausdorff space, and the Borel $\sigma-$ algebra $\mathcal{B}_p$ is (much) larger than the finite cylinder $\sigma-$ algebra $\mathcal{F}_0$. Moreover, the Baire $\sigma-$ algebra of $\mathcal{T}_p$ is strictly smaller than the Borel $\sigma-$ algebra.*

The reason this causes difficulties is subtle, and has to do with what we mean by a stochastic process for continuous time:

Definition 33.16. A **stochastic process** $X(t)$ with values in $\mathbb{R}$ and time index set T is a measure μ on the finite cylinder $\sigma-$ algebra $\mathcal{F}_0$. A **version** of a stochastic process is a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ with a map $\Phi : \Omega \to \mathbb{R}^R$ where $X(\omega, t) = \Phi(\omega)(t)$, such that the measure $\Phi_*(P) = \mu$. Two versions are **equivalent** iff they map to the same measure μ on $\mathcal{F}_0$.

Remark 33.4. Equivalently, making use of a beautiful theorem of Caratheodory, the image set $\Phi(\Omega)$ has $\mu-$ outer measure one. However, this may perfectly well be a *non*measurable subset of $\dot{\mathbb{R}}^T$. See Example 31.

The problem is that when T is uncountable, a process can have many equivalent versions which are not equal. Indeed, it can happen that the ranges of two such version maps Φ_1, Φ_2 are disjoint.

For example, one can find a version (the usual one) of Brownian motion with all paths continuous, and another with all paths discontinuous. See [Fis87], pp. 228-230.

A strengthening of the Kolmogrov theorem, due to (in reverse chronological order) Dudley-Nelson-Kakutani-Doob, helps resolve this.

Replacing $\mathbb{R}$ by its one-point compactification $\dot{\mathbb{R}}$, then by Tychonoff's product theorem $\dot{\mathbb{R}}^T$ is a compact Hausdorff space. Write $\mathcal{B}$ for the Borel $\sigma-$algebra of $\dot{\mathbb{R}}^T$.

Theorem 33.28. *(Nelson) Assume the hypotheses of Kolmogorov's theorem. Then there exists a unique regular Borel measure (for the product topology) μ on $\dot{\mathbb{R}}^T$ which extends the measure on the finite cylinder $\sigma-$algebra $\mathcal{F}_0$ given in that theorem to the Borel $\sigma-$algebra $\mathcal{B}$ for the product topology $\mathcal{T}_p$.*

We call this the **regular extension**. In other words, by Nelson's theorem. there exists a *unique regular Borel* version of the process.

Remark 33.5. In probability theory we often think of this collection of values $\{P[(X_1 \in A_1) \wedge (X_2 \in A_2)]\}_{A_i \in \mathcal{B}(\mathbb{R})}$. itself to be the joint distribution, rather than the measure. The question then becomes, when do they define a measure? More precisely, given random variables X_1, X_2 defined on perhaps different probability spaces, thus $(\Omega_i, \mathcal{A}_i, P_i)$ for $i = 1, 2$, and a collection of such numbers called $\{P[(X_1 \in A_1) \wedge (X_2 \in A_2)]\}_{A_i \in \mathcal{B}(\mathbb{R})}$., then when can we find a single space $(\Omega, \mathcal{A}, P)$, and random variables $\widehat{X}_1, \widehat{X}_2 : \Omega \to \mathbb{R}$, so that now we have a map $\Phi : \Omega \to \mathbb{R}^2$ defined by $\Phi(\omega) = (X_1(\omega), X_2(\omega))$ and the pushed-forward measure $\Phi_*(P)$) such that (i) the distribution of X_I and $\widehat{X}_I$ are the same; and (ii).... (INCOMPLETE)

33.10. **Construction of Brownian motion; continuity of paths.**

34. Choosing a point randomly from an amenable. or nonamenable (!), group.

Remark 34.1. Invariant means on noncompact amenable groups (for example, $\mathbb{Z}^n$ or $\mathbb{R}^n$, or more generally any noncompact abelian group) provide natural examples of finitely additive measures with no countably additive extension. See §48.

An examination of this for other noncompact groups brings us to the idea of **amenability**, and to the beginning of a long and fascinating story.

For any locally compact group there is a natural countably additive measure called **Haar measure**; if the group is infinite this is infinite, and is then only unique up to normalization (i.e. up to multiplication by a positive constant). the Haar measure for $\mathbb{Z}^d$ is counting measure, while the Haar measure on $\mathbb{R}^d$ is Lebesgue measure.

Definition 34.1. An **invariant mean** on a group G with Haar measure m is a continuous positive normalized translation invariant linear functional on $L^\infty(G, m)$.

A group G is termed **amenable** (a pun; it should be "a*mean*able!") iff there exists an invariant mean on it (on $l^\infty(G)$ if the group is discrete; on $L^\infty(G, m)$ where m is Haar measure if G is a continuous group which is locally compact, so Haar measure exists). [Gre69]

There are several equivalent notions; here for simplicity we restrict to discrete groups:

–There exists a finitely additive probability measure on G;

Furst defn of amenable.

Harmonic projection.

Save rest for after boundary- in examples section!

34.1. **Choosing a point randomly from a nonamenable group, or from hyperbolic space.** Boundary at infinity.

Equivalent notions of non-amenable.

Basic example: F_2.

Appendix: Aside: Cayley graph of a semigroup or group. F_2 to $\mathbb{Z}^2$. Generators and relations.

Simplest case: finitely generated discrete.

μ on generators. Top invariant mean: Harmonic function. Harm projection. Equivalent defs. Boundary at ∞. Boundary values. Mokobodski mean.

Test functions: group actions. Random/ Markov ergodic theorem.

Fractal sets, again (top invt measures; IFS)

34.2. **Invariant means, Mokobodski and Fubini and Ergodic Thm.**

35. Appendix: Linear Algebra, Vector Calculus, Lie Groups and Differential Equations

We begin by developing basic material on Linear Algebra needed throughout these Notes. We are guided here in particular by what we need for the following sections (further "minicourses") on Vector Calculus and on Ordinary Differential Equations: some basics on the determinant and trace; on canonical forms of matrices, on the exponential of matrices, and on linear flows. This fits nicely into an introduction to Lie Groups and Algebras.

Our overall approach in these Notes has been to take a "modern" approach to dynamics, where connections with ergodic theory, probability and fractal geometry are emphasized, rather than the more "traditional" approach of stressing ODEs, the local theory of vector fields, iteration of maps of the interval and so on. In our treatment below of ODEs we emphasize certain links with dynamics, more to develop certain themes than to engage in a study of either particular examples or of the qualitative theory. These deep matters are extensively developed in many texts, and as a relative outsider we ourselves have learned only bits of this material. Another interesting direction would make serious links of this initial study of ODEs with the dynamics and ergodic theory. But this would necessitate background (for us) twice as deep and broad, and a book at least twice this length (stable manifolds and structural stability; much more on the Thermodynamic Formalism; Pesin Theory...) See e.g. [KH95] for some further places to start on that wide and deep adventure.

In the Differential Equations minicourse we present an introduction to the classical theory of differential equations encountered in undergraduate math courses, in one and higher dimensions. However our perspective is more at the level of a graduate student or a reseacher in related areas, in particular who already has some background in analysis and algebra. An undergraduate should skip along and learn what is accessible to them now, without demanding of themselves yet an understanding of every step. A more advanced reader should also skip along, trying to not be too irritated by too much explanaton of things already known. Hopefully there will still be parts that are new or interesting.

The first minicourse is on Linear Algebra, but before moving on to that we highlight some disparate questions which can serve as motivation for the overall minicourses in Vector Calculus and in ODE. Our primary overall aim is to introduce the connections with dynamics, in particular in the Vector Calculus section to look at some examples of vector fields, and the tools used: line integrals, curl and divergence, Stokes' Theorem; in the ODE section to describe the passage from flows to vector fields and back again, in both the stationary and nonstationary settings.

In the course of this, we try to look at each concept from different or complementary points of view: both algebraically and analytically, and geometrically. Here are some motivating questions we shall encounter along the way:

–What is the meaning of the determinant beyond its having a complicated formula?

–What is the geometrical significance of the trace?

–How can we understand orientation? Why are there only two orientations for $\mathbb{R}^n$, for each $n \geq 1$?

–How many inner products are there? How are they related?

–How far does the definition of the exponential map extend? To the complex numbers? To matrices? To manifolds?

–What is the analogue of the rotation flow for an indefinite bilinear form?

–What is the tangent space of a continuous group?

–What is the meaning of the Lie bracket?

–What is the geometric meaning of the vector product, its connection with the determinant, and its relation to 2-forms and surface area?

–What is the link between the vector product in $\mathbb{R}^3$ and the Lie algebra of the group of space rotations?

–What is the derivative of a flow?

–What is the geometrical meaning of an ordinary differential equation?

–How can one classify linear flows in $\mathbb{R}^2$ up to (linear) conjugacy?

–How can one prove that indeed, a flow preserves volume iff its vector field has divergence zero? (Hint: this has to do with the relationship between trace and determinant!)

–What is a nonstationary flow, and what does this have to do with nonautonomous vector fields?

–What is Picard iteration in a concrete setting? What is the link to Taylor's series?

–How does a differential 2-form in two dimensions naturally define a pair of differential equations in one dimension, and what does exactness of the form have to do with so-called exact differential equations, and harmonic conjugates? What are some examples in Electrostatics?

35.1. **Minicourse on Linear Algebra, Lie groups and Lie algebras.**

35.2. **Two definitions of the determinant.**

Algebraic definition: Let A be an $(n \times n)$ real or complex matrix. We begin with the usual algebraic definition, which is inductive on n. For $n = 1$, $A = [a] = [A_{11}]$ and $\det A$ is just the number a. For $n = 2$, $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, and we set $\det(A) = ad - bc$.

This is extended as follows: we define a matrix with entries $S_{ij} \in \{1, -1\}$ as follows: $S_{ij} = (-1)^{i+j}$. To visualize this, we write simply the corresponding signs, in a checkerboard pattern:

$$S = \begin{bmatrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{bmatrix}$$

The *ij minor* $A(ij)$ of A is defined to be the $(n-1) \times (n-1)$ matrix formed by removing the i^{th} row and j^{th} column of A.

Then we *expand along the top row* by forming the sum of $(\pm 1)\det A(1j)$, where the signs are given by the top row of S, i.e.

$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} \det A(1j).$$

Similarly we define the expansion along any row $\sum_{j=1}^{n}(-1)^{i+j}\det A(ij)$ or indeed any column.

It turns out these are equal, giving the same number whatever row or column chosen.

Note that this algorithm also works for the (2×2) case!

Geometric definition:

Definition 35.1. Let M be an $(n \times n)$ real matrix. Then

$$\det M = (\pm 1)(\text{factor of change of volume}$$

where we take $+1$ if M preserves orientation, -1 if that is reversed.

Theorem 35.1. *We have the following consequences:*
(i) $det(AB) = (detA)(detB)$*;*
(ii) A *is invertible iff* $detA \neq 0$*.*

Theorem 35.2. *The algebraic and geometric definitions are equivalent.*

Proof. For A (2×2), note that the factor of change of volume is the area of the image of the unit square, that generated by the standard basis vectors $(1, 0)$ and $(0, 1)$, which equals the area of the parallelogram with sides the matrix columns, (a, c) and (b, d).

Case 1: $c = 0$. Then the matrix is upper triangular and its determinant agebraically is ad. But the parallelogram area is (base)(height)$= ad$ as well.

Note that instead of using the formula for the area of a parallelogram, we can transform this to the diagonal (rectangle) case, by an operation of *sliding* the top of parallelogram in the direction of its base, the vector $(a, 0)$.

General Case: We reduce to Case 1 as follows, *not* by rotating (also possible!) but by sliding the far side of the parallelogram along the direction (b, d). A simple computation shows the area is indeed $ad - bc$.

Higher dimensions: We note that the above "sliding" operations can be done algebraically by an operation of column reduction, equivalently, multiplying on the right by an elementary matrix of determinant one. This reduces to the upper diagonal case, and beyond to the diagonal case if desired.

The same procedure works in $\mathbb{R}^3$ and beyond.

□

35.3. **Orientation.** We may be accustomed to thinking of a certain basis as having positive orientation and another negative, but this has no intrinsic meaning: what does make sense is to say that two given bases have the same or different orientation. As we shall explain, there are only two choices,

Thus, given $\mathbb{R}^n$, we let $\widehat{\mathcal{B}}$ denote the collection of all bases. The change from one basis $\mathcal{B}_1$ to another $\mathcal{B}_2$ is given by an invertible matrix A. By definition the collection of such matrices is $GL(n, \mathbb{R})$. Now the group GL has a (non-normal) subgroup of index 2, GL^+. These are characterized by $\det A > 0$. We say these are the *orientation-preserving* matrices. Letting GL act on the bases $\widehat{\mathcal{B}}$, we define two bases $\mathcal{B}_1, \mathcal{B}_2$ to have the *same orientation* iff one is taken to the other by an element of GL^+. Since

this subgroup has index 2, there are only these two choices, so the second case is expressed by saying they have *opposite* orientation.

Then, choose one basis $\mathcal{B}_1$ we declare (arbitrarily) that this has *positive* orientation. The GL^+-orbit of this defines $\widehat{\mathcal{B}^+}$, the bases with positive orientation, and the complement defines $\widehat{\mathcal{B}^-}$, the bases with *negative orientation.* Note that $\widehat{\mathcal{B}^-}$ is the GL^+-orbit of any $\mathcal{B}_2$ not in $\widehat{\mathcal{B}^+}$.

35.4. **Eigenvectors and eigenvalues.**

Lemma 35.3. *Let $T : V \to V$ linear and assume that $(\mathbf{v}_1, \dots \mathbf{v}_n)$ are eigenvectors with eigenvalues $(\lambda_1, \dots, \lambda_n)$. If all the eigenvalues are distinct, then the vectors are linearly independent.*

Proof. (we learned this nice proof from Theorem 5.6 of Axler [Axl97]). Since they are eigenvectors, by definition each is nonzero. If we consider the sequence $(\mathbf{v}_1, \dots \mathbf{v}_n)$ then if these are *not* linearly independent, there must be a least index k such that $\mathbf{v}_k$ can be written as a linear combination of the previous vectors, so $\mathbf{v}_k = \sum_{j=1}^{k-1} a_j \mathbf{v}_j$. Multiply this by the eigenvalue for v_k, giving $\lambda_k \mathbf{v}_k = \sum_{j=1}^{k-1} \lambda_k a_j \mathbf{v}_j$. Now apply T to the equation $\mathbf{v}_k = \sum_{j=1}^{k-1} a_j \mathbf{v}_j$, giving $\lambda_k \mathbf{v}_k = \sum_{j=1}^{k-1} \lambda_j a_j \mathbf{v}_j$. Subtracting the two equations we have $\mathbf{0} = \sum_{j=1}^{k-1} (\lambda_k - \lambda_j) a_j \mathbf{v}_j$. Since the vectors $(\mathbf{v}_1, \dots \mathbf{v}_{k-1})$ are linearly independent, k having been the least such index, each coefficient $(\lambda_k - \lambda_j) a_j = 0$. However the eigenvalues are distinct so $\lambda_k - \lambda_j \neq 0$ for all j. Therefore $a_j = 0$ for all j. But then $\mathbf{v}_k = \mathbf{0}$, a contradiction.

□

Corollary 35.4. *(Axler Cor. 5.9) A linear operator on a vector space of dimension n has at most n distinct eigenvalues.*

□

Theorem 35.5. (Fundamental Theorem of Algebra, complex case) *Every polynomial of degree n, $p(z) = a_0 + a_1 z + \cdots + a_n z^n$ has a factorization, unique up to order, into a constant times n linear factors $p(z) = c_0(z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_n)$.*

(Fundamental Theorem of Algebra, real case) *If the entries a_j are real, then $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ has a factorization, unique up to order, into a constant times n factors wihich are either linear or irreducible quadratic: $p(x) = c_0(x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_m)(x^2 + \beta_1 x + \gamma_1) \cdots (x^2 + \beta_l x + \gamma_l)$, with $\beta_j, \gamma_j \in \mathbb{R}$ and the* discriminant *$\beta^2 - 4\gamma < 0$ (so roots are complex).*

For proofs see e.g. [Axl97], [Ahl66], [MH87], [GP74].

Lemma 35.6. *The complex roots of the quadratic factors of a real polynomial occur in conjugate pairs, λ and $\overline{\lambda}$.*

Proof. The roots of $x^2 + \beta x + \gamma$ are, from the quadratic formula (or directly from completing the square), $\frac{1}{2}(-\beta \pm \sqrt{\beta^2 - 4\gamma})$. If $C = (\beta^2 - 4\gamma) < 0$ then this has two imaginary roots, $\pm iA$ where $A = \sqrt{|C|}$, whence the roots are $\frac{1}{2} \cdot (-\beta \pm iA)$. □

Theorem 35.7. *Any linear operator on a vector space of dimension $n \geq 1$ has at least one eigenvalue.*

Proof. (Axler Theorem 5.10)

Let $\mathbf{v} \neq \mathbf{0}$ in V and consider the $(n+1)$ vectors $(\mathbf{v}, T(\mathbf{v}), \ldots, T^n(\mathbf{v}))$. By the Corollary these are linearly dependent, hence there exist $a_i \in \mathbb{C}$ not all 0, such that

$$\mathbf{0} = a_0\mathbf{v} + a_1 T(\mathbf{v}) + \cdots + a_n T^n(\mathbf{v}).$$

Let m be the largest index k of a_k such that $a_k \neq 0$. Define a polynomial of degree m from this:

$$p(z) = a_0 + a_1 z + a_2 z + \cdots + a_m z^m$$

so the above equation is

$$\mathbf{0} = p(T)\mathbf{v}.$$

Now factor the complex polynomial by the Fundamental Theorem of Algebra:

$$p(z) = a_0 + a_1 z + a_2 z + \cdots + a_m z^m = c(z - \lambda_1) \cdots (z - \lambda_m)$$

where $c \neq 0, \lambda_k \in \mathbb{C}$.

Thus

$$\mathbf{0} = p(T)(\mathbf{v}) = c(T - \lambda_1 I) \cdots (T - \lambda_m I)(\mathbf{v}) = c(T - \lambda_1 I) \circ (T - \lambda_2 I) \cdots \circ (T - \lambda_m I)(\mathbf{v})$$

Hence either $(T - \lambda_m I)(\mathbf{v}) = \mathbf{0}$, or $(T - \lambda_m I)((T - \lambda_m I)(\mathbf{v})) = \mathbf{0}$, and so on, whence one of the operators $(T - \lambda_k I)$ has a nontrivial kernel. This shows there is an eigenvalue. □

The usual proof is of course the following; Axler's argument is not only more beautiful but simpler, as it avoids the use of determinants altogether:

Standard proof: The characteristic polynomial of T is $p(z) = \det(T - xI)$. This has a root λ by the Fundamental Theorem of Algebra, whence $0 = p(\lambda) = \det(T - \lambda I)$ so the operator $(T - \lambda I)$ is noninvertible, equivalently there exists a nonzero vector $\mathbf{v}$ in its kernel. Then $(T - \lambda I)(\mathbf{v}) = \mathbf{0}$ whence $\mathbf{v} \neq \mathbf{0}$ is an eigenvector and λ its eigenvalue.

Upper triangular form.

The *diagonal* of an $(n \times n)$ complex matrix A is $(A_{11}, A_{22}, \ldots A_{nn})$. It is in *upper triangular form* iff all elements below the diagonal (those for which $i > j$)are 0, e.g.:

$$\begin{bmatrix} 1 & -1 & 0 & 2 \\ 0 & 2 & 1 & 7 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

It is easy to check that:

Lemma 35.8.

(i)The eigenvalues of upper triangular A are the diagonal elements.
(ii) A is upper triangular iff for all k, $A\mathbf{v}_k$ is in the span of $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$.
(iii)Equivalently to (ii), the span of $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$ is invariant for each k.
(iv) If $T : V \to V$ is a linear operator, then if we can find a basis satisfying (ii),equivalently (iii), then in this basis T will be an upper triangular matrix. □

If A is an $(n \times n)$ matrix perhaps not upper triangular, then the transformation defined by multiplying column vectors on their left by A, we say *A has an upper triangular form* iff we can find an invertible matrix B such that $B^{-1}AB = U$ is upper triangular: the columns of the basis-change matrix B are just the new basis vectors. The next theorem says we can always do this, over $\mathbb{C}$:

Theorem 35.9. *In a complex vector space V of dimension n, given a linear operator T, there exists a basis which puts T into upper triangular form.*

Proof. (Adler p. 84) The proof is by induction on the dimension n. For $n = 1$ this is clear. By Theorem 35.7 since this is over $\mathbb{C}$, there exists at least one eigenvalue λ. We consider the map $T - \lambda I$. Since this is noninvertible, the image U is a subspace of V of strictly smaller dimension.

We claim that U is T-invariant. Let $\mathbf{u} \in U$, then $T\mathbf{u} = T\mathbf{u} - \lambda\mathbf{u} + \lambda\mathbf{u} = (T - \lambda I)\mathbf{u} + \lambda\mathbf{u}$. Now the first term is in U since $\mathbf{u} \in V$, by the definition of U as the range; the second is also, whence $T\mathbf{u}$ is. So U is invariant for the map T.

Since U has smaller dimension, we can apply the induction hypothesis to $T : U \to U$ and find a basis $(\mathbf{u}_1, \ldots, \mathbf{u}_m)$ of U such that for each k, $T(\mathbf{u}_k)$ is in the span of $(\mathbf{u}_1, \ldots, \mathbf{u}_k.)$ We complete this to a basis of V, $(\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{v}_1, \ldots, \mathbf{v}_l)$. We claim this will put $T : V \to V$ in upper triangular form. We calculate $T(\mathbf{v}_k)$.

Now $T(\mathbf{v}_k) = T(\mathbf{v}_k) - \lambda\mathbf{v}_k + \lambda\mathbf{v}_k = (T - \lambda I)(\mathbf{v}_k) + \lambda\mathbf{v}_k$. but from the definition of U, $(T - \lambda I)(\mathbf{v}_k))$ is in $U = \text{span}(\mathbf{u}_1, \ldots, \mathbf{u}_m)$, while $\mathbf{v}_k \in \text{span}((\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{v}_1, \ldots, \mathbf{v}_k)$. This shows that $T(\mathbf{v}_k) \in \text{span}((\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{v}_1, \ldots, \mathbf{v}_k)$, so we are done.

One surprising point here is that we don't need to use the $\mathbf{v}_j's$ of lesser index. But that is precisely because $(T - \lambda I)(\mathbf{v}_k)$ is in U, so is in the span of the $\mathbf{u}_j's$. □

Lemma 35.10. *If a matrix A has eigenvector $\mathbf{v}$ with eigenvalue λ, then $C = B^{-1}AB$ has eigenvalue λ for the eigenvector $\mathbf{w} = B^{-1}\mathbf{v}$.*

Proof. $C\mathbf{w} = B^{-1}AB\mathbf{w} = B^{-1}ABB^{-1}\mathbf{v} = B^{-1}A\mathbf{v} = B^{-1}\lambda\mathbf{v} = \lambda\mathbf{w}$. □

Lemma 35.11.
(i)$det(AB) = det(A)det(B)$.
(ii)$det(B^{-1}AB) = detA$.
(iii) The characteristic polynomials of A and $\widehat{A} = B^{-1}AB$ are equal.

Proof. We proved (i) geometrically in ???. Part (ii) follows.

For (iii), $B^{-1}AB - zI = B^{-1}(A - zI)B$ so

$$p_{\widehat{A}}(z) = \det(B^{-1}AB - zI) = \det(B^{-1}(A - zI)B) = \det(A - zI) = p_A(z).$$

□

We note that:

Proposition 35.12. *If U is upper triangular, then:*
(i) $det(U)$is the product of its diagonal elements.
(ii) U is invertible iff it has no 0's on the diagonal.
(iii) The eigenvalues of U are its diagonal elements.

(iv) det(U)is the product of its eigenvalues.

For A any $(n \times n)$ matrix, then

(v) det(A) is the product of its eigenvalues.

(vi) for A any matrix and $U = B^{-1}AB$ upper triangular, then A and U have the same eigenvalues and the same characteristic polynomials.

Proof. To calculate the determinant we expand along the first column of U. At the first step we have U_{11}. The determinant is U_{11} times $\det(U_1)$ where (U_1) is the smaller matrix with the first column and row removed. By induction we are done, so $\det(U) = \Pi_{j=1}^{n} U_{jj}$. Part (ii) follows, since U is invertible iff $\det(U) \neq 0$.

For (iii), the eigenvalues are the roots of $p_U(z) = \det(U - zI)$. Now U is upper triangular so $U_z \equiv (U - zI)$ is upper triangular. From part (i), the determinant of U_z is the product of its diagonal elements. The diagonal element $(U_z)_{ii}$ is 0 iff $z = \lambda_i$. This proves the claim.

Part (iv) follows from (i) and (iii).

To prove (v), by Theorem 35.9, there exists B such that $U = B^{-1}AB$ is upper triangular. From Lemma 35.10, U and A have the same eigenvalues. From Lemma 35.4 they have the same determinant. By part (iv) $\det U$ is the product of its eigenvalues, so this passes over to A.

Part (vi) follows from the previous lemmas.

□

Proposition 35.13. *If A and B are $(n \times n)$ upper triangular, then so is AB.*

Proof. Now $(AB)_{ij} = \sum_{k=1}^{d} A_{ik}B_{kj}$ and this is zero if the i^{th} row of A has zeros where the j^{th} column of B is nonzero. This means for k such that $k < i$ and $k < j$ which always holds if $j < i$.

For a transparent geometric proof, draw a graph (a two-step Bratteli diagram) with alphabets $\{1, 2, \ldots n\}$ at times $0, 1, 2$ and a path from i to k at time 0 iff $A_{ik} \neq 0$, a path from k to j at time 1 iff $B_{ik} \neq 0$. That the matrices are upper triangular means these graphs flow upwards or across. This then holds for the composed paths, which corresponds to the composed matrices.

□

Remark 35.1. Once we have defined the exponential of a matrix, a corollary of Proposition 35.12 will be that $e^{\mathrm{tr}A} = \det(e^A)$. This has an interesting physical and geometrical interpretation! See Theorem 35.35.

35.5. **Exponentiation of matrices.** The general context for this is the relationship between Lie groups and Lie algebras, which we have encountered above. Recalling Remark 23.5, a Lie group is a manifold with a smooth group structure. Lie groups can be finite or infinite dimensional, in which case they are based on Banach manifolds, see §37.13 below. The prototype of a finite dimensional Lie group is a matrix group; this encompasses a large part of the theory, as general Lie groups are studied by means of their *representations*, which are simply matrix groups which are factors.

The Lie algebra $\mathfrak{g}$ of a Lie group G is the tangent space at the identity element e; by translation, the tangent space at any other point is isomorphic to this. There is a

map which takes lines through $\mathbf{0}$ in $\mathfrak{g}$ to geodesics in G; this is called the exponential map, and sends $\mathfrak{g}$ to G; the image may not be all of G, as one has:

Theorem 35.14. *There is a map, the* exponential map $\exp : \mathfrak{g} :\to G_e$ *which is onto the largest connected subgroup containing* e.

There is an abstract definition of *Lie algebra*: it is is a vector space with an additional operation, the *Lie bracket*. One can construct a group for which this is the tangent space at e, and then the Lie bracket operation captures the infinitesimal noncommutativity of the group. See Definition 35.14 below.

Restricted to a one-dimensional subspace of the Lie algebra, the image of the exponential map is a curve in the group, the geodesic referred to above:

Definition 35.2. Given $a \in \mathfrak{g}$, then $\{\exp(ta) : t \in \mathbb{R}\}$ is the *one-parameter subgroup generated by* a. The element a (or any positive multiple $r \cdot a$) is called an *infinitesimal generator* for this one-parameter subgroup of G.

For matrix groups, the exponential map is in fact just the exponentiation of matrices, defined by the power series for exp. We define, for $A \in \mathcal{M}_n(K)$, $K = \mathbb{R}$ or $\mathbb{C}$:

$$e^A = \sum_{k=0}^{\infty} A^k/k! = I + A + A^2/2 + A^3/3! + \dots$$

As we shall see, this always converges.

Definition 35.3. Given a set X and a function $\tau : X \times \mathbb{R} \to X$, note that fixing $t \in \mathbb{R}$ then $\tau_t(x) = \tau(x,t)$ defines a map $\tau_t : X \to X$. Thus $\{\tau_t\}_{t\in\mathbb{R}}$ is a collection of maps on X.

We say τ defines a *flow* on X iff
(i) τ_0 is the identity map and
(ii) τ_t satisfies the *flow property*

$$\tau_{t+s} = \tau_s \circ \tau_t.$$

In algebraic terms, this is a special case of the much more general notion of group action, see Definition 2.1, as it is an *action of the additive group* $(\mathbb{R}, +)$. This is also known as a *one-parameter group of transformations*.

We think of the variable t as *time*; then τ_t is called the *time-t map* of the flow.

(There is a wider concept called a *nonstationary* or *nonautonomous* flow, which does not satisfy the flow property; see Defintition 37.16 below.)

The *orbit* of a point $x \in X$ is $\{\tau_t(x) : t \in \mathbb{R}\}$.

If X is a vector space V, then τ_t is a *linear flow* iff each map τ_t is linear. Note that by the flow property plus the fact that τ_0 is the identity, τ_t is bijective as its inverse is τ_{-t}. Thus each τ_t is a linear isomorphism of V.

Example 34. **(Rotation Flow 1)**

Consider the group of rotations of the plane, setting $a = \cos(2\pi t), b = \sin(2\pi t)$ and $R_t = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ and defining $G = \{R_t : t \in \mathbb{R}\} \cong \mathbb{T}^1 = \mathbb{R}/\mathbb{Z}$.

Noting that $R_{t+s} = R_s \circ R_t = R_t \circ R_s$, this is a flow.

The next result will show that these maps are of the form e^{tA}, for a certain matrix A, making the connection with Lie algebras. As we shall see, in fact all linear flows arise in this way.

Proposition 35.15. *(Rotation flow) For* $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, *then*

$$e^{tA} = R_t = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix}.$$

Proof. (First proof) For $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ we note that the powers of A have period 4, with $(A^0, A^1, A^2, A^3, \dots) = (I, A, -I, -A, \dots)$. We separate the Taylor series into even and odd terms. Writing $c = \cos t$ and $s = \sin t$, this gives:

$$\begin{aligned} &\exp(tA) = \sum_{k=0}^{\infty} (tA)^k/k! = \\ &I + tA + (tA)^2/2 + (tA)^3/3! + (tA)^4/4! + \cdots = \\ &(I + (tA)^2/2 + (tA)^4/4! + \dots) + (tA + (tA)^3/3! + (tA)^5/5! + \dots) = \\ &I(1 - t^2/2 + t^4/4! - t^6/6! + \dots) + A(t - t^3/3! + t^5/5! - \dots) = \\ &\begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix} + A \begin{bmatrix} s & 0 \\ 0 & s \end{bmatrix} = \begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix} + \begin{bmatrix} 0 & -s \\ s & 0 \end{bmatrix} = \begin{bmatrix} c & -s \\ s & c \end{bmatrix} \end{aligned} \tag{127}$$

as claimed.

(Second proof)

For $z \in \mathbb{C}$ with $z = a + bi$, we define a map $\Psi : \mathbb{C} \to \mathcal{M}_2(\mathbb{R})$ by $z \mapsto Z \equiv \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. Note that for $z = c + di$ and $W = \psi(w)$, $w \cdot z = (ac - bd) + (ad + bc)i$, so $\Psi(wz) = \Psi(w)\Psi(z) = WZ$. In other words, Ψ is a field isomorphism from $(\mathbb{C}, +, \cdot)$ to $(\mathcal{M}_2(\mathbb{R}), +, \cdot)$. □

Remark 35.2. It will follow (see below for the precise definitions) that the one-dimensional vector space of matrices $\mathfrak{g} = \left\{ \begin{bmatrix} 0 & -t \\ t & 0 \end{bmatrix} \right\}_{t \in \mathbb{R}}$. is the Lie algebra of the rotation group.

To study the general case, first we need to prove convergence:

Lemma 35.16.
(i)For any square matrix M with entries in $\mathbb{C}$ or $\mathbb{R}$ the series $\exp(M) \equiv \sum_{k=0}^{\infty} M^k/k!$ converges.
(ii)Let V be a Banach space (a complete normed vector space). Then the same holds for any continuous linear transformation $T : V \to V$.

To prove the Lemma, recall the proof for the real Taylor series of e^x, $\exp(x) = \sum_{k=0}^{\infty} x^k/k! = 1 + x + x^2/2 + x^3/3! + \dots$.

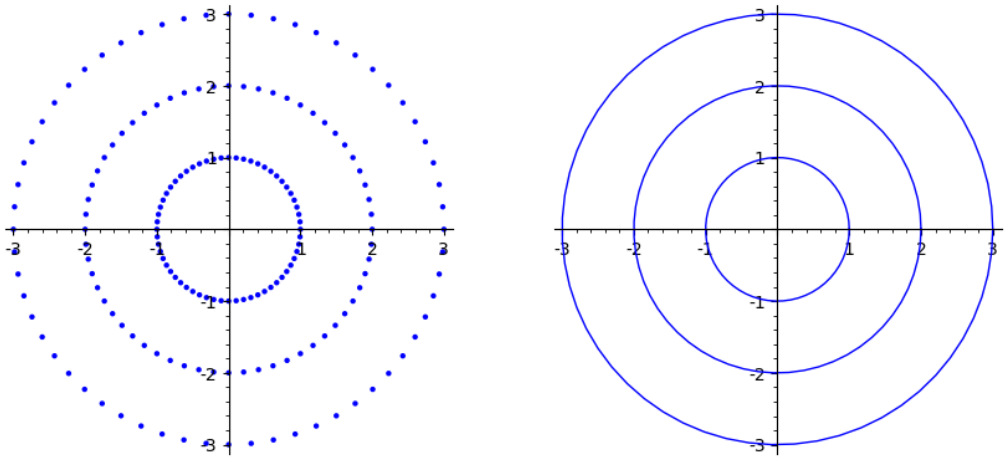


FIGURE 75. Some discrete and continuous-time orbits of the rotation flow, Example 35.15.

Fixing $x \in \mathbb{R}$, let m be so large that $|x|/m < 1/2$. Then for any $k \geq 0$, $|x|^{n+k}/(m+k)! < (1/2)^k$ whence the series $\sum_{k=0}^{\infty} |x|^k/k!$ converges, by comparison with the geometric series $\sum(1/2)^k$. Therefore the series for $x < 0$ also converges, by splitting it into odd and even terms; alternatively, we note that it is an alternating series with the terms decreasing in magnitude, whence it converges.

We are using here the completeness of the reals. We recall:

Definition 35.4. Given a metric space (X, d), $(x_n)_{n\geq 0}$ is a *Cauchy sequence* iff given $\varepsilon > 0$, there exists N such that $\forall n > N$, $d(x_n, x_{n+k}) < \varepsilon$ for all $k > 0$.

The metric space is *complete* iff every Cauchy sequence has a limit point.

The completeness of the reals. can be stated in several equivalent ways:
-any bounded increasing sequence has a limit;
-any Cauchy sequence converges;
-any bounded sequence has a least upper bound.

The axioms for the real numbers can be summarized as follows:
The reals $\mathbb{R}$ *are a complete (totally) ordered field.*

In fact this characterizes them uniquely (i.e. up to field isomorphism).

We note that completeness of $\mathbb{R}$ immediately implies that of $\mathbb{R}^n$, since a Cauchy sequence will be Cauchy in its components. In particular, $\mathbb{C}$ is complete, as the complex plane is homeomorphic to the real plane $\mathbb{R}^2$, and the space $CalM_{m\times n}(\mathbb{R})$ of $(m \times n)$ matrices is complete as it is homeomorphic to $\mathbb{R}^{mn}$, and similarly for complex entries.

Any normed finite-dimensional vector space is complete as all norms are equivalent in this case (see Lemma 35.41) , and convergence of coordinates is clearly equivalent to convergence in the sup (L^∞) norm.

Thus in particular $\mathcal{M}_{m\times n}(\mathbb{R}) = \mathcal{M}_{m\times n}(\mathbb{R})$ is complete for the operator norm. Recall this is defined as follows.

Let V, W be vector spaces (perhaps infinite dimensional) with norms

$$||\cdot||_V, ||\cdot||_W$$

. Then for $T : V \to W$ linear, we define

$$||T||_{\text{op}} = \sup_{\{\mathbf{v}: ||\mathbf{v}||_V = 1\}} ||T(\mathbf{v})||_W = \sup_V ||T(\mathbf{v})||_W / ||\mathbf{v}||_V.$$

This norm has the following very useful property:

Lemma 35.17. *The operator norm is* submultiplicative, *i.e.* $||T \circ S|| \leq ||S|| \cdot ||T||$.

Proof. This is immediate, but it is also instructive to draw a picture. □

Note that, applied to the complex numbers, this corresponds to the basic fact $|zw| \leq |z| \cdot |w|$.

Now to show $\exp(z) = \sum_{k=0}^{\infty} z^k/k! = 1+z+z^2/2+z^3/3!+\dots$, we show the sequence of partial sums is Cauchy by noting that the tail of this series, $|\sum_{k=N}^{\infty} z^k/k!| \leq \sum_{k=N}^{\infty} |z|^k/k! \to 0$ as $N \to \infty$.

Similarly, for any square matrix M, or more generally for any continuous linear operator T on a Banach space V, $\sum_{k=0}^{\infty} T^k/k!$ converges in exactky the same way, now using the operator norm to estimate the tail.

In our particular case, we consider the isomorphism for multiplication by w and A:

$$\begin{array}{ccc} \mathbb{C} & \xrightarrow{w\cdot} & \mathbb{C} \\ {\scriptstyle\Psi}\downarrow & & \downarrow{\scriptstyle\Psi} \\ \mathcal{M}_2 & \xrightarrow[A\cdot]{} & \mathcal{M}_2 \end{array}$$

Then for the exponential map we have, where $\mathbb{C}^*$ denotes the multiplicative group of nonero complex numbers,

$$\begin{array}{ccc} \mathbb{C} & \xrightarrow{w\cdot} & \mathbb{C}^* \\ {\scriptstyle\Psi}\downarrow & & \downarrow{\scriptstyle\Psi} \\ \mathcal{M}_2(\mathbb{R}) & \xrightarrow[A\cdot]{} & GL_2(\mathbb{R}) \end{array}$$

Now $\Psi(i) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. We recall

Lemma 35.18. *(Euler's formula) For all* $t \in \mathbb{R}$, $e^{it} = \cos t + i \sin t$.

Proof. The proof is much llike that for A above: since we know $\exp(z) = \sum_{k=0}^{\infty} (z)^k/k!$ converges, substituting $z = it$ and noting that $(i^0, i^1, i^2, i^3, \dots) = (1, i, -1, -i, \dots)$,

we separate the Taylor series into even and odd terms, and have:

$$\exp(it) = \sum_{k=0}^{\infty}(it)^k/k! =$$
$$1 + it + (it)^2/2 + (it)^3/3! + (it)^4/4! + \cdots = \tag{128}$$
$$(1 + (it)^2/2 + (it)^4/4! + \dots) + (it + (it)^3/3! + (it)^5/5! + \dots) =$$
$$(1 - t^2/2 + t^4/4! - t^6/6! + \dots) + i(t - t^3/3! + t^5/5! - \dots) = \cos t + i \sin t$$

as claimed. □

Now given the isomorphism $\Psi : \mathbb{C} \to \mathcal{M}_2(\mathbb{R})$, this is equivalent to convergence of the matrix series: $\exp(A) = \sum_{k=0}^{\infty}(tA)^k/k! = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix}$ shown above.

This same method works for any $A \in \mathcal{M}_2(\mathbb{R})$ which corresponds to complex multiplication. These are the linear transformations which are *conformal*: they preserve angle and orientation.

Here are some key properties of the exponential map:

Theorem 35.19. *For the map* $\exp : \mathcal{M}_n(\mathbb{C}) \to \mathcal{M}_n(\mathbb{C})$*:*
(a) For B *invertible,* $e^{BAB^{-1}} = Be^AB^{-1}$.
(b) If $AB = BA$*, then* $e^{A+B} = e^Ae^B$.
(c) $e^{-A} = (e^A)^{-1}$.
(d)If $\mathbf{v}$ *is an eigenvector of* A *with eigenvalue* λ*, then* $\mathbf{v}$ *is an eigenvector of* e^A *with eigenvalue* e^λ.

Before giving the proof we note that:

Corollary 35.20. *(Linear flows) For any* $(n \times n)$ *matrix* A*, then* $\tau_t = e^{tA}$ *defines a linear flow on* $\mathbb{R}^n$.

Proof. From part (b), since tA and sA commute, $e^{(t+s)A} = e^{sA}e^{tA}$. Furthermore for $\tau_t \equiv e^{tA}$. note that $\tau_0 = e^{0I} = I$. □

Here are some illustrative examples:

Example 35. (Exponential Repellor)

Proposition 35.21. *For* $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$*, then* $e^{tA} = \begin{bmatrix} e^t & 0 \\ 0 & e^t \end{bmatrix} = e^t \cdot I$*. The orbits flow out from the origin at exponential speed along straight lines.*

Example 36. (Hyperbolic Flow)

Proposition 35.22. *For* $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$*, then* $e^{tA} = \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}$ *The orbits are hyperbolas, level curves of the function* $F(x, y) = 2xy$.

The graph of $F : \mathbb{R}^2 \to \mathbb{R}$ is a saddle: a hyperboloid surface.

Example 37. (Hyperbolic Rotation Flow)

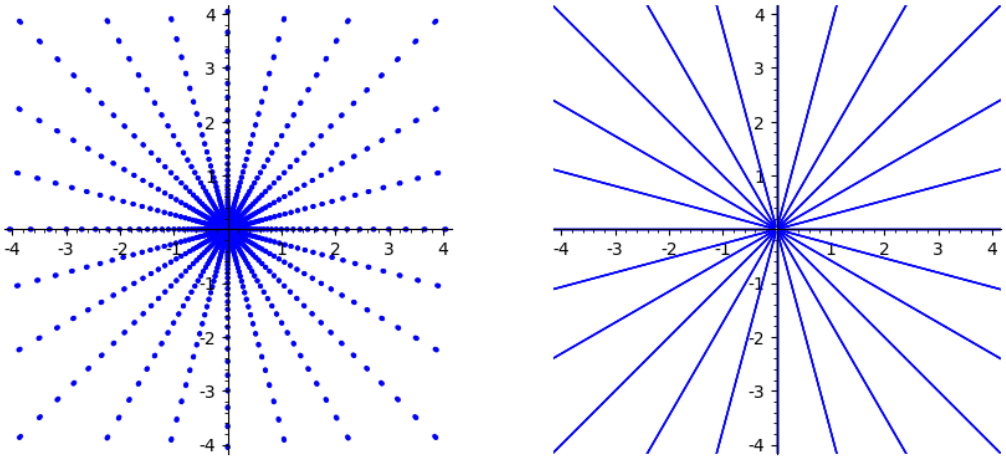


FIGURE 76. Some discrete and continuous orbits of the Exponential Repellor flow, Example 35. Points move away from origin exponentially fast.

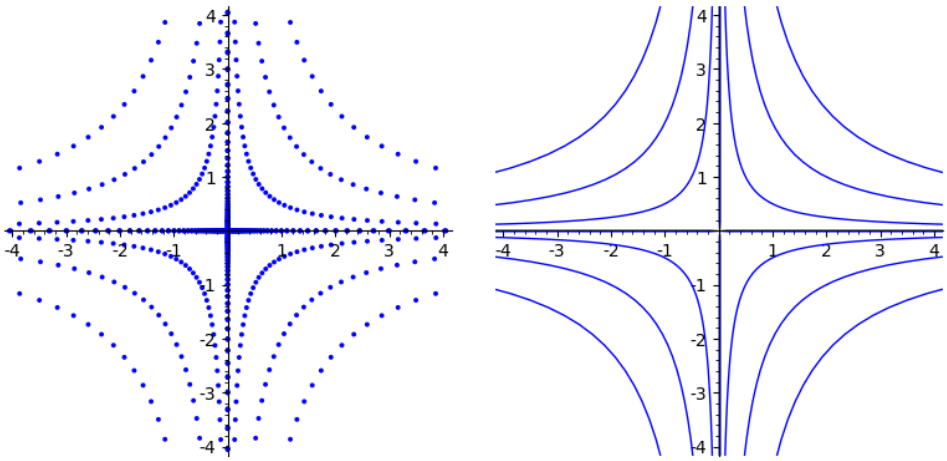


FIGURE 77. Some discrete and continuous orbits of the hyperbolic flow of Example 36.

Proposition 35.23. *For* $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, *we have* $e^{tA} = \begin{bmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{bmatrix}$ *The orbits are level curves of the hyperboloid* $F(x, y) = x^2 - y^2 = (x + y)(x - y)$.

Note that these matrices are reminiscent of the usual rotation matrices $\begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix}$. But this is not the real reason for calling this by the strange name of "hyperbolic rotation flow". Strange, because the words *hyperbolic* and *rotation* are usually contradictory, but the reason for this name is there are two distinct interpretations: one with respect to the Eucidean metric where we have a hyperbolic flow, in fact isometrically conjugated to the previous flow by a rotation of $\pi/4$;the second with respect to an

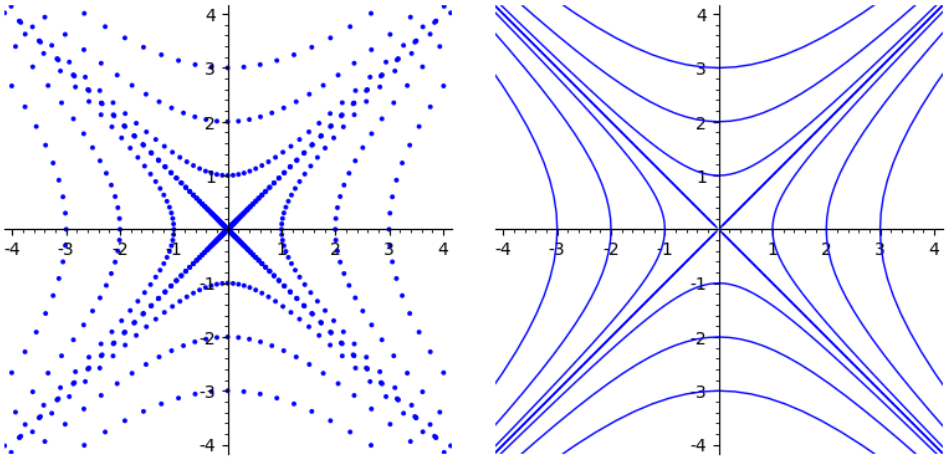


FIGURE 78. Some discrete and continuous-time orbits of the hyperbolic rotation flow of Example 37.

indefinite (Lorenz) metric. This metric is preserved by the flow, so can be interpreted as a rotation flow for that metric. See Example 52 below, where we give the proof.

As mentioned above, the orbits in both cases are level curves for a hyperboloid surface, the graph respectively of $F(x,y) = 2xy$ and $F(x,y) = x^2 - y^2$; these are quadratic forms for indefinite metrics; see Example 52. The matrices in the last example are called *hyperbolic rotations*, and in this case the quadratic form is *indefinite*. The Lorenz metric gives a two-dimensional (one space plus one time) model of Special Relativity, see Example 37 and §23.5.

Example 38. (Exponential Spiral)

Proposition 35.24. *For* $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, *then* $e^{tA} = e^t R_t$. *and the solution curves are exponential spirals.*

Proof. (Proof via matrices) For $B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, $C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, then $BC = CB$ hence $e^A = e^B e^C$, and also $e^{tA} = e^{tB} e^{tC}$, but $e^{tB} = e^t I$ while $e^{tC} = R_t$.

and for $B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, we note that $BC = CB$. Thus by (b) of Theorem 35.19, $e^{tA} = e^{t(B+C)} = e^{tB} e^{tC}$.

This is $\begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix} \begin{bmatrix} e^t & 0 \\ 0 & e^t \end{bmatrix} = e^t R_t$, and the solution curves are exponential spirals.

(Proof via complex numbers)

Repellor(TO DO...) □

Example 39. (Node: exponential repellor with parabola orbits)

Proposition 35.25. *For* $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$, *then the flow is* $e^{tA} = \begin{bmatrix} e^t & 0 \\ 0 & e^{2t} \end{bmatrix}$

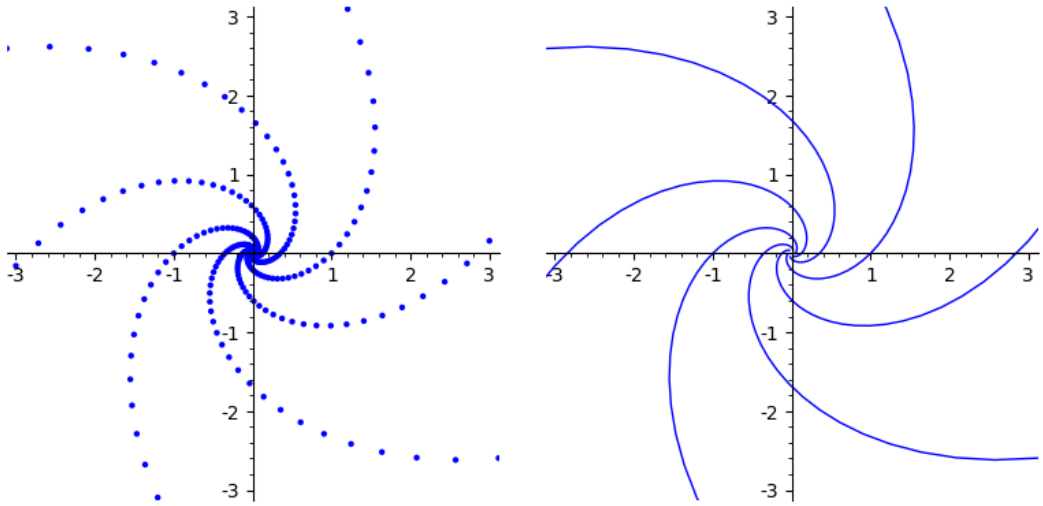


FIGURE 79. Exponential spirals for e^{tA}, of Example 38.

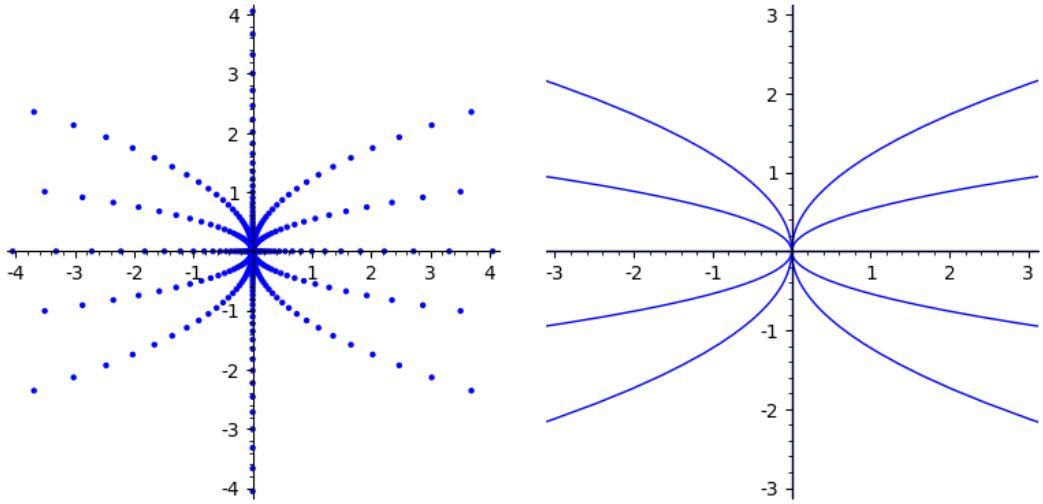


FIGURE 80. NodeRepellor: orbits for e^{tA}, of Example 39.

So the orbit of e.g. (a, b) is $(x(t), y(t)) = (e^t a, e^{2t} b)$ whence $(x/a)^2 \cdot b = x^2 \cdot (b/a^2) = y$, giving a family of parabolas which fill the plane. $(0, 0)$ is a *fixed point* for the flow, and all other points move away from the origin at exponential speed along these curves.

Example 40. (Linear Shear Flow)

Definition 35.5. A square matrix is *nilpotent* iff $N^k = 0 \cdot I$ for some $k \geq 0$.

For example, taking $N = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ then $N^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Now $N = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ is nilpotent. $e^{tA} = I + tA = \begin{bmatrix} 1 & 0 \\ t & 1 \end{bmatrix}$, which defines a (vertical) *Shear Flow* on the plane.

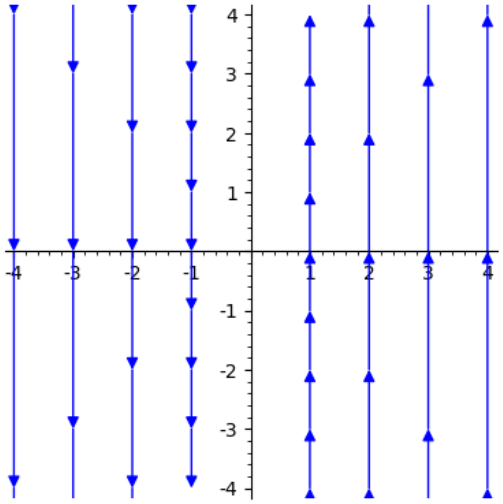


FIGURE 81. Vertical shear vector field of Example 40.

Example 41. (Improper Node Repellor)

The next example is a composition of two linear flows: the exponential repellor flow of Example 35 and the vertical shear flow of Example 40.

Proposition 35.26. *For* $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = I + N$, *the flow is given by* $e^{tA} = \begin{bmatrix} e^t & 0 \\ te^t & e^t \end{bmatrix}$.

Proof. We have $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = I + N$. Now $N = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ is nilpotent; I and N commute. By (b) of Theorem 35.19, in this case the addition of matrices is taken to composition of the flows. Thus $e^{tA} = e^{tI}e^{tN}$. As above, $e^N = I + N$, and $e^{tN} = I + tN = \begin{bmatrix} 1 & 0 \\ t & 1 \end{bmatrix}$ so

$$e^{tA} = \begin{bmatrix} e^t & 0 \\ 0 & e^t \end{bmatrix} \begin{bmatrix} 1 & 0 \\ t & 1 \end{bmatrix} = e^t \cdot \begin{bmatrix} 1 & 0 \\ t & 1 \end{bmatrix} = \begin{bmatrix} e^t & 0 \\ te^t & e^t \end{bmatrix}.$$

□

So the orbit of e.g. (a, b) is $(x(t), y(t)) = (e^t a, e^{2t} b)$ whence $(x/a)^2 \cdot b = x^2 \cdot (b/a^2) = y$, giving a family of parabolas which fill the plane. ??? $(0, 0)$ is a *fixed point* for the flow, and all other points move away from the origin at exponential speed along these curves.

This type of example is called an *improper node* e.g. in [HS74], and a *degenerate node* in [HK03].

Example 42. (Skew Hyperbola)

The next example similar to the previous one as it is is a composition of two linear flows: now replacing the exponential repellor by the hyperbolic flow of Example 36, again with the vertical shear flow of Example 40.

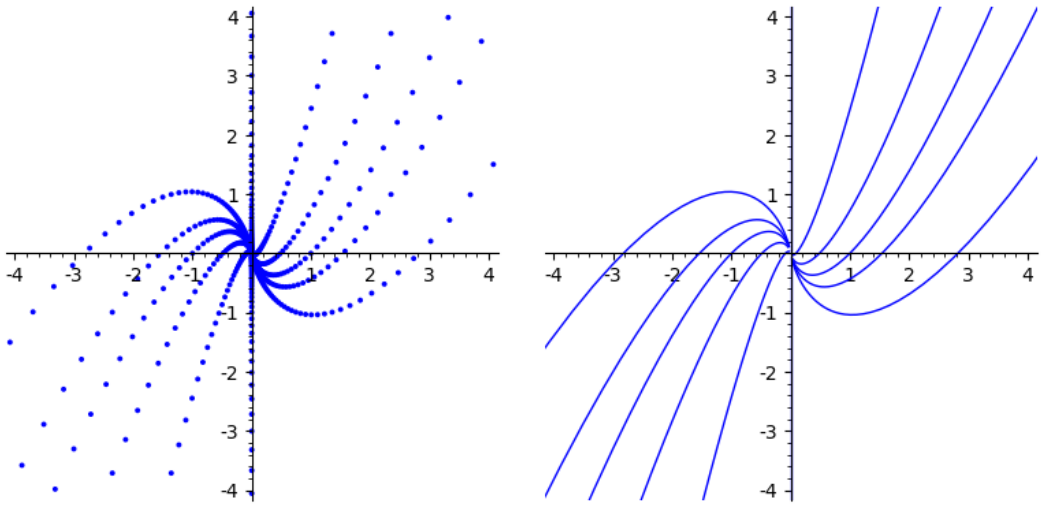


FIGURE 82. Improper Node Repellor: orbits for e^{tA} of Example 41.

Proposition 35.27. *For* $A = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} = I+N$, *the flow is given by* $e^{tA} = \begin{bmatrix} e^t & 0 \\ te^{-t} & e^{-t} \end{bmatrix}$.

Proof. We have $A = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} = K + N$ for $K = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

. Now $N = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ is nilpotent; again K and N commute. By (b) of Theorem 35.19, in this case the addition of matrices is taken to composition of the flows. Thus $e^{tA} = e^{tK}e^{tN}$. As above, $e^N = K + N$, and $e^{tN} = K + tN = \begin{bmatrix} 1 & 0 \\ t & -1 \end{bmatrix}$ so

$$e^{tA} = \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ t & -1 \end{bmatrix} = e^t \cdot \begin{bmatrix} 1 & 0 \\ t & -1 \end{bmatrix} = \begin{bmatrix} e^t & 0 \\ te^{-t} & e^{-t} \end{bmatrix}.$$

□

So the orbit of e.g. (a, b) is $(x(t), y(t)) = (e^t a, e^{-t}(ta+b)$ whence ??? giving a family of hyperbolas which fill the plane. $(0, 0)$ is a fixed point for the flow, and all other points move either toward or away from the origin at exponential speed along these curves.

This example is because the curl of the vector field is $+1$, which is interesting to understand (think of it projectively, i.e. the action on lines through the origin; sometimes they move counterclockwise and sometimes clockwise, but on average the motion is counterclockwise.) This contrasts with the square hyperbolic flows of Examples 36 and 37 which have curl 0, as the two projective rotations cancel out.

Since the matrices K and N commute, and the curl is the difference of the off-diagonal entries, the positive curl comes from the shear flow generated by N.

Classification of linear flows in the plane. The behavior of these flows can be classified, up to linear conjugacy, in terms of conjugacy classes of the matrix A. First we need the next lemma.

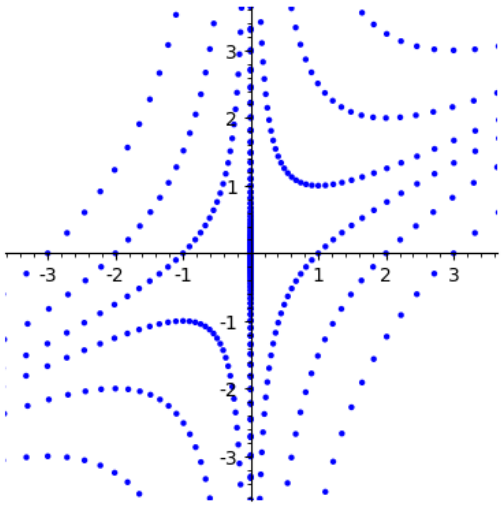


FIGURE 83. Skew Hyperbola: orbits for e^{tA} of Example 42.

The case of $SL(2,\mathbb{C})$ was considered in Lemma 25.1 and of $SL(2,\mathbb{R})$ in Lemma 35.28. The statements and proofs are similar; we include this for completeness.

Lemma 35.28. *Given $A \in GL(2,\mathbb{R})$, then either:*
(i) there are (up to nonzero multiples) two linearly independent eigenvectors $\mathbf{v}_0, \mathbf{v}_1$ or
(ii) there is one eigenvector $\mathbf{v}$, the eigenspace of which has dimension 1 or 2.
In case (i), the eigenvalues λ_0, λ_1 satisfy $\lambda_0 \cdot \lambda_1 = det(A)$, and A is similar to a diagonal matrix D with those entries. Thus $D = B^{-1}AB$ via a change-of-basis matrix $B \in GL(2,\mathbb{R})$ with columns $\mathbf{v}_0, \mathbf{v}_1$. $A = \pm\lambda I$ iff $\lambda \equiv \lambda_0 = \lambda_1$.

There are two subcases:
(ia) $(trA)^2 - 4 > 0$, and the eigenvalues are real, and
(ib)$(trA)^2 - 4 < 0$, and they are imaginary.
In case (ii), $(trA)^2 - 4 = 0$ and there is a double real eigenvalue. Then there exists a rotation matrix R such that $R^{-1}AR = T$ with $T = \pm det(A) \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix}$. Moreover A is conjugate in $GL(2,\mathbb{C})$ to $H^+ = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and to $H^- = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.

Proof. The characteristic polynomial of $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is: $p_A(z) = \det(A - zI) = z^2 - \mathrm{tr}A \cdot z + \det A$. The characteristic polynomial is

$$p_A(\lambda) = \det(A - \lambda I) = \lambda^2 - (\mathrm{tr}A)\lambda + \det A = \lambda^2 - (\mathrm{tr}A)\lambda + 1$$

with roots

$$\lambda^{\pm} \equiv \frac{\mathrm{tr}A \pm \sqrt{(\mathrm{tr}A)^2 - 4}}{2}.$$

Defining $\alpha = \sqrt{(\mathrm{tr}A)^2 - 4}$, then since the entries are real, if $(\mathrm{tr}A)^2 - 4 > 0$ we have $\lambda^{\pm} = (\mathrm{tr}A \pm \alpha)/2$. If $(\mathrm{tr}A)^2 - 4 < 0$ then writing $\beta = \sqrt{4 - (\mathrm{tr}A)^2} > 0$ we have

$\lambda^{\pm} = (\mathrm{tr}A \pm \beta i)/2$. Lastly if $(\mathrm{tr}A)^2 - 4 = 0$ then $\lambda^{\pm} = (\mathrm{tr}A)/2$. Note that in all three cases, $\lambda^+ \cdot \lambda^- = 1$, verifying what we already know from Lemma 25.1.

We know from Lemma 25.1 that, defining as above B to have a columns the eigenvectors, this gives a change-of-basis matrix in $GL(2, \mathbb{C})$. We now show in fact we can take the eigenvectors to be real and B to be in $SL(2, \mathbb{R})$. Now for an eigenvector (without loss of generality) $\mathbf{v} = \begin{bmatrix} z \\ 1 \end{bmatrix}$ we have

$$A\mathbf{v} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} z \\ 1 \end{bmatrix} = \begin{bmatrix} az + b \\ cz + d \end{bmatrix} = \begin{bmatrix} \lambda z \\ \lambda \end{bmatrix}$$

so $cz + d = \lambda$ whence $\lambda \in \mathbb{R} \implies z \in \mathbb{R}$. Thus up to multiplication by a constant in $\mathbb{C}^*$, the eigenvectors are real, whence $B \in SL(2, \mathbb{R})$.

$D = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$

□

Lemma 35.28

TO DO: trace and determinant ...

Some lemmas.

To prove part (b) of Theorem 35.19, we need:

Lemma 35.29. *(Binomial coefficient)*

$$(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

where the binomial coefficient*, read "n choose k (without order)", is*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

.

Proof. Let us consider the number of ways we can choose k objects from a set of n objects. There are two ways to do this, *with order* or *without order*. For example, if $k = n$, we want to choose n objects from a set with n objects. There are $n!$ ways to do this with order, and this equals the number of n-element sequences one can form from the set. That is also the number of permutations of the set. There is one way to choose without order, to take all of them, and this equals $n!$ divided by the number of permutations, or $n!/n! = 1$.

Choosing now k from n, first with order, labelling the choices by $0, 1, \ldots (k-1)$, then there are $n - 0 = n$ ways to choose the first, $(n-1)$ to choose the second, $(n - (k-1)) = n - k + 1$ to choose the last, giving $n \cdot (n-1) \cdots (n-k+1)$ ways total. This equals $n!/(n-k)!$.

To choose without order, we divide this by the number of permutations of the k objects, giving $n!/(n-k)!k!$.

Another way to state this is that the difference between choosing without order or with order is to count subsets, or sequences, with distinct elements.

The notation for the first of these is

$$\binom{n}{k} = \frac{n \cdot (n-1) \cdots (n-k+1)}{k!} = \frac{n!}{k!(n-k)!}$$

.

Next we see why this is indeed the binomial coefficient. Considering $(x+y)^n = (x+y)(x+y)\cdots(x+y)$ then note that the coefficient for $x^k y^j$ where $k+j=n$ or equivalently $j = n-k$ is the number of ways we can choose k letter $x's$ from the n pairs, given that we choose one letter x or y from each pair. Order is not important here so this is indeed equal to $\binom{n}{k}$.

For n fixed these are the numbers on the n^{th} row of Pascal's triangle. Note that $\binom{n}{k} = \binom{n}{n-k}$ and that $\sum_{k=0}^{n} \binom{n}{k} = 2^n$, since that is the total number of paths to make n choices of either letter. □

Lemma 35.30. *For real numbers, $e^{x+y} = e^x e^y$.*

Proof. We have on the left-hand side $e^{x+y} = \sum_{n=0}^{\infty} (x+y)^n/n!$ and on the right-hand side $e^x e^y = \left(\sum_{j=0}^{\infty} x^j/j!\right)\left(\sum_{k=0}^{\infty} y^k/k!\right)$.

On the left-hand side the term with $x^j y^k$ appears with coefficient

$$\binom{n}{k} = \frac{n!}{k!j!} = \frac{(j+k)!}{k!(j)!},$$

all divided by $n! = (j+k)!$ in the Taylor's series. This gives

$$\frac{(j+k)!}{k!j!(j+k)!} = \frac{1}{k!j!}.$$

On the right-hand side the coefficient for $x^j y^k$ is also $\frac{1}{k!j!}$.

Thus the (infinite) collections of terms are identical and since the series converge, the limits agree. □

Proof of Theorem. For (b),$(BAB^{-1})^n = (BA^nB^{-1})$ so $B(\sum_{k=0}^{n} A^k/k!)B^{-1} = \sum_{k=0}^{n}(BAB^{-1})^k/k!$ and since the sums converge, this holds in the limit.

Part (c) is proved in exactly the same way as for real numbers, given that $AB = BA$.

For (d), $e^A e^{-A} = e^{A-A} = e^{0 \cdot I} = I$ so this holds.

For (e), if for $\mathbf{v} \neq \mathbf{0}$) we have $A\mathbf{v} = \lambda \mathbf{v}$ for some $\lambda \in \mathbb{C}$, then
$(\sum_{k=0}^{n} A^k/k!)\mathbf{v} = (\sum_{k=0}^{n}(A^k\mathbf{v})/k!) = (\sum_{k=0}^{n}(\lambda^k\mathbf{v})/k!) = (\sum_{k=0}^{n} \lambda^k/k!)\mathbf{v} \to e^{\lambda}\mathbf{v}$.

□

In the next series of results we apply the existence of upper triangular form for complex matrices proved in Theorem 35.9.

Proposition 35.31.
(i) If A is upper triangular, then so is A^n for any $n \geq 1$.
(ii) If A is upper triangular, then: e^A is upper triangular.
(iii) If A is upper triangular, then: $e^{trA} = det(e^A)$.

Proof. From Proposition 35.13, if A is upper triangular, then so is A^2, whence by induction A^n is upper triangular for any $n \geq 1$, proving (i). Hence from the Taylor series, so is e^A. For (iii), for A upper triangular, note that from the Taylor series,

the entries on the diagonal of e^A are $\exp(A_{jj})$. Now we know from Proposition 35.13 that for A upper triangular, then $\det(A) = \Pi_{j=1}^n A_{jj}$. Thus

$$e^{\mathrm{tr}A} = \exp(\sum_{j=1}^{n} A_{jj}) = \Pi_{j=1}^n \exp(A_{jj}) = \det(e^A)$$

□

Lemma 35.32.
(i) $det(AB) = det(A)det(B)$.
(ii) $det(B^{-1}AB) = det(A)$.

Proof. Part (i) can be proved from the geometric definition of determinant: $\det(A) = (\pm 1) \cdot$ (factor of change of volume) where we have 1 if A preserves orientation, -1 if not.

Then (ii) follows from this. □

Lemma 35.33. *The characteristic polynomial is invariant for conjugacy,*

Proof. The characteristic polynomial is invariant for conjugacy, since

$$\det(B^{-1}AB - \lambda I) = \det(B^{-1}(A - \lambda I)B) = \det(A - \lambda I)$$

where we have used Lemma 35.32.

□

Proposition 35.34. *Given the characteristic polynomial*

$$p_A(z) = det(A - \lambda I) = c_0(z - \lambda_1)(z - \lambda_2)\cdots(z - \lambda_n) = a_0 + a_1 z + \cdots + a_n z^n,$$

then
(i) $c_0 = 1$, $a_0 = \Pi\lambda_i$ *and* $a_{n-1} = \sum \lambda_i$.
(ii) $a_0 = det(A)$ *and* $a_{n-1} = tr(A)$.

Proof. $p_A(\lambda) = \det(A - \lambda I) = c_0(z - \lambda_1)(z - \lambda_2)\cdots(z - \lambda_n) = a_0 + a_1 z + \cdots + a_n z^n$ so $c_0 = 1$, and the coefficient a_1 is given by choosing from each factor the constant, λ_i, so equals $\Pi\lambda_i$. The coefficient a_{n-1} of z is given by choosing z from all but one of the factors, then adding these, whence $a_{n-1} = \sum \lambda_i$. This proves (i).

For part (ii), by Theorem 35.9 we can find a matrix B for which $B^{-1}AB = U$ is upper triangular.

From Proposition 35.12, $\det(U)$ is the product of its diagonal elements. From Lemma 35.33, these are the eigenvalues of U, and from Lemma 35.8 these are also the eigenvalues of A.

We know that $\det(U) = \det(B^{-1}AB) = \det(A)$

□

Theorem 35.35. *If A is upper triangular, then:*
(a) so is e^A, with diagonal entries $e^{A_{ii}}$.
(b) Trace and determinant are related for upper triangular A by: $e^{trA} = det(e^A)$.

Proof. Part (a) is straigtforward. If A is upper triangular, then the determinant is the product of the diagonal entries. From Theorem 35.9, we know that over the complexes, A can be put in upper triangular form. So this follows from (a). □

Corollary 35.36.
(ii) Both trace and determinant are invariants for conjugacy. Combining this with Proposition 35.34 proves (ii).

Corollary 35.37. *The trace is preserved by conjugation.*

Corollary 35.38.
(c) For any $(n\times n)$ real matrix A, e^A is invertible and orientation-preserving, i.e. exp : $\mathcal{M}_n(\mathbb{R}) \to GL^+(n,\mathbb{R})$

Proof. We have: $\det(e^A) = e^{\mathrm{tr}A} > 0$.

□

35.6. **Inner product spaces and symmetric linear maps.** In this section we examine the existence of eigenvectors in a specific case of linear transformations of aspect (real) inner product space: symmetric and orthogonal maps. Both depend on the choice of inner product. See Definition 6.2. Here we take a more general approach.

Let V be a vector space over $\mathbb{R}$. A **bilinear form** is a function $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ which is linear in each coordinate. It is **positive** iff for every $\mathbf{v}$ one has that $\langle \mathbf{v}, \mathbf{v}\rangle \geq 0$, it is **positive definite** iff this equals zero only when $\mathbf{v}$ is the zero vector $\mathbf{0}$, and is **positive semidefinite** otherwise. It is a **symmetric** bilinear form iff $\langle \mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{w}, \mathbf{v}\rangle$ for all $\mathbf{v}, \mathbf{w}$.

An **inner product** on a (real) vector space V is a positive definite symmetric bilinear form. A **norm** is a real-valed function $||\cdot||$ on V satisfying:
(i) $||a\,\mathbf{v}|| = |a|\,||\mathbf{v}||$ *(homogeneity)*;
(ii) $||\mathbf{v}+\mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}||$ *(triangle inequality)*;
(iii)$||\mathbf{v}|| \geq 0$, and $||\mathbf{v}|| = 0$ iff $\mathbf{v} = \mathbf{0}$ (*positive definiteness).*

Given an inner product $\langle\cdot,\cdot\rangle$ there is a standard way to define an associated norm, setting $||\mathbf{v}|| = \langle \mathbf{v}, \mathbf{v}\rangle^{\frac{1}{2}}$.

The **standard inner product** on $\mathbb{R}^n$ is $\langle \mathbf{v}, \mathbf{v}\rangle = \sum_{i=1}^n v_i w_i$. The **Euclidean norm** is the associated norm; thus $||\mathbf{v}|| = (\sum_{i=1}^n v_i^2)^{\frac{1}{2}}$. (This is the l^2-norm; see §6.2.) There are many other possible norms, which are characterized geometrically by the shape of the unit ball about the origin. In what follows, we shall need both the Euclidean and the l^1-norm $||\mathbf{v}||_1 = \sum_{i=1}^n |v_i|$.

Lemma 35.39. *Given the standard inner product on $\mathbb{R}^n$, $\langle \mathbf{v}, \mathbf{w}\rangle = ||\mathbf{v}||||\mathbf{w}||\cos\theta$ where θ is the angle between $\mathbf{v}$ and $\mathbf{w}$.*

Proof. We first consider $\mathbb{R}^2$.

Let $\mathbf{v} = (a,b)$ and $\mathbf{w} = (c,d)$, so $\mathbf{v}\cdot\mathbf{w} = ac + bd$. Taking first the case $\mathbf{v} = (1,0)$, the equality holds. Next consider the general case, but with $||\mathbf{v}|| = ||\mathbf{w}|| = 1$. By applying a rotation matrix R, we can return to the first case. Since R is orthogonal, it preserves the inner product. This remains true for general vectors by linearity, proving the claim.

Note that since $\cos(\theta) = \cos(-\theta)$, the sign of the angle between $\mathbf{v}$ and $\mathbf{w}$ does not matter.

Next we consider $\mathbb{R}^n$. We take the plane generated by $\mathbf{v}, \mathbf{w}$, but this is isometric to $\mathbb{R}^2$, so we simply apply the first case. (Here it is important that the sign of the angle doesn't matter in the formula, since there is no natural way to define the sign of an angle if $n > 2$!)

□

Definition 35.6. On a general inner product space V, we *define* length and the modulus of the angle from the inner product by the above equation.

Remark 35.3. One of the most important things the inner product tells us is which vectors are orthogonal. Indeed, if we choose a basis and then decide those are orthonormal, this defines an inner product; see....below.

Lemma 35.40. *(Schwarz inequality)*
$|\langle \mathbf{v}, \mathbf{w} \rangle| \leq ||\mathbf{v}|| \, ||\mathbf{w}||$.

Proof. By the above definition, $|\langle \mathbf{v}, \mathbf{w} \rangle| = ||\mathbf{v}||||\mathbf{w}|| \cos\theta \leq ||\mathbf{v}|| \, ||\mathbf{w}||$. Since $||\mathbf{v}|| \, ||\mathbf{w}||$ does not depend on the definition of angle, neither does the validity of the proof. □

Definition 35.7. Two norms $||\cdot||_a$ and $||\cdot||_b$ are **equivalent** iff there exist constants $0 < c_1 < c_2$ such that for every $\mathbf{v}$, $c_1||\mathbf{v}||_b \leq ||\mathbf{v}||_a \leq c_2||\mathbf{v}||_b$ (iff they generate the same topology).

A basic fact (which we learned from Proposition 2.1 of [HS74]) is:

Lemma 35.41. *On a finite-dimensional vector space, all norms are equivalent.*

Proof. If V is a vector space of dimension one, and $||\cdot||$ is a norm, let us choose a basis vector $\mathbf{e}$ of norm one. Then for $a > 0$, $||a\mathbf{e}|| = a||\mathbf{e}||$. For another norm $||\cdot||'$, setting $c = ||\mathbf{e}||'$, then for all vectors, $||\mathbf{v}||' = c||\mathbf{v}||$. Thus the norm is unique up to constant multiples. Note that the map $a\mathbf{e} \mapsto a$ defines an isometry from V to $\mathbb{R}$ with the usual metric, and with the norm $|x|$ on $\mathbb{R}$ just the modulus (absolute value) of a number.

Now let V be a vector space of dimension n, with norm $||\cdot||$. Define $f(\mathbf{v}) = ||\mathbf{v}||$. We claim that $f : V \to [0, \infty)$ is a continuous function.

We fix a basis $(\mathbf{e}_1, \ldots \mathbf{e}_n)$ for V. We write $||\cdot||_2$ for the Euclidean norm in this basis; that is, given $\mathbf{v} = \sum_{k=1}^n v_k \mathbf{e}_k$, then $||\mathbf{v}||_2 = (\sum v_k^2)^{\frac{1}{2}}$.

We define $a_k = ||\mathbf{e}_k||$. By the triangle inequality for the norm $||\cdot||$, $||\mathbf{v}|| \leq \sum ||v_k \mathbf{e}_k|| = \sum |v_k| a_k$. For a sequence of vectors, $\mathbf{v}^{(i)} \to \mathbf{0}$ iff for all $1 \leq k \leq n$, $\mathbf{v}_k^{(i)} \to \mathbf{0}$. In this case, $||\mathbf{v}^{(i)}|| \to 0$, whence f is continuous.

Now since V is topologically isomorphic to $\mathbb{R}^n$, the unit sphere $S_1(\mathbf{0})$ is compact. The continuous image of a compact set is compact, whence the image $f(S_1)$ is $[c_1, c_2]$ for some $c_1 \geq 0$. By compactness, these values are assumed; thus, there exists $\mathbf{v}_1 \in S_1$ such that $||\mathbf{v}_1|| = c_1$. By the norm property of positive definiteness, one cannot have $c_1 = 0$ since $\mathbf{v}_1 \neq \mathbf{0}$.

Thus for all v, we have, as desired:

$$c_1||\mathbf{v}||_2 \leq ||\mathbf{v}|| \leq c_2||\mathbf{v}||_2.$$

□

It is useful to have the following two identities for a real vector space V with inner product $\langle \mathbf{v}, \mathbf{w} \rangle$:

Proposition 35.42. *(Polarization Identity)*

$$\langle \mathbf{v}, \mathbf{w} \rangle = 1/4(||\mathbf{v} + \mathbf{w}||^2 - ||\mathbf{v} - \mathbf{w}||^2)$$

Proposition 35.43. *(Parallelogram Law)*

$$||\mathbf{v} + \mathbf{w}||^2 + ||\mathbf{v} - \mathbf{w}||^2 = 2||\mathbf{v}||^2 + 2||\mathbf{w}||^2$$

Proof. Both proofs are simple calculations; the ony difference is in one we add and in one subtract the quantities:
(Polarization Identity): $||\mathbf{v} + \mathbf{w}||^2 - ||\mathbf{v} - \mathbf{w}||^2 = \langle \mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w} \rangle - \langle \mathbf{v} - \mathbf{w}, \mathbf{v} - \mathbf{w} \rangle = ||\mathbf{v}||^2 + ||\mathbf{w}||^2 + 2\langle \mathbf{v}, \mathbf{w} \rangle - \big(||\mathbf{v}||^2 + ||\mathbf{w}||^2 - 2\langle \mathbf{v}, \mathbf{w} \rangle\big) = 4\langle \mathbf{v}, \mathbf{w} \rangle$.
(Parallelogram Law)

$$||\mathbf{v} + \mathbf{w}||^2 + ||\mathbf{v} - \mathbf{w}||^2 = \langle \mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w} \rangle + \langle \mathbf{v} - \mathbf{w}, \mathbf{v} - \mathbf{w} \rangle$$
$$= ||\mathbf{v}||^2 + ||\mathbf{w}||^2 + 2\langle \mathbf{v}, \mathbf{w} \rangle + ||\mathbf{v}||^2 + ||\mathbf{w}||^2 - 2\langle \mathbf{v}, \mathbf{w} \rangle = 2||\mathbf{v}||^2 + 2||\mathbf{w}||^2.$$

□

Note that geometrically, the two quantities being squared are the lengths of the diagonals of a parallelogram with sides $\mathbf{v}$, $\mathbf{w}$. Note also that if $\mathbf{v}, \mathbf{w}$ are perpendicular, in the first the lengths are equal so we do get 0; in the second, we get Pythagoras' Theorem (which is therefore a special case).

The main interest of these laws seems to be the following converse:

Theorem 35.44. *Let V be a real vector space with norm $||\cdot||$. Then this norm comes from an inner product if and only if the Parallelogram Law holds.*

Proof. (Beginning!) The idea of the proof is simple: we *define* $\langle \mathbf{v}, \mathbf{w} \rangle$ by the Polarization Identity and then use the hypothesis to show this satisfies the axioms for an inner product.

thus we define

$$\langle \mathbf{v}, \mathbf{w} \rangle \equiv 1/4\big(||\mathbf{v} + \mathbf{w}||^2 - ||\mathbf{v} - \mathbf{w}||^2\big)$$

whence

$$4\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle \equiv ||\mathbf{u} + \mathbf{v} + \mathbf{w}||^2 - ||\mathbf{u} + \mathbf{v} - \mathbf{w}||^2$$

It is clearly commutative. We must show bilinearity.

It turns out this argument is not so easy! It is known as the Jordan-von Neumann Theorem. See [Tes09], Theorem 0.19; also see StackExchange: norms-induced-by-inner-products-and-the-parallelogram-law.

□

Given finite-dimensional spaces V, W, we recall the passage between linear transformations and matrices. Writing $\mathcal{M}_{(m \times n)}$ for the vector space of all $(m \times n)$ matrices, then a matrix $M \in \mathcal{M}_{(m \times n)}$ defines a linear map on the space of column vectors by left multiplication. That is, $\mathbf{v} \mapsto M\mathbf{v}$ sends the column vectors $\mathcal{M}_{(n \times 1)} \cong \mathbb{R}^n$ to $\mathcal{M}_{(m \times 1)} \cong \mathbb{R}^m$. (The proof that this is linear uses the distributive law for matrix multiplication!)

The converse depends on a choice of basis in each space.

Lemma 35.45. *Given a basis $\mathcal{B} = (\mathbf{v}_i)_{i=1}^n$ of V, we define a map $\Phi_{\mathcal{B}} : V \to R^n$ by sending the basis $\mathcal{B}$ to the standard basis $(\mathbf{e}_i)_{i=1}^n$ of $\mathbb{R}^n$. This extends to an isomorphism.*

Proof. By definition a basis satisfies two properties: it spans the space, and is linearly independent. Since it spans, a vector $\mathbf{v} \in V$ can be expressed as $\mathbf{v} = \sum_{i=1}^n v_i \mathbf{v}_i$; by linear independence this expression is unique. We extend the map Φ from $\mathcal{B}$ by linearity; by the uniqueness this is well-defined. Equivalently, $\Phi_{\mathcal{B}}$ is defined to send a vector to its $\mathcal{B}$-coordinates, thus $\Phi_{\mathcal{B}} : \mathbf{v} \mapsto (v_1, \ldots, v_n)$ where $\mathbf{v} = \sum_{i=1}^n v_i \mathbf{v}_i$. □

Corollary 35.46. *These are equivalent, given a finite-dimensional vector space V:*
(i) Choice of a basis with n elements;
(ii) Choice of an isomorphism $\Phi : V \to \mathbb{R}^n$. □

Proposition 35.47. *Given two vector spaces V, W of dimensions n, m, then a choice bases $\mathcal{B}, \mathcal{C}$ defines by the above correspondence a linear isomorphism from $L(V, W)$, the vector space of all linear transformations, to $\mathcal{M}_{(m\times n)}$.*

Proof. Given a linear transformation $A : V \to W$, with bases $\mathcal{B} = (\mathbf{v}_i)_{i=1}^n$, $\mathcal{C} = (\mathbf{w}_i)_{i=1}^m$ of V, W, and letting $(\mathbf{e}_i)_{i=1}^n$, $(\mathbf{f}_i)_{i=1}^m$ denote the standard bases of $\mathbb{R}^n, \mathbb{R}^m$ with $\Phi_{\mathcal{B}}, \Phi_{\mathcal{B}}$ the above maps, then we define $M \in \mathcal{M}_{(m\times n)}$ such that the following diagram commutes:

$$\begin{array}{ccc} V & \xrightarrow{A} & W \\ \downarrow{\scriptstyle \Phi_{\mathcal{B}}} & & \downarrow{\scriptstyle \Phi_{\mathcal{C}}} \\ \mathcal{M}_{(n\times 1)} & \xrightarrow{M} & \mathcal{M}_{(m\times 1)} \end{array}$$

That is, M is the linear transformation on the spaces of column vectors such that

$$\begin{array}{ccc} \mathbf{v}_i & \xmapsto{A} & A(\mathbf{v}_i) \\ \downarrow{\scriptstyle \Phi_{\mathcal{B}}} & & \downarrow{\scriptstyle \Phi_{\mathcal{C}}} \\ \mathbf{e}_i & \xmapsto{M} & M\mathbf{e}_i \end{array} .$$

Now we define a matrix $M \in \mathcal{M}_{(m\times n)}$ such that the $\mathcal{C}$-coordinates of $M\mathbf{e}_i$ is its i^{th} column. This defines a linear transformation as noted above, and is the only such map since $\mathcal{B}, \mathcal{C}$ are bases. □

Next we consider a vector space with an inner product, but no choice of basis.

Lemma 35.48. *Let V be a finite-dimensional vector space with inner product $\langle \cdot, \cdot \rangle$. Suppose that $\langle \mathbf{u}, \mathbf{w} \rangle = \langle \mathbf{u}', \mathbf{w} \rangle$ for all $\mathbf{w} \in V$. Then $\mathbf{u} = \mathbf{u}'$.*

Proof. Subtracting, $0 = \langle \mathbf{u}, \mathbf{w} \rangle - \langle \mathbf{u}', \mathbf{w} \rangle = \langle \mathbf{u} - \mathbf{u}', \mathbf{w} \rangle$ for all $\mathbf{w}$ so taking $\mathbf{w} = \mathbf{u} - \mathbf{u}'$, by positive definiteness of the inner product, $\mathbf{u} - \mathbf{u}' = \mathbf{0}$. □

Lemma 35.49.
(a)Let V be a finite-dimensional inner product space, then these are equivalent:
(i) $\mathcal{B} = (\mathbf{v}_i)_{i=1}^n$ is an orthonormal basis.
(ii) For a basis $\mathcal{B} = (\mathbf{v}_i)_{i=1}^n$, then for $\mathbf{v} = \sum_{i=1}^n v_i \mathbf{v}_i$, we have $v_i = \langle \mathbf{v}, \mathbf{v}_i \rangle$.

(iii) The inner product $\langle \cdot, \cdot \rangle$ is the pullback from $\mathbb{R}^n$ of the standard inner product by the map $\Phi_{\mathcal{B}}$.
(b) Given vector spaces V, W with inner products $\langle \cdot, \cdot \rangle_V, \langle \cdot, \cdot \rangle_W$, orthonormal bases $\mathcal{B} = (\mathbf{v}_i)_{i=1}^n$, $\mathcal{C} = (\mathbf{w}_i)_{i=1}^m$, and $A : V \to W$ linear, then the matrix entries of M in Proposition 35.47 are $M_{ij} = \langle \mathbf{w}_j, A(\mathbf{v}_i) \rangle_W = \mathbf{f}_j^t M \mathbf{e}_i$, where $(\mathbf{e}_i)_{i=1}^n$, $(\mathbf{f}_i)_{i=1}^m$ are the standard bases of $\mathbb{R}^n, \mathbb{R}^m$.

Definition 35.8. Given finite-dimensional spaces V, W, with inner products $\langle \cdot, \cdot \rangle_V, \langle \cdot, \cdot \rangle_W$, and given a linear transformation $A : V \to W$, we define a map $A^t : W \to V$ by $A^t(\mathbf{v}) = \mathbf{u}$ where $\mathbf{u}$ satisfies $\langle A^t(\mathbf{v}), \mathbf{u} \rangle_V = \langle \mathbf{v}, A(\mathbf{u}) \rangle_W$. We call A^t the **transpose** of A.

The definition makes sense, as:

Lemma 35.50. *This is a uniquely defined linear map.*

Proof. We apply Lemma 35.48, taking $\mathbf{u} = A^t(\mathbf{v})$.

□

In the case $V = W$, a linear transformation is termed an **operator**. This operator is **symmetric** iff $A = A^t$. An operator Q is **orthogonal** iff $Q^t Q = I$.

An **isometry** of two vector spaces with inner products is a linear isomorphism which takes one inner product to the other.

Via the map from linear transformations $\text{Ł}(V, W)$ to matrices $\mathcal{M}_{(m \times n)}$, the transpose map from $\text{Ł}(V, W)$ to $\text{Ł}(W, V)$ is taken to the usual transpose of a matrix, from $\mathcal{M}_{(m \times n)}$ to $\mathcal{M}_{(n \times m)}$, where $M^t_{ji} = M_{ij}$.

In particular, a symmetric operator A on V corresponds to a symmetric $(m \times m)$ matrix: such that $M_{ij} = M_{ji}$ for all i, j.

Note that given an orthonormal basis $\mathcal{B} = (\mathbf{v}_i)_{i=1}^n$, then the map $\mathbf{v} \mapsto (v_1, \ldots, v_n)$ where $v_i = \langle \mathbf{v}, \mathbf{v}_i \rangle$ defines an isometry from V to $\mathbb{R}^n$ with the standard inner product. The self-isometries of an inner product space are just the orthogonal transformations, and correspond to the **orthogonal matrice**s, those satisfying $M^t M = M M^t = I$.

Now we examine these matters more closely.

Lemma 35.51.
(i) Given finite-dimensional spaces V, W, with inner products $\langle \cdot, \cdot \rangle_V, \langle \cdot, \cdot \rangle_W$ and given a linear transformation $A : V \to W$, then $(A^t)^t = A$. We have $(AB)^t = B^t A^t$.
(ii) Given a bilinear form $\langle \cdot, \cdot \rangle$ on V, we define $\langle \cdot, \cdot \rangle_A$ by $\langle \mathbf{v}, \mathbf{w} \rangle_A = \langle \mathbf{v}, A\mathbf{w} \rangle$. This is a bilinear form. If $\langle \cdot, \cdot \rangle$ is a symmetric bilinear form and A is symmetric then $\langle \cdot, \cdot \rangle_A$ is also a symmetric bilinear form. If $\langle \cdot, \cdot \rangle$ is an inner product and if $B = A^t A$ then $\langle \cdot, \cdot \rangle_B$ is symmetric and positive semidefinite; it is positive definite (hence an inner product) iff A is invertible. A linear transformation preserves the inner product iff it is orthogonal.
(iii)Given the standard inner product $\mathbf{v}^t \mathbf{w} = \mathbf{v} \cdot \mathbf{w} = \langle \mathbf{v}, \mathbf{w} \rangle$ on $\mathbb{R}^n$, then a symmetric invertible matrix A defines a second inner product $\langle \mathbf{v}, \mathbf{w} \rangle_A = \mathbf{v}^t A \mathbf{w}$. Conversely, given some inner product $(\mathbf{v}, \mathbf{w})$, there exists a symmetric invertible A such that $(\mathbf{v}, \mathbf{w}) = \langle \mathbf{v}, \mathbf{w} \rangle_A$.

Proof. For any linear transformation $A : V \to W$, $\langle \mathbf{v}, A^t\mathbf{w}\rangle_V = \langle A^t\mathbf{w}, \mathbf{v}\rangle_V = \langle \mathbf{w}, A\mathbf{v}\rangle_W = \langle A\mathbf{v}, \mathbf{w}\rangle_W$ so indeed $(A^t)^t = A$.

$\langle (AB)^t\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{v}, AB\mathbf{w}\rangle = \langle A^t\mathbf{v}, B\mathbf{w}\rangle = \langle B^tA^t\mathbf{v}, \mathbf{w}\rangle$ whence $(AB)^t = B^tA^t$.

If $\langle \cdot, \cdot \rangle$ is a symmetric bilinear form and $A = A^t$ then $\langle \mathbf{v}, \mathbf{w}\rangle_A \equiv \langle \mathbf{v}, A\mathbf{w}\rangle = \langle A\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{w}, A\mathbf{v}\rangle \equiv \langle \mathbf{w}, \mathbf{v}\rangle_A$.

We have $(A^tA)^t = A^tA$ so $B = A^tA$ is symmetric. Now $\langle \mathbf{v}, \mathbf{v}\rangle_{A^tA} = \langle \mathbf{v}, A^tA\mathbf{v}\rangle = \langle A\mathbf{v}, A\mathbf{v}\rangle \geq 0$, and this is zero iff $A\mathbf{v} = \mathbf{0}$, so it is positive definite iff A is invertible.

A preserves the inner product iff for all $\mathbf{v}, \mathbf{w}$, $\langle \mathbf{v}, \mathbf{w}\rangle = \langle A\mathbf{v}, A\mathbf{w}\rangle \equiv \langle \mathbf{v}, \mathbf{w}\rangle_{A^tA}$. Letting $\mathbf{v}, \mathbf{w}$ be the elements of a basis, this determines A^tA uniquely, so it is the identity transformation.

The first part of (*iii*) is just (*i*) restated for matrices. For the converse, we claim that the equation $(\mathbf{v}, \mathbf{w}) = \mathbf{v}^tA\mathbf{w}$ determines A uniquely. For this, take $\mathbf{v} = \mathbf{e}_i$, $\mathbf{w} = \mathbf{e}_j$ where these are the standard basis vectors; then $A_{ij} \equiv \mathbf{e}_i^tA\mathbf{e}_j$ defines an $(n \times n)$ matrix A; by linearity, the equation $(\mathbf{v}, \mathbf{w}) = \mathbf{v}^tA\mathbf{w}$ holds for all $\mathbf{v}, \mathbf{w}$. Since $(\mathbf{v}, \mathbf{w}) = (\mathbf{w}, \mathbf{v})$ we have that $A_{ij} = A_{ji}$. If A is not invertible there exists $\mathbf{v} \neq \mathbf{0}$ such that $A\mathbf{v} = \mathbf{0}$ but then $(\mathbf{v}, \mathbf{v}) = 0$, a contradiction.

□

Definition 35.9. Given a vector space V over a field K and a linear map $T : V \to V$, then an *eigenvector* for T is a nonzero $\mathbf{v} \in V$ such that there exists $\lambda \in K$ (possibly zero) such that $T\mathbf{v} = \lambda\mathbf{v}$.

If $V = \mathbb{R}^n$, and the linear transformation is defined by an $(n \times n)$ matrix M, then since $\mathbb{R}^n$ naturally embeds in $\mathbb{C}^n$, M can be considered to act on $\mathbb{C}^n$ as well. In this case, as in Definition 16.1, we allow both eigenvectors and eigenvalues of M to be complex.

One way to find an eigenvector can be to find a maximal vector, as we now see.

Definition 35.10. We define a norm on the linear space $L(V, W)$ of all linear operators from one normed space V to another W by $||A|| = \sup_{\mathbf{v}\in V} ||A\mathbf{v}||/||\mathbf{v}|| = \sup_{||\mathbf{v}||=1} ||A\mathbf{v}||$. This is called the **operator norm**. One of its main useful properties is that it (clearly) behaves nicely under composition: $||AB|| \leq ||A||\,||B||$. This is called **submultiplicativity**.

Given a linear transformation A, we say a vector $\mathbf{v}$ is **maximal** iff $||\mathbf{v}|| = 1$ and $||A\mathbf{v}|| = ||A||$.

In a finite-dimensional space, maximal vectors exist: the unit ball is norm compact; since the function $\mathbf{v} \mapsto \langle \mathbf{v}, \mathbf{v}\rangle$ is continuous there exists some $\mathbf{v}$ in the unit sphere where $||A\mathbf{v}||$ attains its maximum value.

Lemma 35.52. *Let V, W be inner product spaces with $A : V \to W$ linear. Then:*
(i) $||A|| = ||A^t||$.
(ii) We consider the map $A^tA : V \to V$, and let $\mathbf{v}$ be a maximal vector. Then $A^tA\mathbf{v} = c\mathbf{v}$ with $c = ||A||^2$. In particular, if A is not the zero transformation, then a maximal vector $\mathbf{v}$ is an eigenvector for A^tA, with nonnegative eigenvalue $||A||^2$.

Proof. Let $\mathbf{v}$ be a maximal vector of norm one. Then $||A\mathbf{v}||^2 = (A\mathbf{v})^t(A\mathbf{v}) = \mathbf{v}^t(A^tA\mathbf{v})$. By the Schwarz inequality, $||A||^2 \leq ||\mathbf{v}||\,||A^tA\mathbf{v}|| = ||A^tA\mathbf{v}|| \leq ||A^tA|| \leq$

$||A^t||\,||A||$, using the submultiplicativity of the operator norm. Thus $||A|| \leq ||A^t||$. Symmetrically we have the reverse, completing the proof of (i).

For (ii), as in the previous proof, $||A||^2 = \langle A^tA\mathbf{v}, \mathbf{v}\rangle \leq ||A^tA|| \leq ||A^t||||A|| = ||A||^2$, using (i).

Now the Schwarz inequality is an equality iff the two vectors are multiples. Thus $A^tA\mathbf{v} = c\mathbf{v}$.

Next, $c = c\langle \mathbf{v}, \mathbf{v}\rangle = \langle c\mathbf{v}, \mathbf{v}\rangle = \langle A^tA\mathbf{v}, \mathbf{v}\rangle = \langle A\mathbf{v}, A\mathbf{v}\rangle = ||A\mathbf{v}||^2 = ||A||^2$.

Thus if A is nonzero, so is $\mathbf{v}$ and so is c, and $\mathbf{v}$ is an eigenvector for A^tA, with eigenvalue $||A||^2$.

□

Lemma 35.53. *Let A be a symmetric nonzero linear transformation on finite-dimensional V. Then:*
(i) A maximal vector $\mathbf{v}$ is an eigenvector for A^2, with eigenvalue $||A||^2$.
(ii) Either $||A||$ or -$||A||$ is an eigenvalue for the operator A. Indeed, either $\mathbf{v}$ from (i) is an eigenvector for A, with eigenvalue $||A||$, or $\mathbf{w} = (A - \lambda I)\mathbf{v}$ is an eigenvector for A, with eigenvalue -$||A||$.

Proof. From the previous lemma, $A^2\mathbf{v} = c\mathbf{v}$, where $c = ||A||^2$. This proves (i).

To prove (ii), writing I for the identity map on V, we know from (i) that $(A^2 - \lambda^2 I)\mathbf{v} = \mathbf{0}$ where $\mathbf{v}$ is our (nonzero) maximal vector and $\lambda = ||A||$. But $(A^2 - \lambda^2 I) = (A + \lambda I) \circ (A - \lambda I)$ so either $(A - \lambda I)\mathbf{v} = \mathbf{0}$ (in which case $\mathbf{v}$ is an eigenvector for A, with eigenvalue $||A||$) or, taking $\mathbf{w} = (A - \lambda I)\mathbf{v}$, we have $(A + \lambda I)\mathbf{w} = \mathbf{0}$ and $\mathbf{w}$ is an eigenvector for A, with eigenvalue -$||A||$.

□

Lemma 35.54. *For symmetric A, suppose $A\mathbf{v} = \lambda\mathbf{v}$. Then A preserves the orthogonal complement $\mathbf{v}^\perp$.*

Proof. If $\langle \mathbf{v}, \mathbf{w}\rangle = 0$ then $\langle \mathbf{v}, A\mathbf{w}\rangle = \langle A\mathbf{v}, \mathbf{w}\rangle = \langle \lambda\mathbf{v}, \mathbf{w}\rangle = 0$. □

Definition 35.11. The **rank** of a linear transformation $A : V \to W$ is the dimension of the image. We write $\mathrm{Im}(A)$ for the image, and $\mathrm{Ker}(A)$ for the kernel of this map.

Lemma 35.55. *For $A : V \to W$, we have*
(i) $dim(V) = dim\ Ker(A) + dim\ Im(A)$.
(ii) $rank(A) = rank(A^t)$.

Proof. By definition, the dimension of a vector space is the number of elements in a basis, which is well- defined (for a nice proof, see Theorem 2.14 of [Axl97]). Take a basis for the subspace $\mathrm{Im}(A)$; that pulls back to a linearly independent set in V, which generates a subspace V_0. The rest of the basis of V is a basis for $\mathrm{Ker}(A)$. This proves (i).

For (ii), we can prove this by a matrix calculation: the image is the column space, whose dimension is preserved by row reduction to echelon form. The image of the transpose is the row space. The rows containing pivots generate the row space for the matrix, while the columns containing pivots generate the column space for the echelon form. So in both cases this is the same number. □

35.7. **Diagonal form for real symmetric matrices: the real spectral theorem.**

Theorem 35.56. *(Real Spectral Theorem) Let V be a real finite-dimensional vector space with inner product. Then:*
(i)For a symmetric operator A of rank $m \leq n$ we can find a set of m orthonormal eigenvectors.
(ii) Given a symmetric matrix A of rank $m < n$, we can find a orthogonal matrix Q such that $Q^{-1}AQ = Q^tAQ = D$ with D diagonal, with k^{th} entry the eigenvalue $\lambda_k = \pm||A||_{V_k}$ and 0 for $m+1 \leq k \leq n$, for subspaces $V_1 \subseteq \dots V_k \cdots \subseteq V_m$ with V_k spanned by the first k vectors.

Proof. Suppose A is nonzero. Lemma 35.53 gives us an eigenvector $\mathbf{v}_1$, which we normalize. By Lemma 35.54 we can consider the restriction of A to $\mathbf{v}_1^\perp$ with the restricted inner product; a fortiori it is still symmetric. If it is the zero operator we stop; if not we get a second eigenvector $\mathbf{v}_2$, which is in $\mathbf{v}_1^\perp$ hence orthogonal to $\mathbf{v}_1$; by induction we continue until we have m normalized vectors $\mathbf{v}_1, \dots \mathbf{v}_m$ in nested subspaces $V = V_1, V_2 = \mathbf{v}_1^\perp, V_3 = \mathbf{v}_1^\perp \cap \mathbf{v}_2^\perp, \dots$. Reordering these so the eigenvalues $\lambda_1, \dots, \lambda_m$ have nondecreasing modulus, then each is a maximal vector on the subspace so by (ii) of Lemma 35.53, $\lambda_k = \pm||A||_{V_k}$.

For (ii), express the matrix A and $\mathbf{v}_1, \dots \mathbf{v}_m$ in the standard basis $\mathbf{e}_1, \dots \mathbf{e}_n$. Form the matrix Q with orthonormal columns, where the first m columns are the coordinates of $\mathbf{v}_1, \dots \mathbf{v}_m$ and the last $n-m$ are further vectors completing these to an orthonormal basis. Then Q is orthogonal, and $Q^{-1}AQ = D$ diagonal, with entries $\lambda_1, \dots, \lambda_m, 0 \dots, 0$. □

We now sketch what is to come. A symmetric matrix is a natural object to consider, as it generalizes the simplest matrices of all: the diagonal matrices. The depth of this analogy only becomes apparent when we consider inner products. To introduce the ideas, note that if L is an invertible linear transformation from a vector space V to a vector space W, then L transports an inner product on W to one on V via the equation $\langle \mathbf{v}, \mathbf{w} \rangle_{(V)} \equiv \langle L(\mathbf{v}), L(\mathbf{w}) \rangle_{(W)}$. Let us take for example the case of $\mathbb{R}^2$, where the standard inner product of $\mathbf{v} = (a, b)$ with $\mathbf{w} = (c, d)$ can be written as a matrix multiplication: $\mathbf{v}^t\mathbf{w} = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix}$, where we are identifying the (1×1) matrix $[\mathbf{v}^t\mathbf{w}]$ with the number $\mathbf{v}^t\mathbf{w}$. The linear map L is given by an invertible (2×2) matrix A and we have $\langle L(\mathbf{v}), L(\mathbf{w}) \rangle = \mathbf{v}^t A^t A \mathbf{w}$. So the inner product so defined is equal to the usual one if $A^tA = I$, the identity matrix, otherwise it is a new inner product, providing a new notion of angle and of norm. And the converse is also true, as we shall show: *any* inner product comes about in this way.

Orthonormal bases will play a natural role here, since an inner product defines what it is to be orthonormal, while conversely declaring a basis to be orthonormal and extending by bilinearity defines an inner product. The collection of all orthonormal bases for a given inner product is preserved by the orthogonal transformations (also called isometries).

Proposition 35.57. *Let A be a linear operator on finite-dimensional V. Then:*
(i)there exists a symmetric nonnegative map B such that $B^2 = A^tA$.

(ii)Assume that A is invertible. Then B above is positive definite. Defining $\widetilde{Q} = AB^{-1}$, then $\widetilde{Q}$ is orthogonal, and $A = \widetilde{Q}B$.
(iii)There exists $\widehat{Q}$ orthogonal and E diagonal such that $A = \widehat{Q}E$.
(iv) The image of the unit sphere in $\mathbb{R}^n$ by A is an ellipsoid.

Proof. From Lemma 35.51, A^tA is symmetric, hence by Theorem 35.56 there exists Q orthogonal such that $A^tA = Q^tDQ$ where D is diagonal. Moreover for this special case of A^tA we know the entries of D are nonnegative, using Lemma 35.52, since each eigenvector $\mathbf{v}_k$ is a maximal vector on its respective subspace V_k. Thus D has a square root E, diagonal with nonnegative entries. Then setting $B = Q^tEQ$ we have $B^2 = A^tA$. Since E is symmetric, B is as well. If A is positive definite, then the diagonal entries are positive, whence B is positive definite.

Thus $(B^{-1})^t = (B^t)^{-1} = B^{-1}$. Setting $\widetilde{Q} = AB^{-1}$, we have $\widetilde{Q}^t\widetilde{Q} = (AB^{-1})^tAB^{-1} = (B^{-1})^t(A^tA)B^{-1} = B^{-1}B^2B^{-1} = I$, so $\widetilde{Q}$ is orthogonal.

For part (ii), from the factorization $A = \widetilde{Q}B = \widetilde{Q}QE$, setting $\widehat{Q} = \widetilde{Q}Q$.

Part (iii) now follows since the image by the diagonal matrix E with entries $a_1 \geq \cdots \geq a_n > 0$ of the unit sphere is the ellipsoid $(x_1/a_1)^2 + \cdots + (x_n/a_n)^2 = 1$, and the image by $\widehat{Q}$ is a rotation of this. □

35.8. **Quadratic forms.** Given an inner product $\langle \cdot, \cdot \rangle$ on a (real) vector space V we defined above the associated Euclidean (or l^2) norm $||\mathbf{v}|| = \langle \mathbf{v}, \mathbf{v} \rangle^{\frac{1}{2}}$. Since norms are so important, it is natural to consider beginning, more generally, with any bilinear form, or equivalently, as we see in the next section, with any two-tensor (see Definition 6.2, §45). For a variety of reasons, one takes the square in the above formula, leading to the following:

Definition 35.12. We give four equivalent definitions of a **quadratic form** $Q(\mathbf{v})$ on a vector space V of dimension n:
(i) Q is a polynomial in $x_1, \ldots, x_n$ with all terms of degree two;
(ii) Beginning with a symmetric bilinear form $(\cdot, \cdot)$ on V, we define $Q(\mathbf{v}) = (\mathbf{v}, \mathbf{v})$ to be the quadratic form *associated to* the bilinear form.
(iii) There is a symmetric $(n \times n)$ matrix A such that for the bilinear form $\langle \mathbf{v}, \mathbf{w} \rangle_A = \mathbf{v}^tA\mathbf{w}$ then

$$Q(\mathbf{v}) = \langle \mathbf{v}, \mathbf{v} \rangle_A = \mathbf{v}^tA\mathbf{v}.$$

(iv) $(n = 2)$:

$$Q(x, y) = ax^2 + by^2 + cxy;$$

$(n = 3)$:

$$Q(x, y, z) = ax^2 + by^2 + cz^2 + d(xy) + e(xz) + f(yz)$$

and so on.

Lemma 35.58. *These are indeed equivalent.*

Proof. $(i) \iff (iv) \iff (iii)$:

A polynomial with all terms of degree two can be written as $Q(x_1, \ldots, x_n) = \sum_{i=1}^n a_ix_i^2 + \sum_{j=1}^n \sum_{i<j} c_{ij}x_ix_j$; we define a symmetric matrix from this by $A_{ij} = c_{ij}$

and $A_{ji} = c_{ij}$ for $i < j$, with diagonal elements $A_{ii} = a_i$. From the matrix we define $\langle \mathbf{v}, \mathbf{w} \rangle_A \equiv \mathbf{v}^t A \mathbf{w}$, which is a symmetric bilinear form. This gives the same quadratic form, $Q(\mathbf{v}) = \mathbf{v}^t A \mathbf{v}$. Conversely, a symmetric matrix defines the polynomial.

$(iii) \iff (ii)$: Choosing a basis for V and a symmetric matrix A and denoting by $\widehat{\mathbf{v}}, \widehat{\mathbf{w}}$ the coordinates in $\mathbb{R}^n$ of $\mathbf{v}, \mathbf{w}$ with respect to this basis, then $\langle \mathbf{v}, \mathbf{w} \rangle \equiv \widehat{\mathbf{v}}^t A \widehat{\mathbf{w}}$ defines a symmetric bilinear form on V, and hence a quadratic form. Given a bilinear form on V,

□

Remark 35.4. Here we are considering quadratic forms over a field, so $Q : K^n \to K$; more generally, this could be a ring, as in number theory, e.g. in the statement of Fermat's Last Theorem. For deep work in ergodic theory related to this see e.g. EinsiedlerWard2010 and references given there.

For $p + q = n$, the form Q is said to be of *signature* (p, q) iff the diagonal has p elements > 0 and $q < 0$. We are for simplicity assuming there are no 0 elements; this is termed a *nondegenerate* (bilinear, and quadratic) form.

Example 43. **(Hyperbolic Rotation 1)** Now we look more closely at the matrix of Example 78, in particular we describe the associated quadratic form.

Let $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

We wish to understand the quadratic form $\mathbf{v}^t A \mathbf{v}$; here we recall

For $A = \begin{bmatrix} a & c/2 \\ c/2 & b \end{bmatrix}$ this is $Q(x, y) = ax^2 + by^2 + cxy$, so for our case

$$Q(x, y) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 2xy.$$

Since A is symmetric, by the Spectral Theorem we can diagonalize it. The characteristic polynomial is $\begin{vmatrix} -\lambda & 1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 - 1 = (\lambda + 1)(\lambda - 1)$, which has eigenvectors $\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ for the eigenvalues $1, -1$ respectively. Defining the change-of-basis matrix B to have these columns, so $B = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, then B takes the eigenbasis for the diagonal matrix D (just the standard basis) to that for A, so we indeed have $B^{-1}AB = D = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ where $B^{-1} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$. We recognize B to be $\sqrt{2} \cdot R_{\pi/4}$ so we could normalize B to the orthogonal matrix $U = R_{\pi/4}$ guaranteed by the Spectral Theorem, giving $U^{-1}AU = D$.

The quadratic form for D is $\widehat{Q}(x, y) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x^2 - y^2$. The idea of indefinite symmetric bilinear forms is that they are associated to a more general notion of metric or distance where negative values are possible. The level curves of the quadratic form still correspond to "circles" as they are in this sense "equidistant" from the origin. Thus the level curves of Q are $xy = c$, which are hyperbolas including the x, y

axes, while the level curves of $\widehat{Q}$ are rotated by $\pi/4$. Writing $\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$ then the change-of-variables due to U is $u = \frac{\sqrt{2}}{2}(x-y), v = \frac{\sqrt{2}}{2}(x+y)$; the asymptotes of the $\widehat{Q}$-hyperbolas are the lines $y = \pm x$.

We return to this example after we have considered the conjugate linear hyperbolic flows e^{tD} and e^{tA}: the hyperbolas are flow orbits and the time parameter gives the hyperbolic distance. Although these are hyperbolic flows for the Euclidean metric, they preserve the indefinite (Lorenz) metric and along each orbit preserve the hyperbolic length. See Example 52.

35.9. **Complex (Hermitian) inner product.** To define a complex vector space, the initial axioms are identical to those for real spaces. The definition for norm is the same as well. The first real difference is noted when introducing inner products, as one needs to introduce the new notion of an *Hermitian inner product*. All axioms are as above except now commutativity (or *symmetry*) (1) is exchanged for:
$(\overline{1})$ $\mathbf{v} \cdot \mathbf{w} = \overline{\mathbf{w} \cdot \mathbf{v}}$ (*conjugate-symmetry*). This implies that $(2a, b)$ are replaced by:
$(\overline{2a})$ $(a\mathbf{v}) \cdot \mathbf{w} = a(\mathbf{v} \cdot \mathbf{w})$ (just like $(2a)$) but now
$(\overline{2b})$ $\mathbf{v} \cdot (a\mathbf{w}) = \overline{a}(\mathbf{v} \cdot \mathbf{w})$.

We note that the Hermitian definition reduces to the real one when the field is $\mathbb{R}$.

We see from $(\overline{1})$ that $\mathbf{v} \cdot \mathbf{v} \in \mathbb{R}$; from $(4a, b)$ we have as before that we have a norm, with $||\mathbf{v}||^2 = \mathbf{v} \cdot \mathbf{v}$.

Motivation for conjugate-symmetry comes from $\mathbb{C}^n$, where the *standard Hermitian inner product* is this: given $\mathbf{v} = (v_1, \cdots, v_n)$ and $\mathbf{w} = (w_1, \cdots, w_n)$ one sets

$$\mathbf{v} \cdot \mathbf{w} = \sum v_i \overline{w}_i. \tag{129}$$

Note for example that for $n = 1$, then with $\mathbf{v} = z \in \mathbb{C}$ we have $||\mathbf{v}|| = (z\overline{z})^{\frac{1}{2}} = |z|$, agreeing with our usual notion of size of a complex number being the modulus.

For general n, the norm is

$$||\mathbf{v}|| = \left(\sum v_i \overline{v}_i \right)^{\frac{1}{2}} = \left(\sum |v_i|^2 \right)^{\frac{1}{2}}$$

.

Remark 35.5. A key reason an inner product is a useful idea is that it allows us to define the geometrical notions not just of size, but of orthogonality and hence angle. Indeed, in a real vector space one can define the angle θ between two vectors $\mathbf{v}$ and $\mathbf{w}$ via the equation

$$\mathbf{v} \cdot \mathbf{w} = ||\mathbf{v}||||\mathbf{w}|| \cos\theta.$$

Note that since $\cos(\theta) = \cos(-\theta) = \cos(2\pi - \theta)$ this does not depend on the orientation of the plane containing $\mathbf{v}, \mathbf{w}$. In words, the magnitude of the angle from $\mathbf{v}$ to $\mathbf{w}$ equals that from $\mathbf{w}$ to $\mathbf{v}$.

We treat Hermitian inner products in the same way: we say complex vectors $\mathbf{v}, \mathbf{w}$ are *orthogonal* iff $\mathbf{v} \cdot \mathbf{w} = 0$. Trying to imitate the real case, we say the *Hermitian angle* between two vectors is defined by the equation

$$|\mathbf{v} \cdot \mathbf{w}| = ||\mathbf{v}||||\mathbf{w}|| \cos\theta$$

so $-\pi/2 \leq \theta \leq \pi/2$.

It is important to note that this is different from their angle as real vectors in $\mathbb{R}^2$. For example, the Hermitian angle between $z, w \in \mathbb{C}^1$ is 0 (as it should be for multiples) from the above equation. Indeed, $|\langle v, w\rangle| = |v\overline{w}|$ while $||v||^2||w||^2 = |v\overline{v}| \cdot |w\overline{w}| = |v\overline{w} \cdot w\overline{v}| = |v\overline{w}|^2$ so the angle is 0.

35.10. **Complex eigenvalues.**

Example 44. **(Rotation flow 2)** We recall that although a rotation matrix has no real eigenvalues or eigenvectors, this changes over the complex field: indeed the real matrix $A = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ where $a^2 + b^2 = 1$ has characteristic polynomial $p(\lambda) = \lambda^2 - (\mathrm{tr}A)\lambda + \det A = \lambda^2 - 2a\lambda + 1$ and so has eigenvalues $a \pm \sqrt{-b^2} = a \pm bi = e^{\pm i\theta}$ with eigenvectors $\begin{bmatrix} 1 \\ -i \end{bmatrix}$ and $\begin{bmatrix} 1 \\ i \end{bmatrix}$ respectively. Now since as explained above the standard Hermitian inner product of $\mathbf{v} = (v_1, \cdots, v_n)$ and $\mathbf{w} = (w_1, \cdots, w_n)$ in $\mathbb{C}^n$ is $\mathbf{v} \cdot \mathbf{w} = \sum v_i \overline{w}_i$, these vectors are perpendicular in $\mathbb{C}^2$, so normalizing them gives an orthonormal basis of eigenvectors.

We recall that an $(n \times n)$ real matrix Q is *orthogonal* iff $QQ^t = Q^tQ = I$, iff its columns are an orthonormal basis. An orthogonal matrix preserves the standard inner product of $\mathbb{R}^n$; equivalently Q is a rotation, reflection or a composition of a rotation and a reflection.

The *adjoint* of a square matrix with complex entries is the conjugate transform $M^* = (\overline{M})^t$; a *unitary* matrix satisfies $U^*U = UU^* = I$; that is equivalent to the columns (and rows) of U forming an orthonormal basis, as in the above example, which gives the unitary change-of-basis matrix

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix}$$

A matrix is *self-adjoint* iff $A = A^*$ (this reduces to being *symmetric* if it has real entries) and is *normal* iff $AA^* = A^*A$. See §35.6, §35.11 below.

The Spectral Theorem (Theorems 35.56 and 35.61) states that an orthonormal basis of eigenvectors exists for any normal matrix.

Now the rotation matrix A above is orthogonal hence a normal matrix, so this applies; we have just found such a basis explicitly. The unitary matrix U defined above then yields the diagonalization $U^*AU = D$ where $D = \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix}$ with commutative diagram

$$\begin{array}{ccc} \mathbb{C}^2 & \xrightarrow{A} & \mathbb{C}^2 \\ \uparrow{\scriptstyle U} & & \downarrow{\scriptstyle U^*} \\ \mathbb{C}^2 & \xrightarrow{D} & \mathbb{C}^2 \end{array} \tag{130}$$

See also Example 24.

Now we move beyond orthonormal bases.

Proposition 35.59. *Let A be a real (2×2) matrix with complex nonreal eigenvalue $\mu = a + bi$. Let $\mathbf{z} = (z_1, z_2) \in \mathbb{C}^2$ be an eigenvector for μ, with real and imaginary parts $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$; that is, $\mathbf{z} = \mathbf{u} + i\mathbf{v}$.*
(i) Then in the basis $(\mathbf{v}, \mathbf{u})$ for $\mathbb{R}^2$, A acts as the matrix $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$.
(ii) For A acting on $\mathbb{C}^2$, the vector $\overline{\mathbf{z}} = \mathbf{u} - i\mathbf{v}$ is an eigenvector with eigenvalue $\overline{\mu}$. The vectors $\mathbf{z}, \overline{\mathbf{z}}$ are a basis for $\mathbb{C}^2$, and in this basis the matrix A equals $D = \begin{bmatrix} \mu & 0 \\ 0 & \overline{\mu} \end{bmatrix} = |\mu| \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix}$ with $|\mu| = (a^2 + b^2)^{1/2}$.

Proof. We start with (ii). Since A has real entries,

$$A\bar{\mathbf{z}} = \overline{A}\bar{\mathbf{z}} = \overline{A\mathbf{z}} = \overline{\mu\mathbf{z}} = \bar{\mu}\bar{\mathbf{z}}.$$

Since $\mu \neq \bar{\mu}$, by Lemma 35.3 the eigenvectors $\mathbf{z}, \bar{\mathbf{z}}$ in $\mathbb{C}^2$ are linearly independent over $\mathbb{C}$. We claim that the real vectors $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^2$ are linearly independent over $\mathbb{R}$. Indeed, note that $\mathbf{u} = \frac{1}{2}(\mathbf{z}+\bar{\mathbf{z}})$ while $\mathbf{v} = \frac{-i}{2}(\mathbf{z}-\bar{\mathbf{z}})$. Now if $\mathbf{u}, \mathbf{v}$ are dependent, one is a multiple of the other, say $\mathbf{v} = a\mathbf{w}$ for some $a \in \mathbb{R}$. Then $\mathbf{u} = \frac{1}{2}(\mathbf{z} + \bar{\mathbf{z}}) = a\mathbf{v} = a\frac{-i}{2}(\mathbf{z} - \bar{\mathbf{z}})$ so $(\bar{\mathbf{z}} + \mathbf{z}) = ai(\bar{\mathbf{z}} - \mathbf{z})$, $(1 + ai)\mathbf{z} + (1 - ia)\bar{\mathbf{z}} = \mathbf{0}$, but since $\mathbf{z}, \bar{\mathbf{z}}$ are independent over $\mathbb{C}$, $ai = -1$ and also $ai = 1$, a contradiction.

Now since $\mu = a + ib$, $A(\mathbf{u} + i\mathbf{v}) = (a + ib)(\mathbf{u} + i\mathbf{v}) = (a\mathbf{u} - b\mathbf{v}) + i(b\mathbf{u} + a\mathbf{v})$, so since A has real entries, $A(\mathbf{u}) = a\mathbf{u} - b\mathbf{v}$ and $A(\mathbf{v}) = b\mathbf{u} + a\mathbf{v}$. Hence A restricted to $\mathbb{R}^2$ has with respect to the basis $(\mathbf{v}, \mathbf{u})$ the matrix $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$.

To finish the proof of (ii), we note that in the basis of eigenvectors $(\mathbf{z}, \overline{\mathbf{z}})$ the matrix A is diagonal with entries the eigenvalues $\mu, \overline{\mu}$, and that $|\mu| = (\mu\overline{\mu})^{1/2} = (a^2 + b^2)^{1/2}$. □

This example is also a special case of the following; see Theorem 3 p. 68 of [HS74] (the first edition) which is where we learned about this. The proof is the same.

Proposition 35.60. *Let A be a real $(d \times d)$ matrix with complex nonreal eigenvalue μ. Thus there is a column vector $\mathbf{z} \in \mathbb{C}^d$ with $A\mathbf{z} = \mu\mathbf{z}$. Then:*
(i)the vector $\bar{\mathbf{z}}$ has eigenvalue $\bar{\mu}$;
(ii) Let $Z_{\mathbb{C}}$ be the space spanned by $\mathbf{z}, \bar{\mathbf{z}}$. Then $Z_{\mathbb{R}} \equiv Z_{\mathbb{C}} \cap \mathbb{R}^d$ is a $2-$dimensional subspace over the reals of $\mathbb{R}^d \subseteq \mathbb{C}^d$, and defining $\mathbf{u}, \mathbf{v}, a, b$ to be the real and imaginary parts of the vector $\mathbf{z}$ and number μ, thus with $\mathbf{z} = \mathbf{u} + i\mathbf{v}$ and $\mu = a + ib$, then $(\mathbf{v}, \mathbf{u})$ is a basis for $Z_{\mathbb{R}}$, and in this basis, A acts as the matrix $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. With respect to the basis $(\mathbf{u}, \mathbf{v})$, it acts as $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$. □

Example 45. Consider the permutation matrix $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$.

This rotates the unit simplex by angle $\theta = 2\pi/3$, and that is part of the affine plane $x + y + z = 1$, so it rotates the parallel linear subspace $x + y + z = 0$ by θ. But what is the eigenvalue and eigenvector? Computing, we see $\det A - \lambda I =$

$\begin{vmatrix} -\lambda & 1 & 0 \\ 0 & -\lambda & 1 \\ 1 & 0 & -\lambda \end{vmatrix} = 1-\lambda^3 = (1-\lambda)(\lambda^2+\lambda+1)$, which has three solutions, $1, \mu = e^{2\pi i/3}$ and $\bar{\mu} = e^{-2\pi i/3}$. Solving for an eigenvector $\mathbf{v}$, we have $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ y \end{bmatrix} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \lambda \\ \lambda x \\ \lambda y \end{bmatrix}$, giving $\mathbf{z} = \begin{bmatrix} 1 \\ \lambda \\ \lambda^2 \end{bmatrix}$, which does satisfy $A\mathbf{v} = \lambda\mathbf{v}$. Writing $\mathbf{z} = \mathbf{v}$ for $\lambda = \mu$, $\bar{\mathbf{z}}$ for $\lambda = \bar{\mu}$ and $\mathbf{1}$ for $\lambda = 1$, we have three eigenvectors $\mathbf{z}, \bar{\mathbf{z}}$ and $\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. This last is the midpoint of the simplex, fixed by the rotation, and is the only nonnegative eigenvector for A. And using the formula in the Lemma, $\mathbf{u} = \mathrm{Re}(\mathbf{z}) = \begin{bmatrix} 1 \\ -1/2 \\ -1/2 \end{bmatrix} = \begin{bmatrix} 1 \\ a \\ a \end{bmatrix}$, $\mathbf{v} = \mathrm{Im}(\mathbf{z}) = \begin{bmatrix} 0 \\ \sqrt{3}/2 \\ -\sqrt{3}/2 \end{bmatrix} = \begin{bmatrix} 0 \\ b \\ -b \end{bmatrix}$, which do form an orthogonal basis for the plane $x+y+z=0$, with A acting in this basis as the matrix $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$, which is the rotation $R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.

We remark that the basis vectors $\mathbf{u}, \mathbf{v}$ being orthogonal is special: it means that A acts as a rotation in the imbedded plane with the usual metric. But that is *not* what the Lemma guarantees in general: if $\mathbf{u}, \mathbf{v}$ are *not* orthogonal then the rotation is seen only after the basis change, so the rotation is around *ellipses* rather than circles, followed by an expansion or contraction if $a^2 + b^2 \neq 1$.

That is, a complex eigenvalue of modulus one does not guarantee a rotation on a real plane, but merely an *elliptical* rotation. In retrospect this is obvious: if A is a rotation, then it has a complex eigenvalue, and any conjugate $B^{-1}AB$ has the same eigenvalues; yet unless the columns of B are orthogonal, the invariant circles have been changed into ellipses. See Theorem 35.56 below (the Spectral Theorem) for related matters.

Exercise 35.1. (1) Verify that this is the case for $A = \begin{bmatrix} 0 & -2 \\ 1 & 2 \end{bmatrix}$: find a (non-orthogonal) basis change matrix B such that $B^{-1}AB = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$. (See p. 56 of [HS74]).
(2) Work out the above for a (4×4) permutation matrix.

35.11. **The Spectral Theorem.** In Theorem 35.56 we gave a first version of the Spectral Theorem, for real symmetric matrices. Here we prove the full statement.

First we repeat Definition 35.8, but now for complex vector spaces:

Definition 35.13. Given a Hermitian inner product $\langle \cdot, \cdot \rangle$ on a complex finite-dimensional V and a linear operator $A : V \to V$, the **adjoint** of A is the linear transformation A^* on V defined by the equation $\langle \mathbf{v}, A\mathbf{w} \rangle = \langle A^*\mathbf{v}, \mathbf{w} \rangle$. (As before, this is a uniquely defined linear map).

(Equivalently, $\langle \mathbf{v}, A^*\mathbf{w} \rangle = \langle A\mathbf{v}, \mathbf{w} \rangle$.)

The operator is **self-adjoint** or **Hermitian** iff $A = A^*$. An operator Q is **unitary** (the complex analogue of orthogonal) iff $Q^*Q = I$, iff Q preserves the Hermitian inner product: for all $\mathbf{v}, \mathbf{w}$, $\langle Q\mathbf{v}, Q\mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle$. It is **normal** iff $A^*A = AA^*$.

35.12. **The Complex Spectral Theorem.**

Theorem 35.61. *(Complex Spectral Theorem)*
(i)An operator A is normal iff the space has an orthonormal basis of eigenvectors.
(ii) Given a self-adjoint matrix A of rank $m < n$, we can find a unitary matrix Q such that $Q^{-1}AQ = Q^*AQ = D$ with D diagonal, with k^{th} entry the eigenvalue $\lambda_k = \pm ||A||_{V_k}$ and 0 for $m + 1 \leq k \leq n$.

35.13. **Singular Value Decomposition.**

35.14. **Canonical forms.**

35.15. **Lie algebras and Lie groups: some examples.** Now the basic property of $\exp : \mathbb{R} \to (0, \infty) = \mathbb{R}^{>0}$ is that it takes addition to multiplication, that is, it maps the group of additive reals $(\mathbb{R}, +)$ to the multiplicative reals $(\mathbb{R} \setminus \{0\}, \cdot)$; it is onto the *positive* reals $\mathbb{R}^{>0} = (0, \infty)$, which is the largest connected component of the multiplicative reals containing $1 = e$. This unusual characterization of the positive reals is what generalizes to other situations.

A *Lie group* G is a group which is also a differentiable manifold, such that the differentiable structure is preserved by the group operations. (In other words, such that the group operations are diffeomorphisms.) The tangent space at the identity element $e \in G$ is called the *Lie algebra* $\mathfrak{g}$ of G. (The convention is to use a a small Gothic letter for the Lie algebra). The Lie algebra $\mathfrak{g}$ is an additve (i.e. commutative) group. in fact a vector space, as it is equal to the tangent space at the identity. The exponential map is defined to send the Lie algebra to the group. This is not surjective; it maps onto the connected component of the identity. Thus for G the multiplicative reals, its Lie algebra is the additive reals, and as noted, exp maps onto the positive subgroup $(0, \infty)$,

For the complexes, the exponential map is onto all of the multiplicative group $\mathbb{C} \setminus \{0\}$. For matrix groups, the image will be the *orientation-preserving* matrices: those with determinant > 0. This corresponds to e^x being positive for $x \in \mathbb{R}$.

Now as for $\mathbb{R}$ or $\mathbb{C}$, for any abelian group the exponential map takes addition to multiplication. However this is not true in general, because of the possible noncommutativity of the group.

The exponential map of a manifold.

Just to set the overall stage, given a Riemannian manifold M, then for each point $p \in M$, there is a map $\exp : T_p(M) \to M$, which sends straight lines though $\mathbf{0}$ in T_p to geodesics in M. In the particular case of Lie groups, each T_p can be identified with the tangent space at the origin, T_e. This is a Lie algebra, with a bracket operation; in the general setting the bracket may depend on the point, as the curvature may change. See [**?**].

The bracket operation and infinitesimal noncommutativity.

This noncommutativity needs to be reflected somehow in the Lie algebra. This is the role of the *Lie Bracket* operation, which represents the *infinitesimal* noncommutativity in the following sense.

Definition 35.14. An element a of the Lie algebra $\mathfrak{g}$ of G is tangent to a geodesic curve starting at the identity element e; this curve is $\exp(tA)$ where the exponential map $\exp : \mathfrak{g} \to G$ is explained shortly. Note that indeed $d/dt_{t=0} \exp(tA) = A$ if the usual formulas apply (which they do!) Now setting $A = \exp(a)$ and $B = \exp(b)$, then the noncommutativity of $g, h \in G$ is measured by $ghg^{-1}h^{-1}$. This is the *commutator* of g and h. However the more natural place to define this operation is not on G but rather on the Lie algebra $\mathfrak{g}$, where for $g = E^A$ and $h = e^B$, the infinitesimal version of this– also called the *commutator* of A and B, now denoted $[A, B]$– is

$$[a, b] = \lim_{t \to \infty} \frac{1}{t^2}(e^{tA}e^{tB}e^{-tA}e^{-tB}). \tag{131}$$

See Proposition 35.64. This will define the bracket operation on $\mathfrak{g}$. For the simplest case, if all $a, b \in \mathfrak{g}$ commute, that is, if $[a, b] = \mathbf{0}$, then $\exp(ta)$ and $\exp(sb)$ commute in G for all s, t; in other words, G is a commutative group.

We have already seen a small introduction to this subject, through our study of $SL(2, \mathbb{C})$ and $SL(2, \mathbb{R})$. For much more background see e.g. [Sam12], [War71], [SR73], [Hal15].

Each element $g \in G$ has a derivative map which takes this to the tangent space at g. One wonders how, and to what degree, the group multiplication might be reflected in the Lie algebra, and the remarkable answer is that this is given by an operation called the *Lie bracket*, which determines the group completely in the sense of its local geometry, near e. Precisely:

Here are the definitions:

Definition 35.15. A Lie bracket $[x, y]$ on a vector space V is an operation on V (a function from $V \times V$ to V) which satisfies the axioms:
– bilinearity;
–anticommutativity: $[y, x] = -[x, y]$;
–the *Jacobi identity*

$$[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = \mathbf{0}$$

The intuition behind the bracket operation is that it should capture the noncommutativity of the group, on the tangent space. We explain this below.

Here are some basic examples of Lie algebras:

Example 46. **(Vector product.)**

Proposition 35.62. *The vector product* $\mathbf{v} \wedge \mathbf{w}$ *on* $\mathbb{R}^3$ *is a Lie bracket.*

To explain this, we present four equivalent definitions of $\mathbf{v} \wedge \mathbf{w}$ (here $\mathbf{i}, \mathbf{j}, \mathbf{k}$ denote the standard basis vectors). The fourth, in Proposition 35.65, shows the vector product to be the commutator of a certain matrix algebra.
(1) (Via the "determinant" formula): This is the usual definition: that

$$\mathbf{v} \wedge \mathbf{w} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = \mathbf{i} \begin{vmatrix} v_2 & v_3 \\ w_2 & w_3 \end{vmatrix} - \mathbf{j} \begin{vmatrix} v_1 & v_3 \\ w_1 & w_3 \end{vmatrix} + \mathbf{k} \begin{vmatrix} v_1 & v_2 \\ w_1 & w_2 \end{vmatrix}$$

(2) (*The geometric definition*):
$\mathbf{v} \wedge \mathbf{w}$ satisfies the following properties:
(i) $\mathbf{z} = \mathbf{v} \wedge \mathbf{w}$ is perpendicular to $\mathbf{v}$ and to $\mathbf{w}$;
(ii) The norm of $\mathbf{z}$ is equal to the area of the parallelogram spanned by $\mathbf{v}, \mathbf{w}$; thus

$$||\mathbf{v} \wedge \mathbf{w}|| = ||\mathbf{v}||||\mathbf{w}|| \cdot |\sin(\theta)|.$$

We remark that θ is the angle *from* $\mathbf{v}$ *to* $\mathbf{w}$, where in the plane this would mean measured in the counterclockwise sense from $\mathbf{v}$ to $\mathbf{w}$; in $\mathbb{R}^3$, together with an orientation, "counterclockwise" is defined by looking down along the thumb for the right-hand rule. Note that since the modulus is taken, this is the same for the angle $-\theta$ from $\mathbf{w}$ to $\mathbf{v}$ and in any case is positive as a norm should be.
(iii) If $\mathbf{z} \neq \mathbf{0}$, then $(\mathbf{v}, \mathbf{w}, \mathbf{z})$ forms a positively oriented basis for $\mathbb{R}^3$.
(3) (*The algebraic definition*): The vector product is a bilinear operation such that $\mathbf{i} \wedge \mathbf{j} = \mathbf{k}$, $\mathbf{j} \wedge \mathbf{k} = \mathbf{i}$, $\mathbf{k} \wedge \mathbf{i} = \mathbf{j}$. (This formula is easy to remember as it follows a circular permutation.)

To prove that (1) $\implies$ (2) we note that for any vector $\mathbf{u}$,

$$\mathbf{u} \cdot (\mathbf{v} \wedge \mathbf{w}) = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} \tag{132}$$

Taking $\mathbf{u} = \mathbf{v}$ in (132) it follows that $\mathbf{v} \cdot \mathbf{z} = 0$, similarly for $\mathbf{w}$, proving (i). Recall that $\begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = \pm$ (volume of the paralellopiped spanned by $\mathbf{u}, \mathbf{v}, \mathbf{w}$), using the fact that $\det M = \det M^t$, where the sign is + iff the map preserves orientation, since the parallelogram is the image of the unit cube, and since we know the determinant gives $\pm$ (factor of change of volume).

Now taking $\mathbf{u} = \mathbf{z} = \mathbf{v} \wedge \mathbf{w}$, since this is perpendicular to the parallelogram spanned by $\mathbf{v}, \mathbf{w}$, this volume is (base area)(height)= (area of parallelogram)$\cdot\ ||\mathbf{z}||$. Then (132) gives $||\mathbf{z}||^2 = ||\mathbf{z}|| \cdot$(area) whence the area is indeed $||\mathbf{z}||$. Lastly, ...

It is clear that both $(1), (2)$ imply (3), but knowing (3) for the basis determines $\mathbf{v} \wedge \mathbf{w}$ for all $\mathbf{v}, \mathbf{w}$, by bilinearity.

Proof of Proposition 35.62: Now from (3) we have an exceptionally easy proof of the Jacobi identity, since by bilinearity it is enough to check this on the basis vectors, and for example

$$[\mathbf{i}, [\mathbf{j}, \mathbf{k}]] + [\mathbf{j}, [\mathbf{k}, \mathbf{i}]] + [\mathbf{k}, [\mathbf{i}, \mathbf{j}]] = \mathbf{0}$$

since each term is $\mathbf{0}$, and similarly for the other cases. □

Example 47. **(Commutator of matrices.)** The $(d \times d)$ matrices over a field K are a Lie algebra with the bracket of A, B defined to be the *commutator*

$$[A, B] = AB - BA.$$

As is easily checked, this satisfies the Jacobi identity. Note that the matrices A and B commute iff $[A, B] = 0$.

Definition 35.16. For $K = \mathbb{R}, \mathbb{C}$ these define $\mathfrak{g}l(d, \mathbb{R}), \mathfrak{g}l(d, \mathbb{C})$.

As we shall see, these are the Lie algebras of the general linear groups $GL(d, \mathbb{R}), GL(d, \mathbb{C})$. More generally, $\mathfrak{g}l(V)$ denotes the Lie algebra of the group $GL(V)$ of all isomorphisms of V.

Example 48. **(Space rotation group and the vector product.)** $SO(3, \mathbb{R})$ is defined to be the orthogonal matrices of determinant one, that is, the rotations of 3−space. This example will show that the two Lie algebras of Examples 46, 47 above are in fact isomorphic.

This is not so hard to see, if we are given a hint: that a vector corresponds to an infinitesimal rotation about that axis.

Now it is easy to figure this out for the standard axes, given what we already know about plane rotations.

Rotation about the z axis by angle $2\pi t$ is thus given by $R_t^k = \begin{bmatrix} a & -b & 0 \\ b & a & 0 \\ 0 & 0 & 1 \end{bmatrix} =$ $\exp(tK)$ where $K_t = \begin{bmatrix} 0 & -t & 0 \\ t & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

The definitions for rotations about the y, x axes are similar.

Then we define a map Φ from $\mathbb{R}^3$ to $SO(3, \mathbb{R})$ the vector space of all orthogonal (3×3) real matrices $M_3(\mathbb{R})$ by sending the standard basis elements $\mathbf{i}, \mathbf{j}, \mathbf{k}$ to these matrices $\mathbf{i} \mapsto I_1$, $\mathbf{j} \mapsto J_1$, $\mathbf{k} \mapsto K_1$ and extending linearly.

Thus

$$\Phi : \begin{bmatrix} a \\ b \\ c \end{bmatrix} \mapsto \begin{bmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{bmatrix}.$$

Let $A\mathbb{R} \to \mathcal{M}_n(\mathbb{R})$. Thus $A(t)$ is a curve in $\mathcal{M}_n$ which can be identified with $\mathbb{R}^{n^2}$. Therefore we write $A'(t)$ for the tamgent vector to this path, which is a matrix. This leads to:

Lemma 35.63. *(product rule for vector and matrix-valued curves)*

(i) Let V be a finite-dimensional inner product space, and let $\gamma, \eta : \mathbb{R} \to V$ be differentiable curves in V. Then

$$(\gamma \cdot \eta)' = \gamma' \cdot \eta + \gamma \cdot \eta'$$

that is, for all t,

$$(\gamma \cdot \eta)'(t) = \gamma'(t) \cdot \eta(t) + \gamma(t) \cdot \eta'(t).$$

(ii) Let $A(t), B(t)$ be differentiable curves in $\mathcal{M}_n$.

Then $(AB)' = A'B + AB'$. The same holds for rectangular matrices, $A \in \mathcal{M}_{n\times k}$, $B \in \mathcal{M}_{k\times m}$.

Proof. We prove (i) for $V = \mathbb{R}^n$ by writing the curve in coordinates, and applying the usual product rule in one dimension multiple times.

We prove (ii) from this: the ij^{th} entry of the matrix $(AB)'$ is the inner product of the i^{th} row of A with the j^{th} column of B. Applying (i) to each entry $(AB)'_{ij}$ gives the result. □

Proposition 35.64. *Here $[A, B] = AB - BA$.*
(i)$(e^{tA}Be^{-tA})'(0) = [A, B]$
(ii)

$$\frac{d}{dt}|_{t=0}(\frac{d}{ds}|_{s=0}(e^{tA}e^{sB}e^{-tA}) = [A, B]$$

(iii)

$$\lim_{t\to\infty} \frac{1}{t^2}(e^{tA}e^{tB}e^{-tA}e^{-tB}) = [A, B]$$

Proof. (i) From the product rule, this holds, since

$$\frac{d}{\mathrm{d}t}|_{t=0}(e^{tA}Be^{-tA}) = Ae^{0A}B - Ae^{-0A} = AB - BA.$$

(ii) Taking the limit $s \to 0$ gives us (i). (iii) follows from this. See [Sal] Lemma 1.4. □

Remark 35.6. Note that part (iii) can be pictured as a quadrilateral of geodesic curves which don't quite meet, because of the noncommutativity; the bracket is the infinitesimal version of this.

Many fascinating examples and results are surveyed in the lecture notes of John Baez and of Dietmar Salamon. [Sal]. See p.56 ff. of Hall [Hal15].

One calculates that:

Proposition 35.65. *Φ takes the vector product to the commutator of matrices, $[A, B] = AB - BA$. This is, moreover, the bracket as defined in* (131).

Proof. We noted above that $\mathbf{i}\wedge\mathbf{j} = \mathbf{k}, \mathbf{j}\wedge\mathbf{k} = \mathbf{i}, \mathbf{k}\wedge\mathbf{i} = \mathbf{j}$, and we calculate that indeed I_1, J_1, K_1 satisfy these same identities: for

$$I_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, J_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, K_1 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

then

$$[I_1, J_1] = K_1, [J_1, K_1] = I_1, [K_1, I_1] = J_1$$

For the last statement see Proposition 35.64. □

Remark 35.7. In summary, the Lie algebra of the rotation group is identified with the Lie algebra of linear vector fields on $\mathbb{R}^n$ with bracket operation defined by Definition 35.14; for $n = 3$ this is isomorphic to $\mathbb{R}^3$ with the vector product.

This makes some intuitive sense, since applying a force orthogonal to a spinning object resuts in the axis moving in a direction which is orthogonal to both the spin axis and the force vector, via the right-hand rule!

Also, this explains a connection to electromagnetism, where the vector product also appears, as the magnetic field can be thought of as arising from the spin of a charged object.

We mention that a rotating body in $\mathbb{R}^3$ intuitively always rotates about an *axis*. In fact this is true in any odd dimension ≥ 3. An explanation is that the characteristic polynomial of an orthogonal, orientation-preserving matrix in odd dimensions always has a real root, giving a real eigendirection; this is the axis. (A fuller explanation should involve the conservation of angular momentum...)

Now given a Lie algebra, with a chosen basis $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$, we can collect the coefficients of the possible brackets $[\mathbf{v}_i, \mathbf{v}_j]$ in terms of this basis. These define the *structure constants* of the Lie algebra (in terms of the basis). Note that by bilinearity, these determine the all brackets $[\mathbf{v}, \mathbf{w}]$, and so the Lie algebra, up to isomorphism. The proof just given is a simple example of showing two Lie algebras are isomorphic in this way!

Note that since the Lie algebra determines the the component containing $\mathbf{e}$ of the Lie group, this means it determines the group locally.

This is because that component is either the full Lie group, or is a subgroup of index 2 of the full group:

(TO DO...)

Theorem 35.66. *The Lie algebra of the orthogonal and special orthogonal groups* $O(n) = \{M \in \mathcal{M}_n(\mathbb{R}) : MM^t = M^tM = I\}$ *and* $SO(n) = \{M \in O(n) : det(M > 0\}$ *is the vector space of skew-symmetric matrices,* $\{A : A+A^t = 0I\}$. *SO is a subrgoup of index two of O; O has two connected components and SO is the connected component containing the identity I.*

Remark 35.8. Flanders [Fla63] pp. 36,37 has a nice elementary approach (without the exponential map) to explaining why skew-symmetric matrices are "infinitesimal rotations".

36. Minicourse on Vector Calculus

36.1. **Review of vector calculus: derivatives and the Chain Rule.** Note to students: We will deal a lot with vector fields; see the online text
https://activecalculus.org/vector/
for some nice computer graphics. See also later in these notes for pictures of some interesting vector fields.

Definition 36.1. Let V, W be Banach spaces (a vector space, possibly infinite-dimensional, on which we have a norm defined; not much is lost by restricting to $\mathbb{R}^n$ with the usual norm derived from the standard inner product). A function (or **map**) $F : V \to W$ is **differentiable** at the point $\mathbf{p} \in V$ iff there exists a linear transformation $L : V \to W$ such that

$$\lim_{\mathbf{h}\to\mathbf{0}} ||F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x}+\mathbf{h}) - L(\mathbf{h})||/||\mathbf{h}|| = 0.$$

We then write $DF_{\mathbf{p}} = L$ and call this the *derivative* of F at $\mathbf{p}$.

Let us relate this to the usual definition for $f : \mathbb{R} \to \mathbb{R}$. Then

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = c$$

We would like to write the same formula for vectors, but the problem is that *we cannot divide vectors.* Or nearly: consider the following: given a linear map $L : V \to V$, so $L\mathbf{v} = \mathbf{w}$, then in some sense

$$\frac{\mathbf{w}}{\mathbf{v}} = L :$$

the ratio "should be" a linear transformation!!

However L is not well-defined by this: many linear maps will solve the equation $L\mathbf{v} = \mathbf{w}$. (The equation only defines L on the one-dimensional subspace generated by $\mathbf{v}$). It is only well-defined if we have this information for a generating set of vectors $\mathbf{v}$. Nevertheless, this explains the intuition behind this definition.

We can rewrite the above equation as: for each $\varepsilon > 0$, there exists $\delta > 0$ such that for $|h| < \delta$,

$$\left| \frac{f(x+h) - f(x)}{h} - c \right| < \varepsilon$$

or

$$\frac{\left| f(x+h) - f(x) - ch \right|}{|h|} < \varepsilon$$

which is a special case of the general formula.

So the idea is that the derivative DF gives the "best linear approximation" at each point.

What this means is the best first-order approximation. The best 0^{th}-order approximation at $\mathbf{x} \in \mathbb{R}^n$ is simply the value of the map, $\mathbf{x} \mapsto \mathbf{p} = F(\mathbf{x})$. If $L = DF|_{\mathbf{x}}$, then the best first-order approximation will be this shifted, thus the affine map $(\mathbf{x} + \mathbf{v}) \mapsto \mathbf{p} + L\mathbf{v}$.

Writing $L(V,W)$ for the collection of linear transformations from V to W, then this is a Banach space with the operator norm. Since $DF : V \to L(V,W)$, then we see that the second derivative is a linear map $D^2F_{\mathbf{x}} : V \to L(V,W)$ and thus $D^2F : V \to L(V, L(V,W))$, and so on.

If we choose a basis for $V = \mathbb{R}^m$ and $W = \mathbb{R}^n$, then L is represented by an $(n \times m)$ matrix. Then $L(\mathbb{R}^m, \mathbb{R}^n)$ can be identified with the matrices $\mathcal{M}_{nm} \sim \mathbb{R}^{mn}$, so $DF : \mathbb{R}^m \to L(\mathbb{R}^m, \mathbb{R}^n)$ can be represented as an $(m \times mn)$ matrix. In the same way the second, third derivatives are defined, with matrices of increasing size. (The only exception is when $m = 1$ or $n = 1$, the gradient or the tangent vector).

Writing $F(\mathbf{x}) = (F_1(\mathbf{x}), \ldots, F_m(\mathbf{x}))$ for $\mathbf{x} = (x_1, \ldots, x_n)$ then the ij^{th}-matrix entry is the partial derivative

$$(DF)_{ij} = \frac{\partial F_i}{\partial x_j}$$

so

$$DF|_{\mathbf{x}} = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial F_m}{\partial x_1} & \cdots & \frac{\partial F_m}{\partial x_n} \end{bmatrix}.$$

The simplest maps to study are $f : \mathbb{R}^1 \to W$ and $V \to \mathbb{R}^1$. We call the first a **curve** in V, usually written $\gamma : \mathbb{R} \to V$. A map $F : V \to \mathbb{R}^1$ we call simply a **function**. For the case of $F : \mathbb{R}^2 \to \mathbb{R}$, we visualize it is two ways, by drawing the **graph** (the subset $\{(x,y,z) : z = F(x,y)\}$) or by drawing the **level curves** of the function. The level curve of level $c \in \mathbb{R}$ is the following subset of the plane $\mathbb{R}^2$:

$$\{(x,y) : F(x,y) = c\}.$$

The derivative of a curve γ at time t is a $(m \times 1)$ matrix .

$$D\gamma|_t = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t)) \end{bmatrix}$$

For the function $F : \mathbb{R}^m \to \mathbb{R}$ the derivative is a $(1 \times n)$ matrix:

$$DF|_{\mathbf{x}} = \begin{bmatrix} \frac{\partial F}{\partial x_1} & \cdots & \frac{\partial F}{\partial x_n} \end{bmatrix}.$$

We call this *matrix notation.*

For a curve, we can define

$$\gamma'(t) = \lim_{h \to 0} \frac{\gamma(t+h) - \gamma(t)}{h}.$$

This is the *tangent vector* to γ at time t. The relationship to the derivative $D\gamma|_t$ is that $D\gamma|_t$ is a column vector with exactly the same entries. We call the tangent vector the *vector notation* where for $n = 1$, $\gamma : \mathbb{R} \to \mathbb{R}^m$ and $\gamma'(t) = (x_1'(t), \ldots, x_m'(t))$ and for $m = 1$, $F : \mathbb{R}^n \to \mathbb{R}$, then $\nabla F = (\frac{\partial F}{\partial x_1}, \ldots, \frac{\partial F}{\partial x_m})$ which is called the *gradient* of F.

These have the same entries but in vector notation they are vectors, elements of $\mathbb{R}^n$ and $\mathbb{R}^m$, while for matrix notation they are $(m \times 1)$ and $(1 \times n)$ matrices respectively.

We call these latter *column vectors* and *row vectors* to distinguish them from each other and from elements of $\mathbb{R}^d$.

One can think of the matrix of partials as consisting of lined-up column vectors (tangent vectors) or row vectors (gradients) respectively.

We have described how the derivative at a point defines a matrix of partial derivatives. The converse is:

Lemma 36.1. *A differentiable map $F : V \to W$ is continuous. For the case $F : \mathbb{R}^n \to \mathbb{R}^m$, the map is differentiable with a continuous derivative iff the partial derivatives exist and are continuous.*

Proof. See any good vector calculus text. □

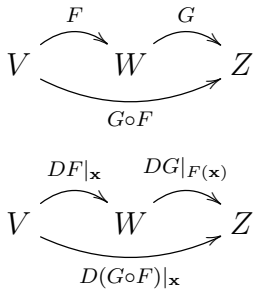
The map F is called $\mathcal{C}^0$ iff it is continuous, and $\mathcal{C}^k$ iff the k^{th} derivative exists and is continuous.

Proposition 36.2. (Chain Rule) *A composition of differentiable maps is differentiable, and the derivative is the composition of the corresponding linear maps.*

That is, for $F : V \to W$ and $G : W \to Z$ then for $G \circ F : V \to Z$ we have:

$$D(G \circ F)|_{\mathbf{p}} = DG|_{f(\mathbf{p})} \circ DF_{\mathbf{p}}.$$

Thus for the finite-dimensional case the chain rule is stated using the product of matrices.



The first example is $\gamma : \mathbb{R} \to \mathbb{R}^3$ and $F : \mathbb{R}^3 \to \mathbb{R}$, where we have in matrix notation:

$$D(F \circ \gamma(t)) = \begin{bmatrix} F_x & F_y & F_z \end{bmatrix} |_{\gamma(t)} \begin{bmatrix} x_1'(t) \\ x_2'(t) \\ x_3'(t)) \end{bmatrix}$$

and in vector notation:

$$(F \circ \gamma)'(t) = \nabla F_{\gamma(t)} \cdot \gamma'(t).$$

At this point we recall:

Proposition 36.3. *(Leibnitz' Rule for curves) Given two differentiable curves $\gamma, \eta : [a, b] \to \mathbb{R}^m$, then $(\gamma \cdot \eta)' = \gamma' \cdot \eta + \gamma \cdot \eta'$.*

Proof. We just write the curves in coordinates, and apply Leibnitz' Rule (the Product Rule) for functions from $\mathbb{R}$ to $\mathbb{R}$. □

Proposition 36.4. *Let γ be a differentiable curve in $\mathbb{R}^m$ such that $||\gamma|| = c$ for some constant c. Then $\gamma \perp \gamma'$.*

Proof. (Proof 1) First we use Leibnitz' Rule. We have $c = ||\gamma||^2 = \gamma \cdot \gamma$ so for all t,

$$0 = (\gamma \cdot \gamma)' = \gamma' \cdot \gamma + \gamma \cdot \gamma' = 2\gamma \cdot \gamma'$$

using commutativity of the inner product.
(Proof 2) We define a function $F : \mathbb{R}^m \to \mathbb{R}$ by $F(\mathbf{x}) = ||\mathbf{x}||^2$. Then since $||\gamma||$ is constant, $c = ||F \circ \gamma||$ whence by the Chain Rule,

$$0 = (F \circ \gamma)'(t) = (\nabla F(\gamma(t)) \cdot \gamma'(t)$$

but $F(\mathbf{x}) = F(x_1, \ldots, x_m) = x_1^2 + \cdots + x_m^2$ whence $\nabla F(\mathbf{x}) = 2(x_1, \ldots, x_m) = 2\mathbf{x}$. Thus $0 = 2\gamma(t) \cdot \gamma'(t)$ as claimed.

□

Note that given a differentiable curve $\gamma : [a, b] \to \mathbb{R}^m$ then γ' is a second curve, whence we can define the higher derivatives $\gamma'' = (\gamma')'$ and so on, if they exist. The most common interpretation comes from physics (which is of course why we have chosen t for the variable in the case of a curve!)

Corollary 36.5. *If $\gamma : [a, b] \to \mathbb{R}^m$ is twice differentiable then if $||\gamma'||$ is constant, we have $\gamma' \perp \gamma''$.*

Proof. We just apply the Proposition to the curve γ'. □

Definition 36.2. If $\gamma(t)$ represents the position of a particle at time t, then the *velocity* of the particle is $\mathbf{v}(t) = \gamma'(t)$, and the *acceleration* is $\mathbf{a}(t) = \mathbf{v}'(t) = \gamma''(t)$. The *speed* is the scalar quantity $||\mathbf{v}||$ (we do not have a special notation for this!).

Corollary 36.6. *If $\gamma : [a, b] \to \mathbb{R}^m$ is twice differentiable and represents the position of a particle at time t, then if the speed $||\gamma'||$ is constant, the acceleration is perpendicular to the curve (i.e. $\mathbf{a} \perp \mathbf{v}$).*

In other words if you are driving a car at a constant speed around a track, the only acceleration you will feel is side-to-side.

If we reparametrize a curve to have speed 1, then the magnitude of the acceleration vector can be used to measure how much it curves: See Definition 36.4 below.

A **vector field** on V is a map $F : V \to V$. We visualize F by drawing the vector $F(\mathbf{x}) = \mathbf{v}_\mathbf{x}$ at location $\mathbf{x}$.

A **domain** is an open subset of $\mathbb{R}^n$. A vector field on a domain $\mathcal{U}$ is simply a such a map defined only on the subset $\mathcal{U}$. The vector field is termed $\mathcal{C}^k$, for $k \geq 0$, iff the map has those properties (again, $\mathcal{C}^0$ means continuous, and $\mathcal{C}^k$ that $D^k F$ exists and is continuous, so $D : \mathcal{C}^{k+1} \to \mathcal{C}^k$).

The simplest maps to study are $f : \mathbb{R}^1 \to W$, $V \to \mathbb{R}^1$, which we have already encountered, and $F : V \to V$. This last is a **vector field** on V. We visualize this last by drawing the vector $F(\mathbf{x}) = \mathbf{v}_\mathbf{x}$ at location $\mathbf{x}$.

This is useful because to draw the graph of the function F we would need four dimensions! Note that we can use this idea for visualizing a complex function $f : \mathbb{C} \to \mathbb{C}$, as a real vector field on the plane $\mathbb{R}^2$.

We say F is a *linear vector field* exactly when $F : V \to V$ is a linear map. Thus for $\mathbb{R}^n$, fixing the standard basis, the vector field is given by a matrix. See the examples and figures in §35.5.

More generally by definition given a topological space X, a *curve* is a map $\gamma : \mathbb{R} \to X$. For values in a manifold M, then given a curve $\gamma : \mathbb{R} \to M$, then the *tangent vector* to the curve at time t is its derivative $\gamma'(t)$ is in $TM_{\gamma(t)}$, so $\gamma' : \mathbb{R} \to TM$. For the case of $M = \mathbb{R}^n$, then this can be taken in coordinates, as $\gamma'(t) = (x_1'(t), \ldots, x_n'(t))$.

We make the distinction between the tangent vector $\gamma'(t) \in \mathbb{R}^n$ and the *derivative* of γ, which takes as values a column vector with those same entries:

$$D\gamma = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t)) \end{bmatrix}$$

Thus $D\gamma : \mathbb{R} \to \mathcal{M}_{n\times 1}$. As an example of this notation we consider the Chain Rule for a curve and a function. Given $\gamma : \mathbb{R} \to \mathbb{R}^n$ and $F : \mathbb{R}^n \to \mathbb{R}$ then the Chain Rule states in matrix form:

$$D(F \circ \gamma(t)) = DF(\gamma(t))D\gamma(t)$$

or in coordinates for example for $n = 3$:

$$D(F \circ \gamma(t)) = \begin{bmatrix} F_x & F_y & F_z \end{bmatrix} |_{\gamma(t)} \begin{bmatrix} x_1'(t) \\ x_2'(t) \\ x_3'(t)) \end{bmatrix}$$

In vector form this is the dot product of the gradient vector field at the point $\gamma(t)$ and the tangent vector to the curve:

$$\mathrm{d}/\mathrm{d}t(F \circ \gamma(t)) = \nabla F(\gamma(t)) \cdot \gamma'(t) = (F_x x' + F_y y' + F_z z')(t),$$

so this number is the same as the entry of the (1×1) matrix above.

36.2. **Flows, velocity vector fields, and differential equations.** At this point, before continuing with Vector Calculus per se, we want to see some illustrative examples of vector fields, and to make the connection with ODEs and with flows. The examples here are principally linear vector fields, with two nonlinear examples at the end.

A *flow* on a space X (for example, on $\mathbb{R}^m$) is a collection of maps $\{\tau_t : t \in \mathbb{R}\}$ satisfying the *flow property* $\tau_{t+s} = \tau_t \circ \tau_s$. A flow is also called a *continuous-time dynamical system*. The space X on which the flow acts is termed *phase space*. The flow describes the time evolution of the system.

A flow defines a collection of curves on X: choosing an initial point $\mathbf{p}$, then $\gamma_{\mathbf{p}}(t) = \tau_t(\mathbf{p})$ is called the *orbit* of the point $\mathbf{p}$.

We define a vector field *tangent to the flow*:

$$F(\mathbf{p}) = \gamma_{\mathbf{p}}'(t).$$

This is a *vector ordinary differential equation* (a vector ODE). This equation links the flow to the vector field.

Say $X = \mathbb{R}^m$, then this is equivalently a *system* of m one-dimensional ODEs.

This *Fundamental Theorem of ODEs* states that one can go in the opposite direction: given a vector field (which is sufficiently smooth: e.g. it is $\mathcal{C}^2$) then given the starting point $\mathbf{p}$ (called the *initial condition*) there exists a unique curve $\gamma_{\mathbf{p}}$ which is tangent to F.

Finding this curve is called *integrating* the vector field, or *solving* the DE, and the curve is a *solution* with that initial condition.

The Fundamental Theorem is also called the theorem of *Existence and Uniqueness for ODEs*.

The actual theorem goes beyond this and says that there exists a unique differentiable flow τ_t whose orbits give these solution curves.

Summarizing, we can differentiate a flow to get a vector field, and conversely can integrate the vector field to get the flow.

A *linear flow* is a flow τ_t on a vector space such that each map τ_t is a linear transformation. In $\mathbb{R}^m$ with the standard basis this means it is given by an $(m \times m)$ matrix.

Rotation flow.

For a basic example, the *rotation flow* on the plane is defined by

$$R_t = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix}.$$

The matrix R_t acting on column vectors gives rotation of the plane by angle t, so $R_{t+s} = R_t \circ R_s$ proving that this defines a flow. See Fig. 85.

There is a beautiful relationship between a linear vector field F given by a matix A and this flow. The flow has the formula

$$\tau_t = e^{tA}.$$

Here the *exponential map* is defined for a matrix M by the power series, extending that from numbers to matrices:

$$\exp(M) = I + M + M^2/2 + \cdots + M^k/k! + \ldots$$

It is not hard to show that this always converges.

The exponential map has the property that if A and B commute, $AB = BA$, then $e^{A+B} = e^A e^B$. As a consequence, $e^{(t+s)A} = e^{tA}e^{tB}$ so this does indeed have the flow property.

For an example, taking $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ we can calculate the power series:

$$e^{tA} = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix} = R_t$$

gives the rotation flow.

Conversely, A is the infinitesimal version of this flow, since $\frac{d}{dt}e^{tA} = Ae^{tA}$ exactly as for real functions, hence at $t = 0$ this equals A.

This helps explain why the curl of a vector field measures the infinitesimal rotation, since for the linear vector field given by $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ the curl $= 2$.

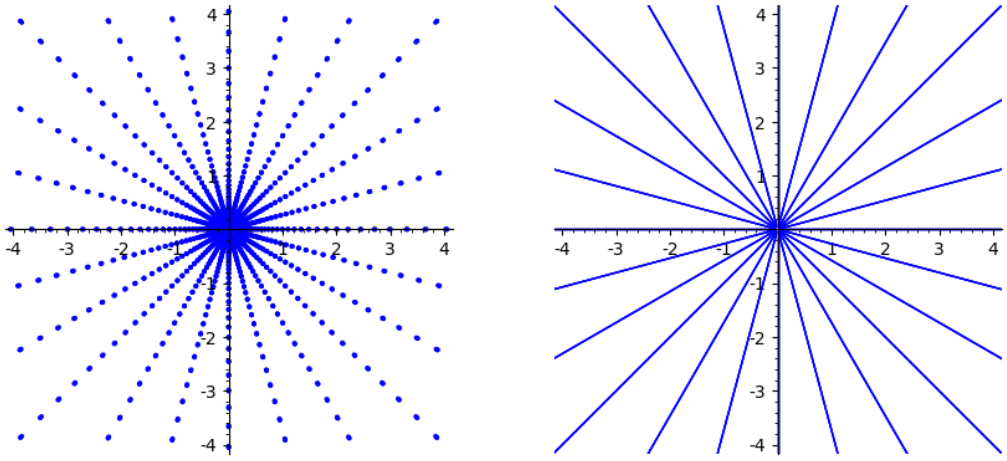


FIGURE 84. Exponential Repellor flow.

The general correspondence from matrices to their exponentials $M \mapsto e^M$ takes us from a *Lie algebra* to a *Lie group*. But that is the beginning of a much longer story!

Here we see some interesting examples of linear flow orbits defined by a matrix A, so the vector fields are tangent to these orbits. Note that in this flow interpretation the vector field is a *velocity vector field* for the flow. A completely different interpretation takes this to be a *force field*. That defines the acceleration of a particle given an inital value of the pair (position, velocity) hence defines a second-order vector DE on (position, velocity) or equivalently (position, momentum) phase space. We see two examples below, of the harmonic oscillator and the pendulum.

(1) Exponential Repellor, Fig. 84

Linear vector field defined by $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Curves are orbits of the Exponential Repellor flow. Points move away from origin exponentially fast.

(2) Rotation flow, Fig. 85

Orbits of the rotation flow, tangent to the linear vector field defined by $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

(3) Hyperbolic Flow, Fig. 86

Orbits of the Hyperbolic Flow, tangent to the linear vector field defined by $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

(4) Exponential spiral flow, Fig. 87

Orbits of the Exponential spiral flow, tangent to the linear vector field defined by $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$.

(5) Node Repellor flow, Fig. 88

Orbits of the Node Repellor flow, tangent to the linear vector field defined by $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$.

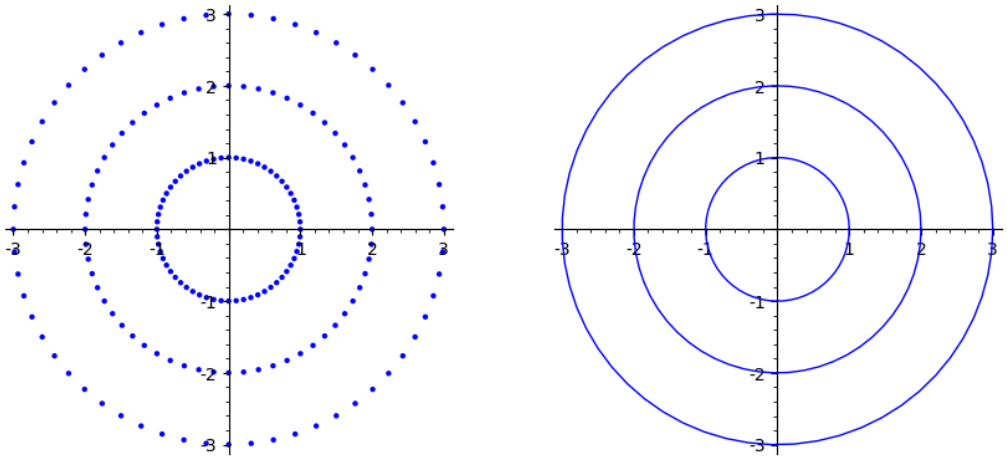


FIGURE 85. Orbits of the rotation flow.

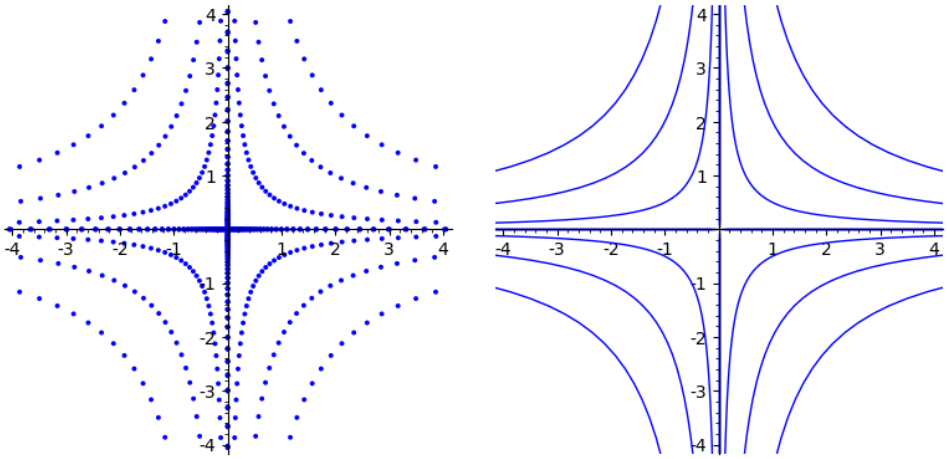


FIGURE 86. vector field for the hyperbolic flow.

(6) Vertical shear flow, Fig. 89

Orbits of the rotation flow, tangent to the linear vector field defined by $A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$.

(7) Improper Node Repellor flow, Fig. 90

Orbits of the Improper Node Repellor flow, tangent to the linear vector field defined by $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.

(8) Nonlinear Rotation flow, Fig. 91

Orbits of a Nonlinear Rotation flow, tangent to a vector field in phase space which models the pendulum.

(9) Exponential growth, Fig. 92

Orbits for a flow describing exponential growth, tangent to the linear vector field defined fronm the equation $y' = ay$.

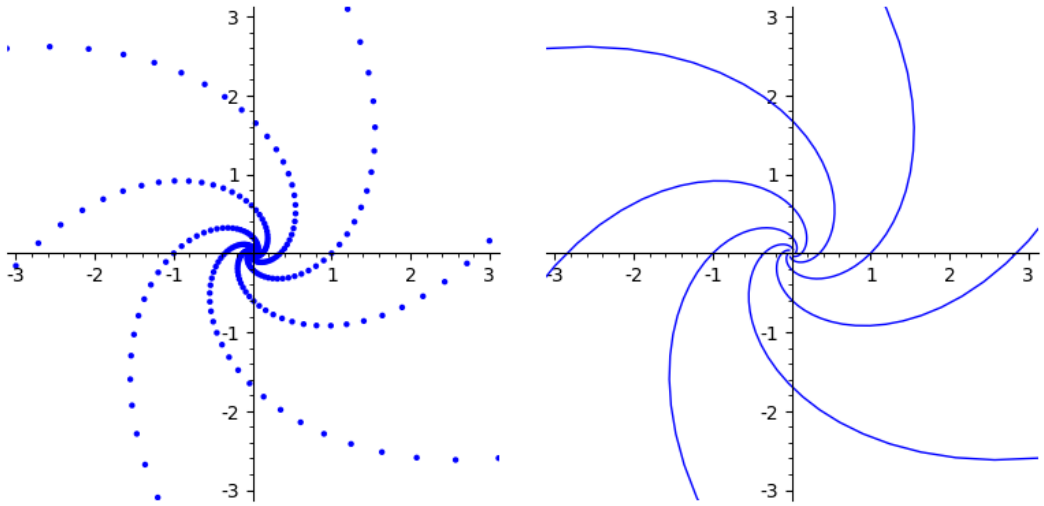


FIGURE 87. Exponential spirals.

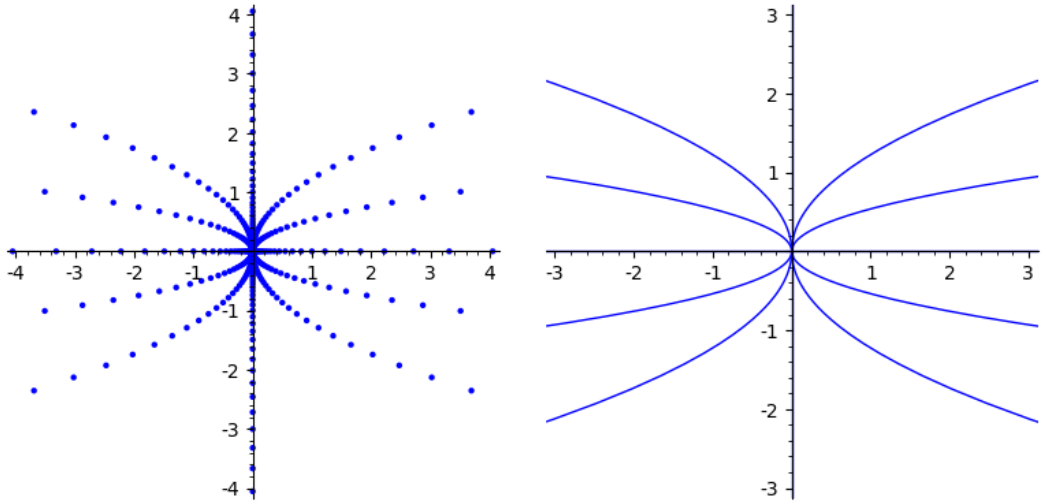


FIGURE 88. NodeRepellor.

36.3. **Review of vector calculus: the line integral.** Given a vector field F on $\mathbb{R}^n$, the *line integral* of F along γ is

$$\int_\gamma F \cdot \mathrm{d}\gamma \equiv \int_a^b F(\gamma(t)) \cdot \gamma'(t)\mathrm{d}t.$$

A line integral gives a weight at each point of the curve which depends not only on the location $\gamma(t)$ but also on the direction, $\gamma'(t)$ with respect to $F(\gamma(t))$: if these two vectors are aligned it gets a positive weight, if opposed it is negative, and if perpendicular it is zero. If for example F gives a *force field*, then the dot product measures the amount of work needed to move in that direction. Thus an ice skater

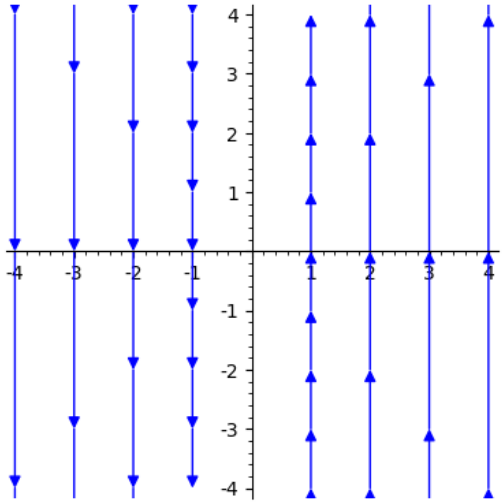


FIGURE 89. Vertical shear vector field.

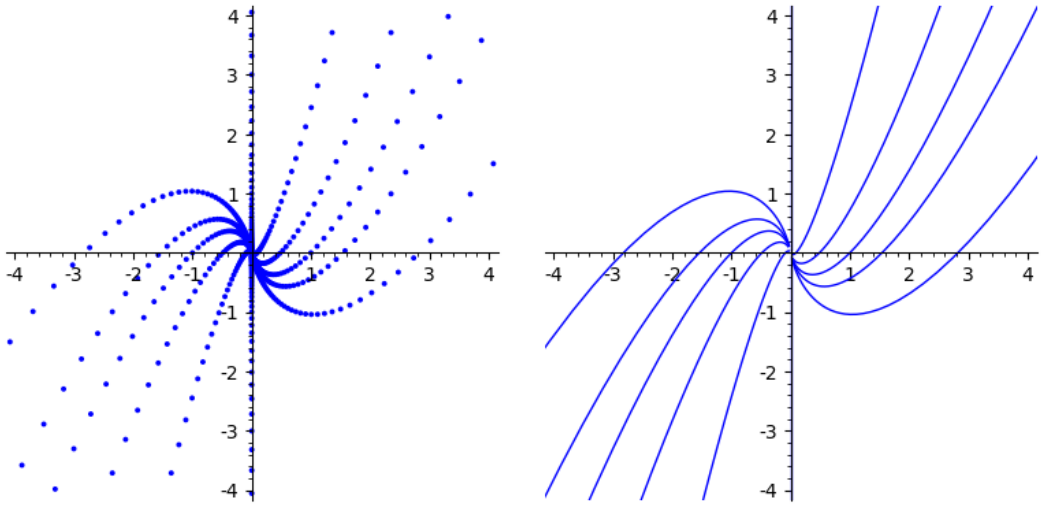


FIGURE 90. Improper Node Repellor.

glides on the ice doing no work, because the plane of the frozen lake is perpendicular to the direction of gravity.

The line integral can also be interpreted as is the integral of the curve with respect to a one-form, the one-form dual to the vector field, just as the dual space F^* is dual to F. We return to this below.

Given a curve $\gamma_1 : [a, b] \to \mathbb{R}^n$, by a *reparametrization* γ_2 of the curve we mean the following: we have an invertible differentiable function $h : [a, b] \to [c, d]$ such that $\gamma_2 = \gamma_1 \circ h$. Note that γ_1 and γ_2 have the same image, and that the tangent vectors are multiples: $\gamma_2'(t) = \gamma_1 \circ h'(t) = \gamma_1'(h(t))h'(t)$. We call this a *positive* or *orientation-preserving parameter change* if $h'(t) > 0$, *negative* or *orientation- reversing* if < 0.

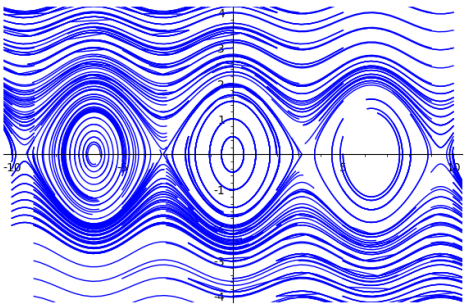


FIGURE 91. The rotation flow represents the time evolution (the dynamics) in phase space for a harmonic oscillator (a weight on a spring); the above flow is tangent to a nonlinear vector field F and models the pendulum. Here phase space has coordinates (x, y) which give the angular position and angular momentum of the pendulum. The linearization at $\mathbf{0}$ of this nonlinear vector field is the derivative $DF|_{\mathbf{0}}$ which is the matrix A of the linear vector field for the rotation flow. In the nonlinear flow, the circles flow clockwise, just as for the rotation flow. The upper curves flow to the right , the lower curves to the left; these correspond to the pendulum no longer oscillating but instead going around and around in one direction when it has high enough angular momentum.

Proposition 36.7.
(i) The line integral is unchanged for an orientation- preserving parametrization. That is,

$$\int_{\gamma_1} F \cdot d\gamma_1 = \int_{\gamma_2} F \cdot d\gamma_2.$$

(ii) For an orientation- reversing parametrization, we change the sign.

Proof. (i) Writing $u = h(t)$, we have $\gamma_2(t) = \gamma_1(h(t)) = \gamma_1(u)$. Since $\mathrm{d}u = h'(t)\mathrm{d}t$ then using the Chain Rule:

$$\begin{aligned} &\int_{\gamma_2} F \cdot \mathrm{d}\gamma_2 \equiv \int_{t=a}^{t=b} F(\gamma_2(t)) \cdot \gamma_2'(t)\mathrm{d}t = \int_{t=a}^{t=b} F(\gamma_1(h(t))) \cdot (\gamma_1 \circ h)'(t)\mathrm{d}t = \\ &\int_{t=a}^{t=b} F(\gamma_1(h(t))) \cdot \gamma_1'(h(t))h'(t)\mathrm{d}t = \int_{u=c}^{u=d} F(\gamma_1(u)) \cdot \gamma_1'(u)\mathrm{d}u = \int_{\gamma_1} F \cdot \mathrm{d}\gamma_1. \end{aligned} \tag{133}$$

(ii) For $h' < 0$, then $h(a) = d, h(b) = c$. The calculation is the same, with that change of the limits of integration:

$$\begin{aligned} &\int_{\gamma_2} F \cdot \mathrm{d}\gamma_2 \equiv \int_{t=a}^{t=b} F(\gamma_2(t)) \cdot \gamma_2'(t)\mathrm{d}t = \\ &\int_{u=d}^{u=c} F(\gamma_1(u)) \cdot \gamma_1'(u)\mathrm{d}u = -\int_{u=c}^{u=d} F(\gamma_1(u)) \cdot \gamma_1'(u)\mathrm{d}u = -\int_{\gamma_1} F \cdot \mathrm{d}\gamma_1. \end{aligned} \tag{134}$$

□

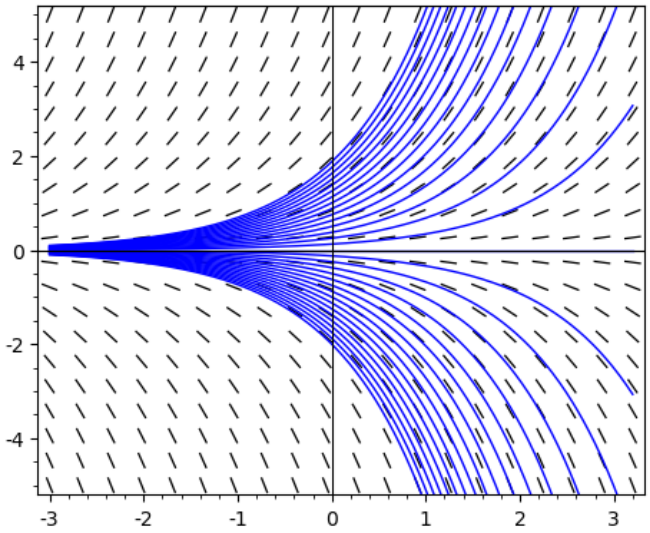


FIGURE 92. Curves tangent to nonlinear vector field $F(x, y) = (1, y)$ (shown here is not the vector field but the *slope field*, all segments of equal length and centered at the points of that slope). These are solution curves to the linear differential equation in one dimension for exponential growth or decay, $y' = ay$, in this case with $a = 1$. The solutions are $y(t) = Ke^{at}$ for $K \in \mathbb{R}$. If $a < 0$ it is exponential decay, and if $a = 0$ is constant. The graph of a solution $y(t)$ is the *image* of a solution curve for a DE in $\mathbb{R}^2$, for the curves $\gamma(t)$ tangent to the vector field. These solution curves are $\gamma(t) = (t, y(t))$ so $\gamma'(t) = (1, y'(t)) = V(\gamma(t)) = (1, y(t))$.

Corollary 36.8. *If $\gamma : [a, b] \to \mathbb{R}^n$ is a path, then writing $\widetilde{\gamma}$ for the orientation-reversed path, we have $\int_{\widetilde{\gamma}} F \cdot d\widetilde{\gamma} = -\int_{\gamma} F \cdot d\gamma$.*

Proof. Define $h : [a, b] \to [a, b]$ by $h(b() = a, h(a) = b$, interpolated linearly. Thus,

$$h(t) = -t + (a + b).$$

Then $\widetilde{\gamma}(t) \equiv \gamma \circ h(t)$. The claim follows from the Proposition.

□

There is a second notion of intergal along a curve, but where we integrate a function rather than a vector field, so there is no dot product:

Definition 36.3. Given $f : \mathbb{R}^n \to \mathbb{R}$, the *line integral of second type* of f along γ is

$$\int_{\gamma} f(\mathbf{v}) \mathrm{d}s \equiv \int_a^b f(\gamma(t))||\gamma'(t)||\mathrm{d}t.$$

Taking the special case of $f \equiv 1$, we define the *arc length* of γ to be:

$$\int_{\gamma} \mathrm{d}s = \int_a^b ||\gamma'(t)||\mathrm{d}t.$$

For an example we already know from first semester Calculus, consider a function $g : [a, b] \to \mathbb{R}$, we consider its graph $\{(x, g(x)) : a \leq x \leq b\}$. We know from Calculus the arc length of this graph is

$$\int_a^b \sqrt{1 + (g'(t))^2} \mathrm{d}x.$$

We claim that the new formula includes this one: parametrizing the graph as a curve in the plane $\gamma(t) = (t, g(t))$. Then $\gamma'(t) = (1, g'(t))$ so $||\gamma'(t)|| = \sqrt{1 + (g'(t))^2}$, whence indeed $\int_\gamma \mathrm{d}s = \int_a^b \sqrt{1 + (g'(t))^2} \mathrm{d}x$ as claimed.

Proposition 36.9.
(i) The line integral of second type of a function along a curve is unchanged for any change of parametrization, independent of orientation. That is,

$$\int_{\gamma_1} f(\mathbf{v}) ds = \int_{\gamma_2} f(\mathbf{v}) ds.$$

Proof. (i) Writing $u = h(t)$, we have $\gamma_2(u) = \gamma_2(h(t)) = \gamma_1(t)$. Then since $\mathrm{d}u = h'(t)\mathrm{d}t$, and using the Chain Rule, we have:

$$\int_{\gamma_2} f(\mathbf{v})\, \mathrm{d}s \equiv \int_{u=c}^{u=d} f(\gamma_2(u))\, ||\gamma_2'(u)||\, \mathrm{d}u = \int_{u=c}^{u=d} f(\gamma_2(h(t)))\, ||(\gamma_2'(h(t))||\, \mathrm{d}u =$$
$$\int_{t=a}^{t=b} f(\gamma_2(h(t)))\, ||\gamma_2'(h(t))||\, h'(t)\, \mathrm{d}t$$

Assuming first that $h' > 0$, this equals (135)

$$\int_{t=a}^{t=b} f(\gamma_1(t))\, ||\gamma_2'(h(t))\, h'(t)||\, \mathrm{d}t = \int_{t=a}^{t=b} f(\gamma_1(t))\, ||(\gamma_2 \circ h)'(t)||\, \mathrm{d}t$$
$$= \int_{t=a}^{t=b} f(\gamma_1(t))\, ||\gamma_1'(t)||\, \mathrm{d}t = \int_{\gamma_1} f(\mathbf{v})\, \mathrm{d}s.$$

If instead $h' < 0$, then we have as before

$$\int_{\gamma_2} f(\mathbf{v})\, \mathrm{d}s \equiv \int_{u=c}^{u=d} f(\gamma_2(u))\, ||\gamma_2'(u)||\, \mathrm{d}u = \int_{u=c}^{u=d} f(\gamma_2(h(t)))\, ||(\gamma_2'(h(t))||\, \mathrm{d}u =$$
$$\int_{t=b}^{t=a} f(\gamma_2(h(t)))\, ||\gamma_2'(h(t))||\, h'(t)\, \mathrm{d}t$$

because, since $h' < 0$, $h(b) = c, h(a) = d$.
Also we now have $||\gamma_2'(h(t))||\, h'(t) = -||\gamma_2'(h(t))\, h'(t)||$ so this is (136)

$$-\int_{t=b}^{t=a} f(\gamma_1(t))\, ||\gamma_2'(h(t))\, h'(t)||\, \mathrm{d}t = \int_{t=a}^{t=b} f(\gamma_1(t))\, ||(\gamma_2 \circ h)'(t)||\, \mathrm{d}t$$
$$= \int_{t=a}^{t=b} f(\gamma_1(t))\, ||\gamma_1'(t)||\, \mathrm{d}t = \int_{\gamma_1} f(\mathbf{v})\, \mathrm{d}s.$$

□

We next see how this can be used to give a *unit speed parametrization* of a curve $\gamma : [a, b] \to \mathbb{R}^n$. Set $l(t) = \int_a^t ||\gamma'(r)|| \, \mathrm{d}r$, so $l(t)$ is the arclength of γ from time a to time t. Note that $l'(t) = ||\gamma'(t)||$. Therefore, if $||\gamma'(t)|| > 0$ for all t, this is invertible. Our parameter change will be given by $h(t) = l^{-1}(t)$, the inverse function.

Proposition 36.10. *Assume that $||\gamma'(t)|| > 0$ for all t. Then the reparametrized curve $\widehat{\gamma} = \gamma \circ h$ has speed one.*

Proof. Now $1 = (l \circ h)'(t) = l'(h(t))h'(t)$ so $||\widehat{\gamma}'(t)|| = ||(\gamma \circ h)'(t)|| = ||(\gamma'(h(t))h'(t)|| = 1$. □

The function l maps $[a, b]$ to $[0, l(\gamma)]]$ whence the parameter-change function h maps $[0, l(\gamma)]$ to $[a, b]$. We keep t for the variable in $[a, b]$ and define $s = l(t)$, the arc length up to time t, so now s is the variable in $[0, l(\gamma)]$ and $h(s) = t$.

The change of parameter gives $\widehat{\gamma}(s) = (\gamma \circ h)(s) = \gamma(h(s) = \gamma(t)$. This indeed parametrizes the curve $\widehat{\gamma}$ is by arc length s.

Note further that

$$\int_\gamma f(\mathbf{v}) \mathrm{d}s \equiv \int_a^b f(\gamma(t))||\gamma'(t)|| \, \mathrm{d}t = \int_0^{l(b)} f(\widehat{\gamma}(s))||\widehat{\gamma}'(s)|| \mathrm{d}s \equiv \int_{\widehat{\gamma}} f(\mathbf{v}) \mathrm{d}s$$

From $s = l(t)$ we have $\mathrm{d}s = l'(t)\mathrm{d}t = ||\gamma'(t)||\mathrm{d}t$. Now we understand rigorously what is $\mathrm{d}s$: it represents the infinitesimal arc length; this helps explain the notation for this type of integral.

Level curves and parametrized curves.

There are two very distinct types of curves we encounter in Vector Calculus: the curves of this section, and the level curves of a function. Next we describe a link between the two:

Proposition 36.11. *Let $G : \mathbb{R}^2 \to \mathbb{R}$ be differentiable and suppose $\gamma : [a, b] \to \mathbb{R}^2$ is a curve which stays in a level curve of G of level c. Then $\gamma'(t)$ is perpendicular to the gradient of G.*

Proof. We have that $G(\gamma(t)) = c$ for alll t. Then by the chain rule, $D(G \circ \gamma)(t) = DG(\gamma(t)D\gamma(t)$. The derivatives here are matrices, with DG a (1×2) matrix (a row vector) and $D\gamma$ a column vector; in vector notation, these are the gradient and tangent vector, so this reads $0 = \frac{\mathrm{d}}{\mathrm{d}t}c = (G \circ \gamma)'(t) = (\nabla G)(\gamma(t)) \cdot \gamma'(t)$. □

Corollary 36.12. *If γ is a curve with $||\gamma(t)|| = c$, then $\gamma' \perp \gamma''$.*

Here is a second, direct proof; see also Corollary 36.6 above:

Proposition 36.13. *For a unit-speed curve γ, then always $\gamma' \perp \gamma''$.*

Proof. $1 = \gamma' \cdot \gamma'$ whence by Leibnitz' Rule,

$$(\gamma' \cdot \gamma')' = 2(\gamma' \cdot \gamma'') = 0.$$

□

This fact allows us to make the following

Definition 36.4. The *curvature* of a twice differentiable curve γ in $\mathbb{R}^n$ at time t is the following. For its unit-speed parametrization $\widehat{\gamma}(s)$ we define the curvature at time s to be $\widehat{\kappa}(s) = ||\widehat{\mathbf{a}}(s)||$; for γ the curvature at time t is $\kappa(t) = (\widehat{\kappa} \circ l)(t) = \kappa(t)$

For example, the curve $\gamma_r(t) = r(\cos t/r. \sin t/r)$ has velocity $\gamma_r'(t) = (-\sin t/r, \cos t/r)$ which has norm one; the acceleration is $\gamma_r''(t) = \frac{1}{r}(\cos(t/r). \sin(t/r)) = -\frac{1}{r^2}\gamma_r(t)$, with norm $\frac{1}{r}$. The curvature is therefore $\frac{1}{r}$. So if the radius of the next curve on the race track is half as much, you will feel twice the force, since by Newton's law, $F = m\mathbf{a}$! This is the physical (and geometric) meaning of the curvature. In differential geometry see p. 59 of [O'N06], For how curvature can be defined for surfaces and manifolds, see e.g. [DC16].

We have seen how a level curve $F = c$ can (sometimes) be filled in by a parametrized curve $\gamma(t)$.

This is for $f : \mathbb{R}^2 \to \mathbb{R}$. For functions on $\mathbb{R}^3$ the notion of level curve is replaced by *level surfaces*. When these can also be parametrized; the exact conditions which permit this are given by the *Implicit Function Theorem*, see §36.12 and vector calculus texts.

36.4. **Conservative vector fields.**

Definition 36.5. By a *region* we mean a connected open set. A vector field F on a region $\Omega \subseteq \mathbb{R}^n$ is *conservative* iff there exists $\varphi : \Omega \to \mathbb{R}$ such that the gradient $\nabla\varphi = F$. Such a function is called a *potential* for F.

Lemma 36.14. *If Ω is connected and φ, ψ are two potentials for F then they differ by a constant.*

Proof.

$$\frac{\partial\varphi}{\partial x} = \frac{\partial\psi}{\partial x} \implies \varphi(x,y) = \psi(x,y) + c(y); \frac{\partial\varphi}{\partial y} = \frac{\partial\psi}{\partial y} \implies \varphi(x,y) = \psi(x,y) + d(x).$$

Subtracting, $c(y) = d(x)$ so this is locally a constant, hence by connectedness is constant. □

Proposition 36.15. *If F is conservative and $\gamma : [a,b] \to \Omega$ with $A = \gamma(a), B = \gamma(b)$ then*

$$\int_\gamma F \cdot d\gamma = \varphi(A) - \varphi(B).$$

Proof.

$$\int_\gamma F \cdot \mathrm{d}\gamma \equiv \int_a^b F(\gamma(t)) \cdot \gamma'(t)\mathrm{d}t.$$

And $F(\gamma(t)) = \nabla\varphi(\gamma(t))$ so

$$F(\gamma(t)) \cdot \gamma'(t) = \nabla\varphi(\gamma(t)) \cdot \gamma'(t) = (\varphi \circ \gamma)'(t)$$

thus

$$\int_\gamma F \cdot \mathrm{d}\gamma = \int_a^b (\varphi \circ \gamma)'(t)\mathrm{d}t = \varphi \circ \gamma(t)|_a^b = (\varphi(\gamma(b) - \varphi(\gamma(a)) = \varphi(A) - \varphi(B).$$

□

Remark 36.1. This says that for conservative vector fields, we can find a potential and then evaluate a line integral in a very simple way, just as in one dimension with the Fundamental Theorem of Calculus. Both of these are special cases of Stokes' Theorem; see below in §???.

Next we review some equivalent conditions for F to be conservative.

Proposition 36.16. *The following are equivalent, for a vector field on a pathwise connected domain $\Omega \subseteq \mathbb{R}^n$:*
(i) F is conservative, i.e. there exists a potential function for F, that is, $\varphi : \Omega \to \mathbb{R}$ such that $\nabla\varphi = F$.
(ii) The line integral is path-independent.
(iii)For γ a piecewise $\mathcal{C}^1$ path which is closed *i.e. $\gamma(a) = \gamma(b)$, the line integral is* 0.

Proof. $(i) \implies (ii)$: From Proposition 36.15,

$$\int_\gamma F \cdot \mathrm{d}\gamma = \varphi(A) - \varphi(B);$$

thus this value only depends on $\varphi(A)$ and $\varphi(B)$, not on the path taken to get there. Hence if there are two paths γ_1, γ_2 with the same initial and final points A, B, then $\int_{\gamma_1} F \cdot \mathrm{d}\gamma_1 = \int_{\gamma_2} F \cdot \mathrm{d}\gamma_2$.
$(ii) \implies (iii)$: If γ is a closed path, then $\gamma(a) = A = \gamma(b) = B$. Define a second path η with the same initial and final points $A = B$ but with $\eta(t) = A$ for all t. Then $\eta'(t) = 0$ so $\int_\eta F \cdot \mathrm{d}\eta = 0$, whence by (ii) also $\int_\gamma F \cdot \mathrm{d}\gamma = 0$.

Another proof is the following: Given a closed path γ, we choose some $c \in [a, b]$ and define $C = \gamma(c)$. Write γ_1 for the path γ restricted to $[a, c]$ and γ_2 for γ restricted to $[c, b]$. Then by (ii) γ_1 and the time-reversed path $\widetilde{\gamma}_2$ have the same initial and final points, so

$$\int_{\gamma_1} F \cdot \mathrm{d}\gamma_1 = \int_{\widetilde{\gamma}_2} F \cdot \mathrm{d}\widetilde{\gamma}_2.$$

Therefore

$$\int_\gamma F \cdot \mathrm{d}\gamma = \int_a^b F(\gamma(t)) \cdot \gamma'(t)\mathrm{d}t =$$
$$\int_a^c F(\gamma(t)) \cdot \gamma'(t)\mathrm{d}t + \int_c^b F(\gamma(t)) \cdot \gamma'(t)\mathrm{d}t = \int_{\gamma_1} F \cdot \mathrm{d}\gamma_1 + \int_{\gamma_2} F \cdot \mathrm{d}\gamma_2 =$$
$$\int_{\gamma_1} F \cdot \mathrm{d}\gamma_1 - \int_{\widetilde{\gamma}_2} F \cdot \mathrm{d}\widetilde{\gamma}_2 = 0.$$

$(iii) \implies (ii)$: We essentially reverse this last argument. We are given that the integral over a closed path is 0. If there are two paths γ_1, γ_2 with the same initial and final points A, B we are to show that $\int_{\gamma_1} F \cdot \mathrm{d}\gamma_1 = \int_{\gamma_2} F \cdot \mathrm{d}\gamma_2$.

As above, we write $\widetilde{\gamma}_2$ for the time-reversed path. Then $\gamma = \gamma_1 + \widetilde{\gamma}_2$ is a closed loop, so

$$0 = \int_\gamma F \cdot \mathrm{d}\gamma = \int_{\gamma_1} F \cdot \mathrm{d}\gamma_1 + \int_{\widetilde{\gamma}_2} F \cdot \mathrm{d}\widetilde{\gamma}_2 = \int_{\gamma_1} F \cdot \mathrm{d}\gamma_1 - \int_{\gamma_2} F \cdot \mathrm{d}\gamma_2 = 0.$$

$(ii) \implies (i)$: We define a function φ by fixing some point A and choosing $\varphi(A)$ arbitrarily. Then we define the other values as follows. Letting $B \in \Omega$, since the region is path connected there exists a piecewise $\mathcal{C}^1$ path $\gamma : [a, b] \to \Omega$ with $A = \gamma(a), B = \gamma(b)$. We set

$$\varphi(B) = \int_\gamma F \cdot \mathrm{d}\gamma.$$

By (ii), this is well-defined as it does not depend on the path.

We claim that $\nabla\varphi = F$, showing the calculation for the case of $F : \mathbb{R}^2 \to \mathbb{R}$. We compute $\frac{\partial\varphi}{\partial x}$ at the point $B = (B_0, B_1)$. Defining a path η by $\eta(t) = B + t\mathbf{e}_1$, then:

$$\frac{\partial\varphi}{\partial x}|_B = \frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}\varphi(\eta(t)) = \lim_{h\to\mathbf{0}} \frac{1}{h}(\varphi(\eta(h)) - \varphi(B)) = \lim_{h\to\mathbf{0}} \frac{1}{h}(\varphi(\eta(h)) - \varphi(B) =$$

$$\lim_{h\to\mathbf{0}} \frac{1}{h}\int_\eta F \cdot \mathrm{d}\eta = \lim_{h\to\mathbf{0}} \frac{1}{h}\int_0^h F(\eta(t)) \cdot \eta'(t)\mathrm{d}t = \lim_{h\to\mathbf{0}} \frac{1}{h}\int_\mathbf{0}^h F(\eta(t)) \cdot (1, 0)\mathrm{d}t =$$

$$\lim_{h\to\mathbf{0}} \frac{1}{h}\int_0^h F(B_0, B_1) + (t, 0)) \cdot (1, 0)\mathrm{d}t = \lim_{h\to\mathbf{0}} \frac{1}{h}\int_0^h P(B_0 + t, 0)\mathrm{d}t =$$

$$\lim \frac{1}{h}\int_0^h P(B_0 + t, 0)\mathrm{d}t = P(B_0, B_1)$$

This shows that $\frac{\partial\varphi}{\partial x}|_B = P(B)$. So $\nabla\varphi = F$.

□

Next we explain where the term "conservative" comes from: from the conservation of energy in mechanics!

Suppose we have an object (a point mass) and a vector field F of forces acting on this object. This will move according to Newton's law $F = m\mathbf{a}$; here F and also the acceleration $\mathbf{a}$ are vector quantities, while the mass m is a positive scalar. If the position of the object in time is given by the curve $\gamma(t)$, then we write $\mathbf{v}(t) = \gamma(t)$ for the velocity and $\mathbf{a}(t) = \mathbf{v}'(t) = \gamma''(t)$ for the acceleration. So Newton's law states

$$F(\gamma(t)) = m\mathbf{a}(t) = m\gamma''(t).$$

Definition 36.6. *Work* is defined in mechanics to be (force) · (distance). This means that the work done by moving a particle against a force is given by that expression. The continuous-time version of this is given by a line integral.

Precisely, we define the *work done* by moving a particle along a path (a curve) γ in a force field F to be $\int_\gamma F \cdot \mathrm{d}\gamma$.

The *kinetic energy* of the particle is $\frac{1}{2}m||\mathbf{v}||^2$.

Proposition 36.17. *The work done by moving along the path γ in a force field F from time a to time b is the difference in kinetic energies, $E_{kin}(b) - E_{kin}(a)$.*

Proof. The work done by moving along the path γ from time a to time b is

$$\int_\gamma F \cdot \mathrm{d}\gamma = \int_a^b F(\gamma(t)) \cdot \gamma'(t)\mathrm{d}t = m\int_a^b \gamma''(t)) \cdot \gamma'(t)\mathrm{d}t$$

Now by Leibnitz' Rule,

$$\gamma''(t)) \cdot \gamma'(t) = \frac{1}{2}(\gamma'(t) \cdot \gamma'(t))' = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}||\mathbf{v}||^2(t)$$

so our integral is

$$\frac{1}{2}m\int_a^b \frac{\mathrm{d}}{\mathrm{d}t}||\mathbf{v}(t)||^2\mathrm{d}t = \frac{1}{2}m||\mathbf{v}(t)||^2|_a^b = \frac{1}{2}m||\mathbf{v}(b)||^2 - \frac{1}{2}m||\mathbf{v}(a))||^2 = E_{\mathrm{kin}}(b) - E_{\mathrm{kin}}(a).$$

□

This is valid for any force field, conservative or not.

Definition 36.7. Given a conservative vector field F, so with potential function φ, we define the *potential energy* of F to be $E_{\mathrm{pot}} = -\varphi$.

Note that the potential energy function of physics has the opposite sign from the potential function used in mathematics, whose gradient gives the field.

The *total energy* of a particle moving in a force field is the sum of the potential and kinetic energies, $E_{\mathrm{tot}} = E_{\mathrm{pot}} + E_{\mathrm{kin}}$. Note that the potential energy at time a depends only on the position $A = \gamma(a)$, so we write this as $E_{\mathrm{pot}}(A)$, while the kinetic energy depends on time and the path, so we write this as $E_{\mathrm{kin}}(a)$, as for the total energy $E_{\mathrm{tot}}(a)$.

Proposition 36.18. *In a conservative force field F, the work done by moving along the path γ from time a to time b is $\varphi(B) - \varphi(A) = E_{pot}(A) - E_{pot}(B)$.*

Proof. This is just Proposition 36.15 restated in the context of mechanics. □

Theorem 36.19. *If a particle moves according to Newton's law $F = m\mathbf{a}$ in a conservative force field, then the total energy is preserved: $E_{tot}(a) = E_{tot}(b)$.*

Proof. We have shown in Proposition 36.17 that the work done (in any field) is

$$\int_\gamma F \cdot \mathrm{d}\gamma = E_{\mathrm{kin}}(b) - E_{\mathrm{kin}}(a).$$

But in a conservative field, we also have a second expression for this: the work done is

$$\int_\gamma F \cdot \mathrm{d}\gamma = \varphi(B) - \varphi(A) = E_{\mathrm{pot}}(A) - E_{\mathrm{pot}}(B).$$

Thus

$$E_{\mathrm{kin}}(b) - E_{\mathrm{kin}}(a). = E_{\mathrm{pot}}(A) - E_{\mathrm{pot}}(B)$$

so

$$E_{\mathrm{tot}}(a) = E_{\mathrm{kin}}(a) + E_{\mathrm{pot}}(A) = E_{\mathrm{kin}}(b) + E_{\mathrm{pot}}(B) = E_{\mathrm{tot}}(b).$$

□

Remark 36.2. Note that we calculated the line integral $\int_a^b F(\gamma(t)) \cdot \gamma'(t)\mathrm{d}t$ in two different ways, in Proposition 36.15 and Proposition 36.17. For the first we used the existence of a potential to rewrite $F(\gamma(t))$ as $\nabla\varphi(\gamma(t))$ and use the Chain Rule; for the second we used Newton's Law to rewrite F as $m\mathbf{a} = m\gamma''$ and apply Leibnitz' Rule.

It is interesting that this are the same two very different techniques applied to give two different proofs of Corollary 36.12 above.

Equality of mixed partials. The next result can be proved using just derivatives, but we like the following "Fubini's Theorem argument", partly because it leads in to Green's Theorem later on:

Lemma 36.20. *If $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentable, then we can change the order in taking two partial derivatives: e.g. for $n = 2$, then*

$$\frac{\partial}{\partial x}\left(\frac{\partial \varphi}{\partial y} f\right) = \frac{\partial \varphi}{\partial y}\left(\frac{\partial \varphi}{\partial x} f\right).$$

Proof. Given two continuous functions $\varphi, \widetilde{\varphi} : \otimes \to \mathbb{R}$ on an open set Ω, then if for every rectangle $B \subseteq \Omega$ we have

$$\int \int_B \varphi \, \mathrm{d}x \, \mathrm{d}y = \int \int_B \widetilde{\varphi} \, \mathrm{d}x \, \mathrm{d}y,$$

then we certainly can conclude that $\varphi = \widetilde{\varphi}$ on Ω.

We define $\varphi(x, y) = \frac{\partial \varphi}{\partial x}(\frac{\partial \varphi}{\partial y} f(x, y))$ and $\widetilde{\varphi} = \frac{\partial \varphi}{\partial y}(\frac{\partial \varphi}{\partial x} f(x, y))$. Our strategy of proof will be to show that for any $B = [a, b] \times [c, d]$ we have the above equality of integrals, and the result will then follow.

Fubini's Theorem tells us that

$$\int \int_B \varphi(x, y) \, \mathrm{d}x \, \mathrm{d}y = \int_c^d \left(\int_a^b \varphi(x, y) \, \mathrm{d}x \right) \mathrm{d}y = \int_c^d \left(\int_a^b \frac{\partial}{\partial x}\left(\frac{\partial}{\partial y} f(x, y)\right) \mathrm{d}x \right) \mathrm{d}y$$

Now

$$\int_a^b \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)(x, y) \mathrm{d}x = \frac{\partial f}{\partial y}(b, y) - \frac{\partial f}{\partial y}(a, y)$$

so the iterated integral equals

$$\int_c^d \frac{\partial}{\partial y} f(b, y) \mathrm{d}y - \int_c^d \frac{\partial}{\partial y} f(a, y) \mathrm{d}y =$$
$$\Big(f(b, d) - f(b, c) \Big) - \Big(f(a, d) - f(a, c) \Big).$$

Next, by Fubini's Theorem:

$$\int \int_B \widetilde{\varphi}(x, y) \, \mathrm{d}x \, \mathrm{d}y = \int_a^b \int_c^d \widetilde{\varphi}(x, y) \, \mathrm{d}y \, \mathrm{d}x = \int_a^b \left(\int_c^d \frac{\partial}{\partial y}\left(\frac{\partial \varphi}{\partial x} f(x, y)\right) \mathrm{d}y \right) \mathrm{d}x$$

This time,

$$\int_c^d \frac{\partial}{\partial y}\left(\frac{\partial}{\partial x} f(x, y)\right) \mathrm{d}y = \frac{\partial f}{\partial x}(x, d) - \frac{\partial f}{\partial x}(x, c)$$

so the iterated integral equals

$$\int_a^b \frac{\partial}{\partial x} f(x, d) \mathrm{d}x - \int_a^b \frac{\partial}{\partial x} f(x, c) \mathrm{d}x =$$
$$\Big(f(b, d) - f(a, d) \Big) - \Big(f(b, c) - f(a, c) \Big)$$

Which equals the previous expression, finishing the proof.

□

Definition 36.8. The *curl* of a vector field $F = (P, Q)$ on $\mathbb{R}^2$ is $\operatorname{curl}(F) = (\frac{\partial}{\partial x}Q - \frac{\partial}{\partial y}P)\mathbf{k}$. The curl of a vector field $F = (P, Q, R)$ on $\mathbb{R}^3$ is

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{vmatrix} = \begin{vmatrix} \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ Q & R \end{vmatrix} \mathbf{i} - \begin{vmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial z} \\ P & R \end{vmatrix} \mathbf{j} + \begin{vmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} \\ P & Q \end{vmatrix} \mathbf{k} = \big(R_y - Q_z, P_z - R_x, Q_x - P_y\big).$$

This can also be written as a vector product, since

$$\mathbf{v} \wedge \mathbf{w} = \mathbf{v} \times \mathbf{w} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix},$$

see Example 46. So one writes

$$\operatorname{curl}(F) = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) \wedge (P, Q, R),$$

which is often abbreviated as

$$\operatorname{curl}(F) = \nabla \wedge F = \nabla \times F.$$

Note that to define the curl of a vector field in $\mathbb{R}^2$, we have to understand that $\mathbb{R}^2$ is identified with the $x - y$ plane embedded in $\mathbb{R}^3$, with the the curl a vector in $\mathbb{R}^3$ which is perpendicular to this embedded plane.

Remark 36.3. Note that these formulas represent the determinant of a matrix of symbols rather than numbers, so only make sense as formulas. Nevertheless some of the properties carry over from the usual situation of a matrix of numbers. For example, multilinearity of the determinant or linearity of the vector product is reflected in linearity of the curl: given two vector fields on $\mathbb{R}^3$, F, G then $\operatorname{curl}(\alpha F + \beta G) = \alpha \operatorname{curl}(F) + \beta \operatorname{curl}(G)$.

The formulas for $\mathbb{R}^2$ and $\mathbb{R}^3$ are connected. To understand this, take the fields $F = (P, Q)$ and $\widehat{F} = (\widehat{P}, \widehat{Q}, \widehat{R})$ with $\widehat{R} \equiv 0$ and with $\widehat{P}(x, y, z) = P(x, y)$ $\widehat{Q}(x, y, z) = Q(x, y)$ whence $\widehat{Q}_z = \widehat{P}_z = 0$ so then $\operatorname{curl}(\widehat{F}) = \big(\widehat{R}_y - \widehat{Q}_z, \widehat{P}_z - \widehat{R}_x, \widehat{Q}_x - \widehat{P}_y\big) = \big(-\widehat{Q}_z, \widehat{P}_z, \widehat{Q}_x - \widehat{P}_y\big) = \big(0, 0, \widehat{Q}_x - \widehat{P}_y\big) = (\widehat{Q}_x - \widehat{P}_y)\mathbf{k}$.

In other words, $\operatorname{curl}(\widehat{F}) = \operatorname{curl}(F)$ in this case.

Proposition 36.21. *If a field F on $\mathbb{R}^2$ is conservative, then the curl is $\mathbf{0}$.*

Proof. This follows immediately from the equality of mixed partials, Lemma 36.20. □

In fact, the curl in $\mathbb{R}^3$ can be understood with the help of that in $\mathbb{R}^2$: if $\widehat{F}$ is constant in some other direction $\mathbf{v}$, then the curl Theis a vector in that direction, giving the curl on the plane perpendicular to $\mathbf{v}$.

This will always be the case for a linear vector field. If $\widehat{F}$ is not linear, then if we take its derivative $\widehat{F}^*$ at a point $\mathbf{p}$, then $\operatorname{curl}(\widehat{F})|_{\mathbf{p}} = \operatorname{curl}(\widehat{F}^*)|_{\mathbf{p}}$:

Theorem 36.22. *Let $F = (P, Q, R)$ be a differentiable vector field on $\mathbb{R}^3$, with derivative $DF_{\mathbf{p}}$ at the point $\mathbf{p}$. Let F^* denote the linear vector field defined by the matrix $DF_{\mathbf{p}}$.*

Then $curl(F)|_{\mathbf{p}} = curl(F^*)|_{\mathbf{0}}$.

The same holds for $\mathbb{R}^2$.

Proof. For the case of $\mathbb{R}^2$, the derivative matrix is $DF = \begin{bmatrix} P_x & P_y \\ Q_x & Q_y \end{bmatrix}$. The curl is calculated from the off-diagonal entries. So curl(F) and curl(F^*) are the same, as they are determined by these entries.

Now the derivative of a linear map is constant, so $DF^*(\mathbf{x}) = DF^*(\mathbf{0}) = DF_{\mathbf{p}} = F^*$ for all $\mathbf{x}$.

$$DF_{\mathbf{p}} = \begin{bmatrix} P_x & P_y & P_z \\ Q_x & Q_y & Q_z \\ R_x & R_y & R_z \end{bmatrix} |_{\mathbf{p}} = F^*$$

□

The curl is a type of derivative, so this makes sense. A sphere in $\mathbb{R}^3$ rotates about an axis; the curl measures the infinitesimal rotation of the vector field, and its vector points along that axis, using the right-hand rule to indicate the direction of the vector.

See the online text https://activecalculus.org/vector/ for some nice illustrations.

36.5. **Angle as a potential.** First we consider the linear vector field V on $\mathbb{R}^2$ defined by $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$; this is tangent to the rotation flow

$$R_t = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix},$$

see Fig. 85.

The derivative of the linear map $V : \mathbb{R}^2 \to \mathbb{R}^2$ at a point $\mathbf{p}$ is $DV_{\mathbf{p}} = A$ for all $\mathbf{p}$, since the derivative of a linear map is constant, with value equal to the matrix itself. Now writing $V = (P, Q)$, $DV = \begin{bmatrix} P_x & P_y \\ Q_x & Q_y \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, so the curl is $Q_x - P_y = 1 + 1 = 2$. Thus by Proposition 36.21 the field V is not conservative.

For a second proof, we calculate the line integral $\int_\gamma V \cdot \mathrm{d}\gamma$ for the curve $\gamma(t) = (\cos t, \sin t)$, $t \in [0, 2\pi]$. This is

$$\int_0^{2\pi} V(\gamma(t)) \cdot \gamma'(t) \mathrm{d}t = \int_0^{2\pi} (-\sin t, \cos t) \cdot (-\sin t, \cos t) \mathrm{d}t = 2\pi.$$

But this is a closed loop, hence by (iii) of Proposition 36.16 is not conservative.

Next we modify V to a nonlinear vector field F, defined everywhere on the plane except at $\mathbf{0}$.

Thus on $\mathcal{U}$ the open set $\mathbb{R}^2 \setminus (0, 0)$ we define

$$F = (P, Q) = \left(\frac{-y}{x^2 + y^2}, \frac{x}{x^2 + y^2} \right)$$

Exercise 36.1. What is $||F(\mathbf{v})||$ for $\mathbf{v} = (x, y)$ in terms of $r = ||\mathbf{v}||$? Calculate the derivative, DF, and use that to verify that $\text{curl}(F) = \mathbf{0}$.

Lemma 36.23. *Verify:*
(i) For $\theta \in [0, +\infty)$ and for $\gamma : [0, \theta] \to \mathcal{U}$ with $\gamma(t) = (\cos t, \sin t)$ then $\int_\gamma F \cdot d\gamma = \theta$.
(ii) For $\theta \in (-\infty, 0]$ then also then $\int_\gamma F \cdot d\gamma = \theta$.

We can use line integrals to measure (more precisely, to *define*!) the number of times a curve in the plane "winds about" a certian point. Here is the definition for the point $\mathbf{0}$:

Definition 36.9. Given a closed curve γ in $\mathbb{R} \setminus \mathbf{0}$, the *winding number* or *index of γ about $\mathbf{0}$* of $I(\gamma; \mathbf{0}) \equiv 1/2\pi \int_\gamma F \cdot \mathrm{d}\gamma$.

Corollary 36.24. *For $\gamma(t) = (\cos(2\pi nt), \sin(2\pi nt))$ with $t \in [0, 1]$, $n \in \mathbb{Z}$ then the winding number of γ about $\mathbf{0}$ is n.*

Exercise 36.2. Let $A = (1, 0)$ and $B = (1, 1)$ and suppose $\gamma : [a, b] \to \mathbb{R} \setminus \mathbf{0}$ with $\gamma(a) = A, \gamma(b) = B$. What are the possible values of $\int_\gamma F \cdot \mathrm{d}\gamma$? Why, precisely?

To define this for a different point $\mathbf{x} \in \mathbb{R}^2$, we would translate F to $F_{\mathbf{x}} = F - \mathbf{x}$ and set $I(\gamma; \mathbf{x}) \equiv 1/2\pi \int_\gamma F_{\mathbf{x}} \cdot \mathrm{d}\gamma$.

Remark 36.4. This provides one way of defining the inside and outside of a curve: $\mathbf{x}$ is on the outside iff $I(\gamma; \mathbf{x}) = 0$, otherwise on the inside. (For $\mathbf{x} \in Im(\gamma)$ it is not defined).

Remark 36.5. Recalling Definition 23.1, if $f : \mathbb{C} \to \mathbb{C}$ is a complex analytic function, with $f = u + iv$, then this defines a vector field $F = (u, v)$ on $\mathbb{R}^2$. We note that in this case the field F has a special form:

$$DF = \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

since f is analytic iff it is complex differentiable, meaning that $f'(z)$ is a complex number $w = a + ib = re^{i\theta}$, giving a dilation times a rotation. This proves the *Cauchy-Riemann equations* $u_x = v_y$, $u_y = -v_x$.

Now the line integral $\int_\gamma F \mathrm{d}\gamma$ is closely related to the *contour integral* of f over γ, written $\int_\gamma f$. The beginnings of the theory are developed in parallel; see e.g. [MH87] p.95 ff. In particular, the winding mumber can be defined using a contour integral. Of course this is only a starting point for the deep and beautiful subject of Complex Analysis.

Conclusion: Despite the fact that we have $\text{curl}(F) = \mathbf{0}$, this field F cannot be conservative because the integral around the closed loop γ with $\theta = 2\pi$ is $\int_\gamma F \cdot \mathrm{d}\gamma = 2\pi$.

We set $\Omega = \mathbb{R}^2 \setminus \{(x, y) : y = 0, x \geq 0\}$, the plane with the positive part of the x-axis removed. We define the *angle function* $\Theta : \Omega \to (0, 2\pi)$ to be the angle of the point (x, y) measured in the counterclockwise direction from this halfline.

Since $\tan(\theta) = \sin(\theta)/\cos(\theta)$,

Note that Θ can be defined as follows [Spi65], p. 73:

$$\Theta(x,y) = \begin{cases} \arctan(y/x) \text{ for } x>0, y>0 \\ \pi + \arctan(y/x) \text{ for } x<0 \\ 2\pi + \arctan(y/x) \text{ for } x>0, y<0 \\ \pi/2 \text{ for } x=0, y>0 \\ 3\pi/2 \text{ for } x=0, y<0 \end{cases}$$

Definition 36.10. Two curves $\gamma, \eta : [a,b] \to \mathbb{R}^m$ are *homotopic* iff there is a continuous function $\Phi : [0,t] \times [a,b] \to \mathbb{R}^m$ such that $\Phi(0,t) = \gamma(t)$ and $\Phi(1,t) = \eta(t)$. If you draw a picture of this you will see that it says that the first curve can be continuously deformed into the second. A curve γ is said to be *homotopic to a point* iff it is homotopic to a constant curve $\eta(t) = \mathbf{p}$ for all t. For an example, the curve $\gamma(t) = (\cos t, \sin t)$ in $\mathbb{R}^2$ is homotopic to a point; however in the domain $\mathcal{U} = \mathbb{R}^2 \setminus \{\mathbf{0}\}$ it is *not*.

A region $\Omega \subseteq \mathbb{R}^2$ is *simply connected* iff it is pathwise connected and has no "holes", meaning every closed curve is homotopic to a point. In the above example, $\mathcal{U} = \mathbb{R}^2 \setminus \{\mathbf{0}\}$ has a "hole" at $\mathbf{0}$.

The basic result is:

Theorem 36.25. *If a region Ω is simply connected, and if curl$(F) = \mathbf{0}$ on Ω, then there exists a primitive φ for F defined on Ω.*

Lemma 36.26. *???*

Example 49. We analyze the important specific example of the angle function Θ. This is a potential function for the field F, but only on the restricted, simply connected domain $\mathbb{R}^2$ minus the positive real axis.

What happens at the limit as the angle goes to 2π is quite interesting, explained geometrically by the graph of Θ.

For the angle function Θ example we carry this out directly. We choose the initial point $A = (-1, 0)$ and connect it to $B \in \Omega$ by a path γ in Ω, defining φ by $\varphi(A) = 0$, and

$$\varphi(B) = \int_\gamma F \cdot \mathrm{d}\gamma.$$

This is well-defined since Ω is pathwise connected, so by (ii) of Proposition 36.16 it is path-independent.

We claim that there exists c such that $\varphi(x,y) + c = \Theta(x,y)$ for all $(x,y) \in \Omega$. (We will find the value of c).

We use the following path to connect A and $B = (x,y)$. We define $\gamma_1(t) = (-1,t)$ for $t \in [0,y]$ and $\gamma_2(t) = (t,y)$ for $t \in [-1,x]$. Note that $\gamma_1' = (0,1)$, $\gamma_2' = (1,0)$. We define $\gamma = \gamma_1 + \gamma_2$. This goes vertically up from A to the point $(-1,y)$ and then horizontally over to B.

We have

$$\varphi(x,y) = \int_{\gamma_1} F \cdot \mathrm{d}\gamma_1 + \int_{\gamma_2} F \cdot \mathrm{d}\gamma_2.$$

To evaluate this we need to recall some facts about inverse trigonometric functions.
We have $\tan(\theta) = \sin(\theta)/\cos(\theta) = y/x$ so $\arctan(y/x) = \theta$;
$\cot(\theta) = \cos(\theta)/\sin(\theta) = x/y$ so $\operatorname{arccot}(x/y) = \theta$.
The domain of definition of tan is $(-\pi/2, \pi/2)$ and of cot is $(0, \pi)$.
So

$$\Theta(x, y) = \operatorname{arccot}(x/y) \text{ for } y > 0$$

takes values $\theta \in (0, \pi)$
and

$$\Theta(x, y) = \operatorname{arccot}(x/y) + \pi \text{ for } y < 0$$

takes values $\theta \in (\pi, 2\pi)$.
We shall show that

$$\varphi(x, y) = \begin{cases} \operatorname{arccot}(x/y) \text{ for } y > 0 \\ \pi \text{ for } y = 0, x < 0 \\ \operatorname{arccot}(x/y) + \pi \text{ for } y < 0. \end{cases}$$

This will prove that $\varphi = \Theta$.
??
Now

$$\int_{\gamma_1} F \cdot \mathrm{d}\gamma_1 = \int_0^y F(-1, t) \cdot (0, 1)\mathrm{d}t = \int_0^y \frac{x}{x^2 + y^2} \circ (-1, t)\mathrm{d}t =$$
$$\int_0^y \frac{-1}{1 + t^2}\mathrm{d}t = \operatorname{arccot}(y) - \operatorname{arccot}(0) = \operatorname{arccot}(y) - \pi/2.$$

And:

$$\int_{\gamma_2} F \cdot \mathrm{d}\gamma_2 = \int_{-1}^x \frac{-y}{x^2 + y^2} \circ (t, y)\mathrm{d}t = \int_{-1}^x \frac{-y}{t^2 + y^2}\mathrm{d}t =$$
$$\int_{u=-1/y}^{u=x/y} \frac{-1}{u^2 + 1}\mathrm{d}u = \operatorname{arccot}(x/y) - \operatorname{arccot}(-1/y).$$

Here we have used the substitution $u = t/y$, so $\mathrm{d}u = 1/y\mathrm{d}t$, $t = uy$, and

$$\frac{-y}{t^2 + y^2} = \frac{-y}{(uy)^2 + y^2} = \frac{-1}{u^2 + 1}.$$

So

$$\varphi(x, y) = \operatorname{arccot} y - \operatorname{arccot}(-1/y) + \operatorname{arccot}(x/y) - \pi/2.$$

We claim that

$$\operatorname{arccot} y - \operatorname{arccot}(-1/y) = \begin{cases} -\pi/2 \text{ for } y > 0 \\ \pi/2 \text{ for } y < 0 \end{cases}$$

To prove this, we calculate that the following derivative is 0 so we get a constant:

$$\frac{\mathrm{d}}{\mathrm{d}y}\Big(\mathrm{arccot}y - \mathrm{arccot}(-1/y)\Big) = \frac{-1}{1+y^2} - -\frac{-1}{1+(\frac{-1}{y})^2}\cdot y^{-2} = \frac{-1}{1+y^2} + \frac{1}{y^2+1} = 0$$

Now $\cot(\pi/4) = 1, \mathrm{arccot}(1) = \pi/4$; $\cot(-\pi/4) = -1, \mathrm{arccot}(-1) = 3\pi/4$
So

$$\mathrm{arccot}y - \mathrm{arccot}(-1/y) = \begin{cases} -\pi/4 - 3\pi/4 = -\pi/2, y = 1 \\ 3\pi/4 - \pi/4 = \pi/2, y = -1 \end{cases}$$

We know that

$$\varphi(x,y) = \mathrm{arccot}(x/y) + \begin{cases} -\pi \text{ for } y > 0 \\ 0 \text{ for } y < 0 \end{cases}$$

$$\varphi(x,y) + \pi = \mathrm{arccot}(x/y) + \begin{cases} 0 \text{ for } y > 0 \\ \pi \text{ for } y < 0 \end{cases}$$

$$\Theta(x,y) = \mathrm{arccot}(x/y) + \begin{cases} 0 \text{ for } y > 0 \\ \pi \text{ for } y < 0 \end{cases}$$

Hence in fact

$$\Theta = \varphi + \pi.$$

But this makes sense since $\varphi(-1,0) = 0$ while $\Theta((-1,0) = \pi$.

To better understand the potential function Θ, draw its level curves; they are rays from the origin, climbing up like a spiral staircase.

Note that for $\gamma(t) = (\cos t, \sin t)$ then

$$\int_0^{2\pi} F(\gamma(t))\cdot\gamma'(t)\mathrm{d}t = \int_0^{2\pi} (\cos t, \sin t)\cdot(-\sin t, \cos t)\mathrm{d}t = 2\pi$$

and also

$$\lim_{t\to 2\pi}\Theta(\gamma(t)) - \Theta(1) = \lim_{B\to\mathbf{0}}\Theta(B) - \Theta(\mathbf{0}) = 2\pi - 0 = 2\pi$$

so the formula $\int_\gamma F\cdot d\gamma = \varphi(B) - \varphi(A)$ is still valid in the limit; it is also valid if we can somehow allow for a "multi-valued function" as a potential!

See §36.11 below for a different view of this potential.

Remark 36.6. The above proof may remind the reader of the statement of Green's Theorem, which we get to below. Indeed, If we define a vector field F by $F = \nabla\mathbf{f}$, then $F = (P,Q)$ where $P(x,y) = \frac{\partial}{\partial x}f(x,y)$ and $Q(x,y) = \frac{\partial}{\partial y}f(x,y)$. For B denote the corners by $A = (a,c), B = (b,c), C = (b,d), D = (a,d)$. Let $\gamma = \gamma_{A,B}+\gamma_{B,C}+\gamma_{C,D}\gamma_{D,A}$ be unit-speed affine paths around the boundary of . Thus $\gamma_{A,B}(t) = (0,c) + t(1,0) = (t,c)$ for $t \in [a,b]$ and so on. So $\gamma'_{A,B} = (1,0)$. We have $\int_\gamma F\cdot\mathrm{d}\gamma_{A,B} = \int_a^b (P,Q)(\gamma_{A,B})\cdot(1,0)\mathrm{d}t = \int_a^b P(t,c)\mathrm{d}t = \int_a^b \frac{\partial}{\partial x}f(t,c)\mathrm{d}t = f(t,c)|_a^b = f(b,c) - f(a,c)$.

So

$$\int_\gamma F\cdot \mathrm{d}\gamma = \big(f(b,c)-f(a,c)\big)+\big(f(b,d)-f(b,c)\big)+\big(f(b,d)-f(a,d)\big)+\big(f(a,c)-f(a,d)\big)$$

which explains geometrically the above calculations: that

$$\int_\gamma F\cdot \mathrm{d}\gamma = \int_{\gamma_{A,B}+\gamma_{B,C}+\gamma_{C,D}\gamma_{D,A}} = 0$$

is equivalent to that

$$\int_{\gamma_{A,B}+\gamma_{C,D}} = -\int_{\gamma_{B,C}+\gamma_{D,A}}$$

, as we traverse the sides of B in a different order.

Thus Green's Theorem (which is overkill!) can be used to prove the above theorem, since $\mathrm{curl}(F) = \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}$ and hence the mixed partials are equal iff $\mathrm{curl}(F) = 0$ while by Green's Theorem,

$$\int\int_B \mathrm{curl}(F)\,\mathrm{d}x\mathrm{d}y = \int_\gamma F\cdot \mathrm{d}\gamma$$

where γ is the curve of the boundary ∂B consisting of four line segments.

Remark 36.7. (On the interpretation of vectors and of vector fields)

There are various possible interpretations of vectors. The two most important are as *movement* (a translation of position of a particle) and *force* (applied to a particle; it is important to not confuse them as these are completely different! For a simple example, consider the six vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{e}, \mathbf{f}$ in the plane defining the vertices of a regular hexagon, so $\mathrm{d} = -\mathbf{a}$ and so on. We prove that $\mathbf{a}+\mathbf{b}+\mathbf{c}+\mathbf{d}+\mathbf{e}+\mathbf{f} = \mathbf{0}$ in two different ways, using these interpretations. First, movement: we consider the path $\mathbf{0}, \mathbf{a}, \mathbf{a}+\mathbf{b}, \ldots, \mathbf{a}+\mathbf{b}+\mathbf{c}+\mathbf{d}+\mathbf{e}+\mathbf{f}$; this walks along a translated hexagon and returns us to $\mathbf{0}$, so that is the total motion. Second, each vector represents a force applied to a particle located at $\mathbf{0}$. Now they cancel pairwise, giving $\mathbf{0}$ as the resultant force.

The same two interpretations arise for a vector field F. Now the movement interpretation is that the field is tangent to the flow of a fluid, and a particle (perhaps an ant on a leaf!) is being carried along the flow lines. The second interpretation is that a particle is moving according to Newton's law $F = M\mathbf{a}$ in this force field.

Further possible interpretations are for example that F represents a magnetic field, or an area element of a surface as the covector for a two-form. But the first two are certainly the most common and important for our intuition.

36.6. **Line integral with respect to a differential form.** The expression

$$\eta = P\mathrm{d}x + Q\mathrm{d}y$$

is called a *differential one-form* on $\mathbb{R}^2$.

Essentially it is a *field of dual vectors*, elements of the dual vector space V^* to $V = R^2$. Thus the one-form η is *dual* to the vector field $F = (P, Q)$, and conversely, F is dual to η. (The term *duality* in math refers to any situation where you can switch back and forth like that; formally there is a action of the two element group $\mathbb{Z}_2$, i.e. there is a permutation, $V \mapsto V^* \mapsto V$....)

Here we recall that given a vector space V, its dual space V^* is defined to be the set of all *linear functionals* on V, that is all $\lambda : V \to \mathbb{R}$ linear. We call λ a *dual vector* or a *co-vector*. Now given choice of an inner product on V, we can think of a co-vector as simply a vector, as follows: defining a function $\lambda_{\mathbf{v}}$ on V by $\lambda_{\mathbf{v}}(\mathbf{w}) = \langle \mathbf{v}, \mathbf{w} \rangle$, we see that indeed $\lambda_{\mathbf{v}} \in V^*$.

This becomes more subtle in higher dimensions, where V^* is replaced by the set of *alternating k-tensors* on $\mathbb{R}^k$, as we explain shortly.

But the first thing to note is that given a one-form η, we define line integral of a curve γ over η, written as follows: $\int_\gamma \eta = \int_\gamma P\mathrm{d}x + Q\mathrm{d}y$ and defined to be simply equal to the line integral with respect to F,

$$\int_\gamma F \cdot \mathrm{d}\gamma = \int_a^b F(\gamma(t)) \cdot \gamma'(t)\mathrm{d}t.$$

Thus to calculate a line integral over a one-form, the first step is to write it out as a standard line integral with respect to the dual field $F = (P, Q)$.

A key fact about line integrals is that the *orientation* of γ is important, since for $\gamma : [a, b] \mapsto \mathbb{R}^2$ with opposite curve $\widetilde{\gamma} = -\gamma$, then as we know, $\int_{\widetilde{\gamma}} F \cdot \mathrm{d}\widetilde{\gamma} = -\int_\gamma F \cdot \mathrm{d}\gamma$.

Thus γ is an oriented curve, and not just the point set $\mathrm{Im}(\gamma)$, the image of the curve.

This is the same as the difference between the Riemann integral $\int_{[a,b]} f(x)\mathrm{d}x$ and the integral $\int_a^b f(x)\mathrm{d}x = F(b) - F(a)$ defined from a primitive F, since in the second case $A = [a, b]$ is treated as an *oriented interval* and we have $\int_b^a f(x)\mathrm{d}x = -\int_a^b f(x)\mathrm{d}x$.

The version for the double integral of a 2-form is to replace in the Riemann integral the symbol $\mathrm{d}x\mathrm{d}y$ by $\mathrm{d}x \wedge \mathrm{d}y$. Again here, for forms, the order of integration makes a difference! For the Riemann integral this is not the case, as by Fubini's Theorem we have $\int\int_A f(x, y)\mathrm{d}x\mathrm{d}y = \int\int_A f(x, y)\mathrm{d}x\mathrm{d}y$, while for forms the sign changes: $\int\int_A f(x, y)\mathrm{d}x \wedge \mathrm{d}y = -\int\int_A f(x, y)\mathrm{d}y \wedge \mathrm{d}x$.

In this respect integration in Vector Calculus (which is really a part of Differential Geometry) is very different from Analysis. In Analysis, the integral only depends on location, not orientation. A basic example is the Riemann integral $\int_{[a,b]} f(x)\mathrm{d}x$ or $\int\int_A f(x, y)\mathrm{d}x\mathrm{d}y$ and so on, defined for a region $A \subseteq \mathbb{R}^2$ with content 0 boundary. But this also holds for the much more general and powerful Lebesgue integral, defined over perhaps very complicated sets, or more generally with respect to a measure, which essentially keeps track of mass, but not of orientation.

For forms the set comes with an orientation, so it must be a nice enough set that this makes sense. Examples include an open set having a smooth boundary, such as a rectanglular solid or a curved version of such an object. The form itself keeps track of orientation as follows: $\mathrm{d}x \wedge \mathrm{d}y = -\mathrm{d}y \wedge \mathrm{d}x$, $\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z = \mathrm{d}y \wedge \mathrm{d}z \wedge \mathrm{d}x = -\mathrm{d}y \wedge \mathrm{d}x \wedge \mathrm{d}z$, and so on. That is, these expressions are *alternating multilinear forms* just like the determinant function. And such a form is a *field of alternating k-tensors*.

A 1-tensor is the same as an element of the dual space. So as we said, a 1-form is just a co-vector field, i.e. a field of co- or dual vectors, elements of the dual space V^* rather than of our vector space V.

In particular, on $\mathbb{R}^n$ the one-form $\mathrm{d}x_k$ is dual to the constant vector field $F(\mathbf{x}) = \mathbf{e}_k$ where $\mathbf{e}_k$ is the standard basis vector. Any one-form can be written as a linear combination of these multiplied by functions. Thus for example on $\mathbb{R}^3$ we can express a one-form η as

$$\eta = P\mathrm{d}x + Q\mathrm{d}y + R\mathrm{d}z.$$

Again, we then define the line integral with respect to a one-form as equal to its line integral over the associated vector field.

Summarizing,

Definition 36.11. Given a vector space V, a *one-tensor* is an element of the dual space V^*. A *differential one-form* η on a vector space V is a function taking values in the one-tensors, so equivalently, $\eta : V \to V^*$. Choice on an inner product associates V to V^*, by sending $\mathbf{v} \mapsto \lambda_{\mathbf{v}} \in V^*$ with $\lambda_{\mathbf{v}}(\mathbf{w}) = \langle \mathbf{v}, \mathbf{w} \rangle$. This is an isomorphism, which depends on the coice of inner product. For k-tensors and forms with $k > 1$ see Definition 45.1 and Definition 45.3.

Remark 36.8. It is not quite true that oriented integration is limited to nice sets; there is an extension to fractal-like sets or measures, called *currents*, in the field of *Geometric Integration Theory*. See Whitney's [Whi15], still worth looking at for some wonderful inspiration; there is a more recent treatment in [KP08].

36.7. **Green's Theorem: Stokes' Theorem in the Plane.** For our approach we follow the outlines of the elementary proof in Guidorizzi's Calculus 3 text: [Gui02]. In my opinion this is (for those who know Portuguese) a good text to teach from, as it is well organized, with correct proofs and good worked-out examples and exercises of a consistent level, but it's not so easy to study from as it is too dry and also because it lacks the beauty of a more advanced and abstract approach. The latter is given in spades in Spivak's beautiful [Spi65] and Guillemin and Pollack's transcendent [GP74]; the approach in these notes is to bridge the way to this very beautiful and powerful more abstract approach while keeping our feet firmly on the ground of simplicity, inspired also by the perspective of the dynamics of flows.

Definition 36.12. Given a simple closed C^1 curve γ in $\mathbb{R}^2$, so $\gamma : [a, b] \to \mathbb{R}^2$ with $\gamma(a) = \gamma(b)$, we define a curve on the circle by
$wh\gamma(t) = \gamma(t)/||\gamma(t)||$. This is just the normalized tangent vector, so to see how the tangent vector turns, we look at how $\widehat{\gamma}$ moves along the unit circle. One can prove (and it makes sense intuitively) that:

Lemma 36.27. *$\widehat{\gamma}$ either goes around once in the clockwose direction or once in the counterclockwise direction.*

We say γ is *oriented positively* if it is a counterclockwise motion, otherwise we say it is *oriented negatively*.

One has the famous *Jordan Curve Theorem*:

Theorem 36.28. *(Jordan) A continuous simple closed curve γ in $\mathbb{R}^2$ partitions the plane into three connected sets:*
–the interior *of the curve, an open set we call K;*

–the image of γ, a closed set, which is the topological boundary of K, so we call it $\partial K = Im(\gamma)$, the boundary *of K;*
–the exterior *of γ, the open set which is the complement of $K \cup \partial K$.*

Proposition 36.29. *If γ is oriented positvely, then the interior region K is to the left of the tangent vector $\gamma'(t)$ for all t where $\gamma'(t)$ exists and is nonzero.*

Unfortunately, we will not prove any of these beautiful results here, as good proofs require a more advanced perspective, bringing in ideas from algebraic or differential topology; see [GP74], and as they are clear intuitively by sketching a few pictures. These ideas also are needed in Complex Analysis. There is a nice treatment relating this to line integrals in the third edition of Marsden-Hoffman: [MH98].

Theorem 36.30. *(Green's Theorem) Let γ be a simple closed positively oriented curve in $\mathbb{R}^2$ with non-empty interior. Write K for the closure of the interior of γ. Let $F = (P, Q)$ be a $\mathcal{C}^1$ vector field defined on some open set $\mathcal{U} \supseteq K$.*

Then

$$\int_\gamma F \cdot d\gamma = \int \int_K curl(F) \cdot \mathbf{k} dx\, dy.$$

equivalently,

$$\int_\gamma P\, dx + Q\, dy = \int \int_K (Q_x - P_y)\, dx\, dy.$$

The proof will be given in stages:

Proof. **Proof for rectangle:** Let $K = [a, b] \times [c, d]$. Write $A = (a, c), B = (b, c), C = [b, d), D = (a, d)$. Let $\gamma = \gamma_1 + \cdots + \gamma_4$ be unit-speed boundary curves traversing the segments in a counterclockwise direction, γ_1 from A to B and so on. Thus $\gamma_1(t) = A + t(1, 0) = (t, c)$ for $t \in [a, b]$, so $\gamma_1' = (1, 0)$. We have

$$\int_{\gamma_1} P\mathrm{d}x + Q\mathrm{d}y = \int_a^b P(t, c)\mathrm{d}t$$

and similarly for the other cases, so

$$\int_\gamma P\mathrm{d}x + Q\mathrm{d}y = \int_a^b P(t, c)\mathrm{d}t + - \int_a^b P(t, d)\mathrm{d}t + \int_c^d Q(b, t)\mathrm{d}t - \int_c^d Q(a, t)\mathrm{d}t =$$
$$\int_a^b P(t, c) - P(t, d)\mathrm{d}t + \int_c^d Q(b, t) - Q(a, t)\mathrm{d}t.$$

On the other hand,

$$\int_K (Q_x - P_y)\mathrm{d}x\mathrm{d}y = \int_c^d \left(\int_a^b \frac{\partial Q}{\partial x} \mathrm{d}x \right) \mathrm{d}y - \int_a^b \left(\int_c^d \frac{\partial P}{\partial y} \mathrm{d}y \right) \mathrm{d}x =$$
$$\int_c^d Q(b, y) - Q(a, y)\mathrm{d}y - \int_a^b P(x, d) - P(x, c)\mathrm{d}x =$$
$$\int_c^d Q(b, t) - Q(a, t)\mathrm{d}t - \int_a^b P(t, d) - P(t, c)\mathrm{d}t$$

which is exactly what we had before!

Proof for right triangle: We take for K the triangle with corners $A = \mathbf{0}$, $B = (1,0)$, $C = (1,0)$ and with boundary curve $\gamma_1 + \gamma_2 - \gamma_3$ with $\gamma_1(t) = (t,0)$, $\gamma_2(t) = (1,t)$ and $\gamma_3(t) = (t,t)$, all for $t \in [0,1]$.. (Here $-\gamma_3$ means the opposite, i.e. orientation-reversed, curve.) Thus $\gamma_1' = (1,0), \gamma_2' = (0,1)$ and $-\gamma_3' = (1,1)$.

We have for $F = (P,Q)$,

$$\int_\gamma F \cdot \mathrm{d}\gamma = \int_\gamma P\mathrm{d}x + Q\mathrm{d}y = \int_0^1 P(t,0) + Q(1,t)\mathrm{d}t - \int_0^1 P(t,t) + Q(t,t)\mathrm{d}t.$$

On the other hand,

$$\int\int_K Q_x - P_y \mathrm{d}x\mathrm{d}y = \int_{y=0}^{y=1}\left(\int_{x=y}^{x=1} \frac{\partial Q}{\partial x}\mathrm{d}x\right)\mathrm{d}y - \int_{x=0}^{x=1}\left(\int_{y=0}^{y=x} \frac{\partial P}{\partial y}\mathrm{d}y\right)\mathrm{d}x =$$

$$\int_{y=0}^{y=1} Q(x,y)|_{x=y}^{x=1}\mathrm{d}y - \int_{x=0}^{x=1} P(x,y)|_{y=0}^{y=x}\mathrm{d}x =$$

$$\int_{y=0}^{y=1} Q(1,y) - Q(y,y)\mathrm{d}y - \int_{x=0}^{x=1} P(x,x) - P(x,0)\mathrm{d}x$$

which equals the line integral! □

Proof for right triangle with one curvy side:

Next we consider a topological triangle with vertices at $A = (a,c)$, $B = (b,c)$, $C = (c,d)$ and with boundary curve $\gamma_1 + \gamma_2 - \gamma_3$ with $\gamma_1(t) = (t,c)$ for $t \in [a,b]$; $\gamma_2(t) = (b,t)$ for $t \in [c,d]$, and $-\gamma_3$ where $\gamma_3(t) = (t, f(t))$ for $t \in [a,b]$.

We assume that f is invertible, with inverse g.

We have for $F = (P,Q)$:

$$\int_\gamma F\cdot\mathrm{d}\gamma = \int_\gamma P\mathrm{d}x + Q\mathrm{d}y = \int_a^b P(t,c)\mathrm{d}t + \int_c^d Q(b,t)\mathrm{d}t - \int_a^b (P,Q)(\gamma_3(t))\cdot(1,f'(t))\mathrm{d}t$$

Here

$$\int_a^b (P,Q)(\gamma_3(t))\cdot(1,f'(t))\mathrm{d}t = \int_a^b P(t,f(t)) + Q(t,f(t))f'(t)\mathrm{d}t$$

so the total is

$$\int_a^b P(t,c)\mathrm{d}t + \int_c^d Q(b,t)\mathrm{d}t - \int_a^b P(t,f(t)) - \int_a^b Q(t,f(t))f'(t)\mathrm{d}t.$$

On the other hand,

$$\int\int_K Q_x - P_y \,\mathrm{d}x\mathrm{d}y = \int_{y=c}^{y=d}\left(\int_{x=g(y)}^{x=b} \frac{\partial Q}{\partial x}\mathrm{d}x\right)\mathrm{d}y - \int_{x=a}^{x=b}\left(\int_{y=c}^{y=f(x)} \frac{\partial P}{\partial y}\mathrm{d}y\right)\mathrm{d}x =$$

$$\int_{y=c}^{y=d} Q(b,y) - Q(g(y),y)\mathrm{d}y - \int_{x=a}^{x=b} P(x,f(x)) - P(x,c)\mathrm{d}x =$$

$$\int_{x=a}^{x=b} P(x,c)\mathrm{d}x + \int_{y=c}^{y=d} Q(b,y)\mathrm{d}y - \int_{x=a}^{x=b} P(x,f(x))\mathrm{d}x - \int_{y=c}^{y=d} Q(g(y),y)\mathrm{d}y$$

We are almost done. Note that each expression has four terms, and the first three of them agree, just changing the variable of integration from time t to the spatial coordinates x and y. It remains to check the last term. This is a substitution,

making use of the inverse function: writing $s = f(t)$, so $t = g(s)$, then $\mathrm{d}s = f'(t)\mathrm{d}t$ whence indeed

$$\int_a^b Q(t, f(t))f'(t)\mathrm{d}t = \int_{s=c}^{s=d} Q(g(s), s)\mathrm{d}s = \int_{y=c}^{y=d} Q(g(y), y)\mathrm{d}y$$

completing the proof.

□

Proof for more complicated regions.

Once we have these special cases we can build up to the general statement of Green's Theorem as follows. First we consider other cases of an open region K with a simple closed piecewise-$\mathcal{C}^1$ boundary curve γ. Using straigt lines, we cut K into pieces of the above forms and add up the results. The key point is that the pieces have nonintersecting interiors, and meet on their boundaries in curves with opposite orientation. The double integrals add as this boundary has content zero so those add; on the line integral side of the equation, the question is why do the boundary intersections always meet in curves with opposite orientation? But this is easy to justify: we prove this by induction on the number of pieces, reducing to two regions. Their boundaries meet in curves with opposite orientations because each is counterclockwise as seen from its own interior, hence opposite as seen from the other region.

The next step is to consider two disjoint simple closed piecewise-$\mathcal{C}^1$ boundary curves γ_1, γ_2 with regions K_1, K_2. If these regions are disjoint, we simply define the boundary of the union $k_1 \cup K_2$ to be γ_1 together with γ_2, which we write as $\gamma_1 + \gamma_2$. The result clearly hols for this case also. Next consider the case where γ_2 is inside of K_1. Then we consider the region $K = K_1 \setminus (K_2 \cup \mathrm{Im}(\gamma_2)$. For example, if the curves are concentric circles, then K is called an *annulus*: a disk with a hole removed from it. We defined the boundary curve to be $\gamma = \gamma_1 - \gamma_2$. That is, the outer curve γ_1 is oriented positively, while the inner curve is oriented negatively.

Note that the resulting boundary curve γ has the property that as we traverse the curve, the region K always occurs on the *left-hand side*.

It is then easy to show by subtracting the two results for γ_1, γ_2 that Green's Theorem still holds.

Note that such a region is now not simply connected.

We do similarly for a disk with k holes removed.

A more formal proof uses the notion of *chains* as developed in [Spi65] or [GP74].

Exercise 36.3. Consider the field

$$F = (P, Q) = \left(\frac{-y}{x^2 + y^2}, \frac{x}{x^2 + y^2}\right)$$

of Exercise 36.1, for the region with two boundary circles of radius 1 and 2. What does Green's Theorem say in this case?

36.8. **The Divergence Theorem in the plane.**

Definition 36.13. Let $F = (P, Q)$ be a $\mathbb{C}^1$ vector field in $\mathbb{R}^2$. The *divergence* of F is defined to be:

$$\mathrm{div}(F) = P_x + Q_y.$$

We shall use the notation: given a vector $\mathbf{v} = (a, b) \in \mathbb{R}^2$, then $\mathbf{v}^* = (b, -a)$ and $\widetilde{\mathbf{v}}^* = (-b, a)$.

For the particular case of $F = (P, Q)$ we write G for $\widetilde{F}^* = (-Q, P)$.

Theorem 36.31. *Let $F = (P, Q)$ be a $\mathbb{C}^1$ vector field in the plane, and let γ be a piecewise $\mathbb{C}^1$, positively oriented simple closed curve, with interior region K. We define $\mathbf{n} = \gamma'^*/||\gamma'||$; this is the* outward normal vector *of γ.*

Then

$$\int_\gamma F \cdot \mathbf{n} ds = \int \int_K div(F) dx dy.$$

The same holds more generally for a finite collection of disjoint such regions $K_1, \dots K_n$ with boundaries $\gamma_1, \dots, \gamma_n$ and then writing $K = \cup K_n$ and $\gamma = \gamma_1 + \cdots + \gamma_n$.

Proof. We place the two statements side-by-side, for γ the boundary curve of K, one for the field F and the other for $G = \widetilde{F}^*$:

Green's Theorem:

$$\int_\gamma G \cdot \mathrm{d}\gamma = \int \int_K \mathrm{curl}(G) \cdot \mathbf{k} \, \mathrm{d}x\mathrm{d}y$$

Divergence Theorem:

$$\int_\gamma F \cdot \mathbf{n} \, \mathrm{d}s = \int \int_K \mathrm{div}(F) \, \mathrm{d}A.$$

Note here that $\mathrm{curl}(G) \cdot \mathbf{k} = \mathrm{div}(F)$, so once the prove the two different types of line integrals are equal, the theorem is proved!

For $\gamma(t) = (x(t), y(t))$, then $\gamma'(t) = (x'(t), y'(t))$, and $\mathbf{n} = \gamma'^*/||\gamma'|| = (y', -x')/||\gamma'|| = (y', -x')/||(y', -x')||$.

Recall that the line integral of second type of a function $f : \mathbb{R}^2 \to \mathbb{R}$ over $\gamma : [a, b] \to \mathbb{R}^2$ is defined to be

$$\int_\gamma f(\mathbf{v}) \mathrm{d}s \equiv \int_a^b f(\gamma(t)) \, ||\gamma'(t)|| \mathrm{d}t$$

where $\mathrm{d}s$ is the element of arclength, $\mathrm{d}s = ||\gamma'(t)|| \mathrm{d}t$. Now for this to make sense, it is enough for the function f to be defined on the image of γ, not necessarily on all of $\mathbb{R}^2$. So when we write the formula

$$\int_\gamma F \cdot \mathbf{n} \, \mathrm{d}s$$

what we mean by this is the line integral of second type of the function f over γ, where f is defined on the image of γ by

$$f(\gamma(t)) = F(\gamma(t)) \cdot \mathbf{n}(t).$$

Thus

$$\int_\gamma F \cdot \mathbf{n} \, \mathrm{d}s \equiv \int_\gamma f(\mathbf{v}) \mathrm{d}s \equiv \int_a^b f(\gamma(t)) ||\gamma'(t)|| \mathrm{d}t.$$

Now writing in components $F = (P, Q)$, we have

$$\int_\gamma F\cdot\mathbf{n}\,\mathrm{d}s=\int_a^b F(\gamma(t))\cdot\mathbf{n}(t)\;||\gamma'(t)||\mathrm{d}t=\int_a^b F(\gamma(t)))\cdot(y',-x')/||\gamma'(t)||\;\;||\gamma'(t)||\mathrm{d}t=$$
$$\int_a^b (P,Q)(\gamma(t))\cdot(y',-x')\,\mathrm{d}t=\int_a^b(-Q,P)(\gamma(t))\cdot(x',y')\,\mathrm{d}t=\int_a^b G(\gamma(t))\cdot\gamma'(t)\,\mathrm{d}t=\int_\gamma G\cdot\mathrm{d}\gamma=$$
$$\int\int_K \mathrm{curl}(G)\cdot\mathbf{k}\,\mathrm{d}x\mathrm{d}y=\int\int_K \mathrm{div}(F)\,\mathrm{d}x\mathrm{d}y.$$

□

Remark 36.9. An explanation is that F is lined up with $\mathbf{n}$, thus producing positive divergence, iff $\widetilde{F}^*$ is lined up with γ', thus producing positive curl. The reason for using $\widetilde{F}^*$ rather than F^* is so the sign matches; the key point is that for $\mathbf{v}=(a,b)$ and $\mathbf{w}=(c,d)$, then $\mathbf{v}^*=(b,-a)$ and $\widetilde{\mathbf{w}}^*=(-c,d)$, and $\mathbf{v}\cdot\mathbf{w}^*=\mathbf{w}\cdot\widetilde{\mathbf{v}}^*$. So $\alpha(\mathbf{v},\mathbf{w})\equiv\mathbf{v}\cdot\mathbf{w}^*$ defines an alternating form; indeed, it equals $\det(\mathbf{v},\mathbf{w})$! See Proposition 45.2 ff. regarding two-tensors.

Using this notation, the last part of the proof can be summarized as:

$$\int_\gamma F\cdot\mathbf{n}\,\mathrm{d}s=\int_a^b F(\gamma(t))\cdot\gamma'^*(t)\mathrm{d}t=\int_a^b\widetilde{F}^*(\gamma(t))\cdot\gamma'(t)\mathrm{d}t=$$
$$\int\int_K\mathrm{curl}(\widetilde{F}^*)\cdot\mathbf{k}\,\mathrm{d}x\mathrm{d}y=\int\int_K\mathrm{div}(F)\,\mathrm{d}x\mathrm{d}y.$$

See p. 79 of [War71] regarding the star operator.

Poincaré's Lemma: Existence of the vector potential.

A key idea of Vector Calculus is to extend the Fundamental Theorem of Calculus in a variety of ways. The first is that if for a vector field F on $\mathbb{R}^n$, we have a function $\phi:\mathbb{R}^n\to\mathbb{R}$ such that $\nabla\varphi=F$, then for a $\mathbb{C}^1$ path $\gamma:[a,b]\to\mathbb{R}^n$ with endpoints $A=\gamma(a), B=\gamma(b)$ then

$$\int_\gamma F\cdot\mathrm{d}\gamma=\int_a^b F(\gamma(t))\cdot\gamma'(t)\mathrm{d}t=\varphi(B)-\varphi(A).$$

This is just like the case in one dimension where given $f:[a,b]\to\mathbb{R}$ and a function F satisfying $F'=f$ then

$$\int f(x)\mathrm{d}x=F(b)-F(a).$$

There is however one important difference: for f Riemann integrable there always exists such a *primitive* or *antiderivative* F, while for higher dimensions this only works if the field F is *conservative*, is which case φ is called a *potential function.*

Equivalently, in differental form notation, say for $F=(P,Q)$ then the form $\eta=P\mathrm{d}x+Q\mathrm{d}y$ has a *primitive* φ such that $\mathrm{d}\varphi=\eta$. This leads to the nice formula

$$\int_\gamma\eta=\int_\gamma P\mathrm{d}x+Q\mathrm{d}y=\int_{\partial\gamma}\mathrm{d}\eta=\int_{B=A}\varphi=\varphi(B)-\varphi(A).$$

The terminology thus that F has a potential φ iff η has a primitive φ.

In one dimension the potential is simply called the primitive, and can be defined from the integral by: $F(x) = \int_{x_0}^{x} f(r)\mathrm{d}r$. This is defined up to a constant, choice of which corresponds to changing the initial point x_0; thus we have made the choice $F(x_0) = 0$.

Exactly the same thing works for the line integrals, where we can attempt to define a potential function in the same way. Again this is defined up to an initial poin A, setting $\varphi(A) = 0$ and $\varphi(B) = \int_a^b F(\gamma(t)) \cdot \gamma'(t)\mathrm{d}t$. This will be defined if there exists a path γ connecting A and B (by definition, iff the region is pathwise-connected) and will be *well*-defined iff this definition is independent of the path chosen. That is one of the equivalent characterizations of conservative field. (To prove that this definition indeed gives a potential, we calculate the partials and show one indeed recovers the field; this is the "hardest" step in proving the equivalence of the conditions).

This result, which extends the Fundamental Theorem of Vector Calculus, itself extends much further, to Green's Theorem, the Divergence Theorem, and Stokes' Theorem in $\mathbb{R}^2$ and $\mathbb{R}^3$. All of this becomes simultaneously much more complicated and much simpler in its most natuarl setting, to the generalized Stokes Theorem on manifolds with boundary.

Why we say "much simpler" is shown by the statement:

$$\int_{\partial B} \eta = \int_B \mathrm{d}\eta$$

or equivalently

$$\langle \partial B, \eta \rangle = \langle B, \mathrm{d}\eta \rangle.$$

The second notation exhibits the integral as a bilinear form, like an inner product. However here the elements on the right-hand side are differential forms, which form a vector space, while on the lef-hand side these are *chains*, parametrized manifolds which can be added, subtracted or multiplied by integers, thus belonging to a *module* (over the ring $\mathbf{z}$) rather than a vector space.

This second equation says that the *boundary operator* ∂ on chains is *dual* to the *exterior derivative* operator d on forms. This relationship can be summarized by saying that these operators are *adjoints*. (Note that this is indeed analogous to the definition of the transpose, or adjoint, of a linear operator!)

The first difficulty hidden by this simple notation is all in the definitions, which are equally abstract and deep. The secondary difficult comes in bridging the abstraction to the concrete versions of Vector Calculus in $\mathbb{R}^2$ and $\mathbb{R}^3$.

We mention two auxilliary points which come up in all these settings. The basic theorem is Stokes, which can be thought of as (and indeed can be called) the *Fundamental Theorem of Vector Calculus*.

We shall need:

Definition 36.14. A differential k- form η is *closed* iff $\mathrm{d}\eta = 0$.

It is *exact* iff there exists a $(k-1)$-form α such that $\mathrm{d}\alpha = \eta$.

The two other results are these:

Theorem 36.32. *(Poincaré Lemma) On a simply connected domain, a closed form is exact.*

Thus the Poincaré Lemma says that for topologically nice doman (simply connected), a primitive always exists; specifically, for one-forms in $\mathbb{R}^n$, we know this, since the dual vector field has a potential φ, and $\nabla\varphi = F$ iff $\mathrm{d}\varphi = \eta = \sum P_i \mathrm{d}x_i$.

The second related result is:

Theorem 36.33. *(Hodge Decomposition) On a simply connected domain, every differential form can be uniquely written as the sum of a closed form and an exact form.*

For vector fields in $\mathbb{R}^n$, we say:

Definition 36.15. A vector field F is *divergence-free* or *incompressible* iff $\mathrm{div}(F) = 0$. It is *curl-free* or *conservative* or *irrotational* iff $\mathrm{curl}(F) = 0$.

The Hodge decomposition then gives:

Theorem 36.34. *(Helmholtz Decomposition) On a simply connected domain, every vector field can be uniquely written as the sum of a two vector fields, one divergence-free and one curl-free.*

Corollary 36.35. *A vector field on a simply connected domain is determined by its divergence and its curl.*

Proof. By the Helmholtz Decomposition, our field$F = F_d + F_c$ where F_d is curl-free and F_c is divergence-free. Then $\mathrm{curl}(F) = \mathrm{curl}(F_c) + \mathrm{curl}(F_d) = \mathrm{curl}(F_c)$ and $\mathrm{div}(F) = \mathrm{div}(F_c) + \mathrm{div}(F_d) = \mathrm{div}(F_d)$. Hence $F = \mathrm{curl}(F) + \mathrm{div}(F)$.??? □

For vector fields on a simply connected domain in $\mathbb{R}^n$, there are two versions of Poincaré's Lemma. The first says that a curl-free vector field has a potential, hence is conservative: if $\mathrm{curl}F = 0$ then there exists φ such that $\nabla\varphi = F$.

The second statement is:

Theorem 36.36. *If div$(F) = 0$, then there exists a field A such that curl$(A) = F$.*

For the proof we need:

Lemma 36.37. *(Derivative under the Integral) Suppose for $\mathcal{U} \subseteq \mathbb{R}^2$ open that $f : \mathcal{U} \to \mathbb{R}$ is continuous, and that $\partial f/\partial y$ exists and is continuous. Define $\varphi(y) = \int_a^b f(x,y)\,dx$. Then*

$$\varphi'(y) = \frac{d}{dy}\int_a^b f(x,y)\,dx = \int_a^b \frac{\partial f}{\partial y}(x,y)\,dx.$$

Example 50. Before the proof, we consider some examples.

Remark 36.10. For these examples, recall that the Gaussian function e^{-x^2} is a well-known example for which the antiderivative cannot be found "in closed form". Roughly this means as a finite formula involving other elementary functions (polynomials, trigonometric functions, log and exp); for a precise statement, which makes use of the notion of a *differential field,* see [Ros68]. (Note that one can however easily give an infinite formula, using Taylor's series.)

Problem: Find $\partial\varphi/\partial x$ and $\partial\varphi/\partial y$ for

$$\varphi(x,y) = \int_0^x e^{-yt^2}\mathrm{d}t.$$

The first is easy, as by the Fundamental Theorem of Calculus we have: $\frac{\partial\varphi}{\partial x} = e^{-yx^2}$.

For the second, we apply the Lemma, giving:

$$\frac{\partial\varphi}{\partial y}\int_0^x e^{-yt^2}\mathrm{d}t = \int_0^x \frac{\partial}{\partial y}e^{-yt^2}\mathrm{d}t = \int_0^x -t^2e^{-yt^2}\mathrm{d}t.$$

Problem: For

$$h(t) = \int_0^{t^2} e^{-tu^2}\mathrm{d}u,$$

calculate $h'(t)$.

Well, this is indeed pretty confusing, as the variable t occurs in two different spots! The trick is to first define a function of two variables

$$\varphi(x,y) = \int_0^x e^{-yu^2}\mathrm{d}u$$

and then compose it with a curve. Now as above $\frac{\partial\varphi}{\partial x} = e^{-yx^2}$, while

$$\frac{\partial\varphi}{\partial y} = \int_0^x -u^2e^{-yu^2}\mathrm{d}u.$$

Defining the curve $\gamma(t) = (t^2, t)$ then $h(t) = \varphi(\gamma(t))$. We then apply the Chain Rule:

$$h'(t) = \nabla\varphi|_{\gamma(t)}\cdot\gamma'(t) = \frac{\partial\varphi}{\partial x}|_{\gamma(t)}x'(t) + \frac{\partial\varphi}{\partial y}|_{\gamma(t)}y'(t) =$$

$$e^{-tt^4}2t + \int_0^{t^2} -u^2e^{-tu^2}\mathrm{d}u = 2te^{-t^5} - \int_0^{t^2} u^2e^{-tu^2}\mathrm{d}u.$$

Proof. (of Lemma) Our proof follows Apostol p. 448 [**?**].

We are given that $\varphi(y) = \int_a^b f(x,y)\mathrm{d}x$ and want to find $\varphi'(y)$. We have:

$$\frac{\varphi(y+h)-\varphi(y)}{h} = \frac{1}{h}\left(\int_a^b f(x,y+h)\mathrm{d}x - \int_a^b f(x,y)\mathrm{d}x\right) = \frac{1}{h}\left(\int_a^b f(x,y+h) - f(x,y)\mathrm{d}x\right)$$

Now by the Mean Value Theorem, for each fixed y there exists $c_y \in [a,b]$ such that

$$f(x,y+h) - f(x,y) = \frac{\partial f}{\partial y}(c_y,y)\cdot h.$$

So

$$\frac{\varphi(y+h)-\varphi(y)}{h} = \frac{1}{h}\left(\int_a^b \frac{\partial f}{\partial y}(c_y,y)\cdot h\ \mathrm{d}x\right) = \int_a^b \frac{\partial f}{\partial y}(c_y,y)\ \mathrm{d}x$$

But by continuity of the partial derivative, $\frac{\partial f}{\partial y}(c_y,y) \to \frac{\partial f}{\partial y}(x,y)$ as $h \to 0$. This gives

$$\frac{\varphi(y+h)-\varphi(y)}{h} \to \int_a^b \frac{\partial f}{\partial y}(x,y)\ \mathrm{d}x$$

as claimed.

□

Proof. (of Theorem)

We give the proof for the easier case of a star-shaped domain.

Let $F = (P, Q, R)$, with $0 = \text{div}F = P_{xx} + Q_{yy} + R_{zz}$.

We want to find a field $A = (L, M, N)$ such that $\text{curl}(A) = F$. Now

$$\text{curl}A = (N_y - M_z, L_z - N_x, M_x - L_y).$$

We consider the simpler case where $L = 0$. Then we have

$$L_z - N_x = -N_x = Q, M_x - L_y = M_x = R$$

Now $\partial N(x, y, z)/\partial x = \text{d}/\text{d}t|_{t=x}(Q(t, y, z)$ so for any inital point $x_0, y, z)$ we have

$$N(x, y, z) = \int_{t=x_0}^{x} Q(t, y, z)\text{d}t + c(y, z)$$

where $c(y, z)$ is constant in x. Similarly

$$M(x, y, z) = \int_{t=x_0}^{x} R(t, y, z)\text{d}t + d(y, z)$$

where $c(y, z)$ is constant in x.

We look for a solution with $c(y, z) = 0$. We know that $P = N_y - M_z$ so subtracting the previous two equations gives

$$P = N_y - M_z = \partial/\partial y \int_{t=x_0}^{x} Q(t, y, z)\text{d}t - \partial/\partial z \int_{t=x_0}^{x} R(t, y, z)\text{d}t + \partial/\partial z d(y, z)$$

Now from the Lemma, taking the derivative inside the integral, this gives

$$\int_{t=x_0}^{x} \partial/\partial y Q(t, y, z)\text{d}t - \int_{t=x_0}^{x} \partial/\partial z R(t, y, z)\text{d}t + \partial/\partial z d(y, z) =$$
$$\int_{t=x_0}^{x} -\partial/\partial y Q(t, y, z) - \partial/\partial z R(t, y, z)\text{d}t + \partial/\partial z d(y, z)$$

Using the fact that $\text{div}F = 0$, we know that $-Q_y - R_z = P_x$ so this is

$$\int_{t=x_0}^{x} -\partial/\partial x P(t, y, z)\text{d}t + \partial/\partial z d(y, z) = P(x, y, z) - P(x_0, y, z) + \partial/\partial z d(y, z)$$

We now have the equation

$$P(x, y, z) = P(x, y, z) - P(x_0, y, z) + \partial/\partial z d(y, z)$$

So we will be done if we can find a function $d(y, z)$ satisfying

$$\partial/\partial z d(y, z) = -P(x_0, y, z)$$

(check all signs!)

So we simply define

$$d(y, z) = \int_{t=z_0}^{z} P(x_0, y, r)\text{d}r$$

giving the first part of the solution, defined up to a constant.

So far we have shown that for $F = (P, Q, R)$ then the field $A = (L, M, N)$ with $L = 0$.

$$N(x, y, z) = \int_{t=x_0}^{x} Q(t, y, z)\mathrm{d}t$$

$$M(x, y, z) = \int_{t=x_0}^{x} R(t, y, z)\mathrm{d}t + d(y, z)$$

where

$$d(y, z) = \int_{t=z_0}^{z} P(x_0, y, r)\mathrm{d}r$$

Putting these together we have shown that given $F = (P, Q, R)$, then for $A = (L, M, N)$ defined by

$$L = 0$$

$$N(x, y, z) = \int_{t=x_0}^{x} Q(t, y, z)\mathrm{d}t$$

$$M(x, y, z) = \int_{t=x_0}^{x} R(t, y, z)\mathrm{d}t + \int_{t=z_0}^{z} P(x_0, y, r)\mathrm{d}r.$$

then we have

$$\mathrm{curl}(A) = F$$

.

□

Remark 36.11. There is a strangeness in the above proof as we arbitrarily chose $L = 0$ and yet somehow found a solution.

This is explained by noting that any solution A above is defined up to addition of a field B with $\mathrm{curl}(B) = 0$. Call the particular solution above A_L. Then if we carry out the above construction assuming instead that $M = 0$ we get solution A_M and if we assume instead $N = 0$ we get solution A_N. But then indeed $\mathrm{curl}(A_L - A_M) = F - F = 0$ and similarly $\mathrm{curl}(A_L - A_N) = 0$, $\mathrm{curl}(A_M - A_N) = 0$.

36.9. **Stokes' Theorem.** Green's Theorem and the Divergence Theorem both turn out to be a special case of the fundamental result of vector calculus: Stokes' Theorem, where the points A, B, C, D are the boundary of the curve γ and get replaced by the boundary of any domain.

$$\int_{\partial\Omega} \omega = \int_{\Omega} \mathrm{d}\omega$$

or, in a different notation,

$$\langle \partial\Omega, \omega \rangle = \langle \Omega, \mathrm{d}\omega \rangle.$$

In this notation, which can be called *functional notation*, $\langle \cdot, \cdot \rangle$ is a *pairing* . A pairing is a bilinear operator, but on the right we have a vector space (of d-forms) and on the left an additive group (of d-chains, generated by d dimensional submanifolds). Here $d = k - 1$ on the left and $d = k$ on the right. The analogous assumption to the field being conservative is hidden here, in that we begin with a $k - 1$-form on the left, like the potential, and take its derivative on the right, like its gradient.

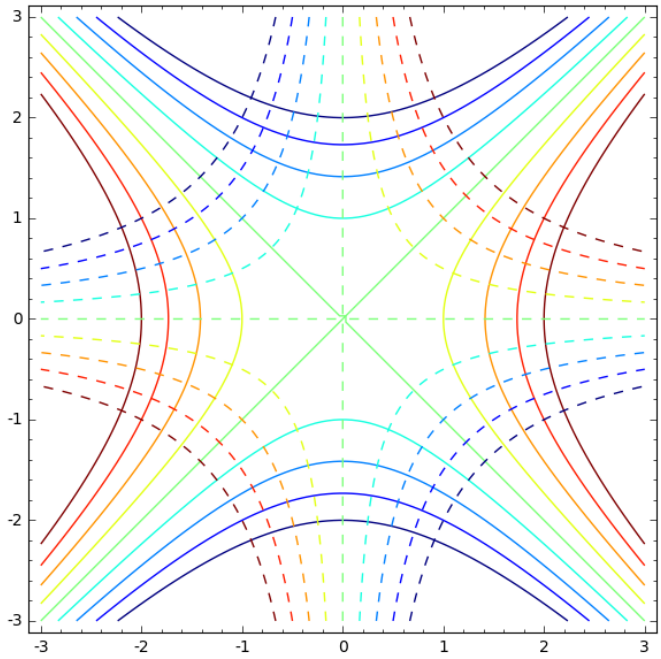


FIGURE 93. Dual families of hyperbolas: Level curves (equipotential curves) for the real and imaginary parts of $f(z) = z^2 = (x+iy)(x-iy) = (x^2 - y^2) + 2(xy)i$

36.10. **Analytic functions and harmonic conjugates.**

Definition 36.16. A function $f : \mathcal{U} \subseteq \mathbb{C} \to \mathbb{C}$ is *(complex) differentiable* at $z \in \mathcal{U}$, with derivative $f'(z)$, iff $f'(z) = \lim_{h\to 0}(f(z+h) - f(z))/h$ exists. It is *complex analytic* iff it is differentiable for every $z \in \mathcal{U}$. See also Definition 23.1 and Remark 36.5 for equivalent conditions.

A function $u : \mathcal{U} \to \mathbb{R}$ is *harmonic* iff u is $\mathcal{C}^2$ and $u_{xx} + u_{yy} = 0$.

We define a linear operator Δ, also written as ∇^2 and called the *Laplacian*, on the vector space $\mathcal{C}^2(\mathcal{U}, \mathbb{R})$ by $\Delta(u) = u_{xx} + u_{yy}$. So u is harmonic iff $\Delta(u) = 0$, iff u is in the kernel of the operator.

The reason for the notation ∇^2 isbeacuse it is notationally suggestive, as we can think of it as the dot product: $\nabla^2\varphi = (\nabla\cdot\nabla)(\varphi) = \nabla\cdot(\nabla(\varphi) = (\delta/\delta x + \delta/\delta y)\cdot(\varphi_x + \varphi_y)$.

Theorem 36.38. *For a complex analytic function $f : \mathcal{U} \to \mathbb{C}$, where $\mathcal{U} \subseteq \mathbb{C}$ is open, with real and imaginary parts $u = \Re(f), v = Im(f)$ so $f = u + iv$, then thought of as real functions on $\mathcal{U} \subset \mathbb{R}^2$,*
(i) these satisfy the Cauchy-Riemann equations *$u_x = v_y$, $u_y = -v_x$;*
(ii) u, v are both harmonic functions;
(iii) their gradient vector fields are orthogonal;
(iv) their families of level curves are orthogonal.

Proof. If $f : \mathbb{C} \to \mathbb{C}$ is a complex analytic function, then by definition the derivative $f'(z) = \lim_{h\to 0}(f(z+h) - f(z))/h$ is a complex number $w = (a + ib) = re^{i\theta}$. Now multiplication by a complex number defines a linear transformation of $\mathbb{C}$ hence of $\mathbb{R}^2$;

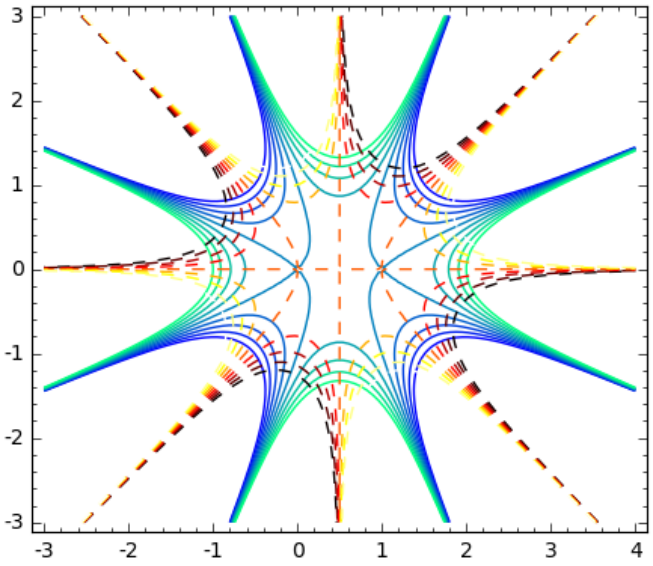


FIGURE 94. Level curves for the real and imaginary parts of $f(z) = z^2(z-1)^2$.

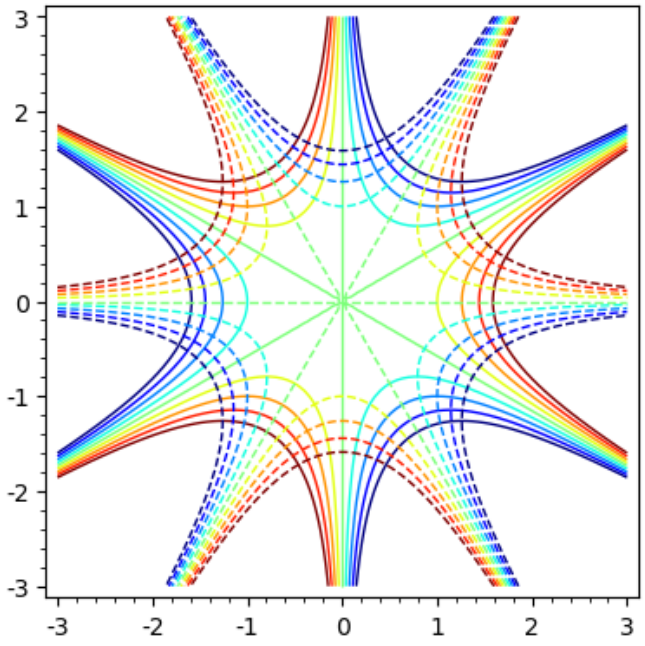


FIGURE 95. Level curves for the real and imaginary parts of $f(z) = z^3$.

since this is a rotation followed by a dilation, this matrix has a special form. Writing $f = u + iv$, then thought of as a map F of $\mathbb{R}^2$, this is the vector field $F = (u, v)$, the

derivative of which is the matrix

$$DF = \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix}.$$

Because we know this is a rotation by θ followed by a dilation by $r \geq 0$, this equals

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = r \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

This proves the Cauchy-Riemann equations $u_x = v_y$, $u_y = -v_x$.

Now $u_x = v_y$ whence $u_{xx} = v_{xy}$ and

$u_y = -v_x$ whence $u_{yy} = -v_{yx}$, giving that

$$u_{xx} + u_{yy} = v_{xy} - v_{yx} = 0$$

by the equality of mixed partials, Lemma 36.20.

Similarly, from $u_x = v_y$ we have that $u_{yx} = v_{yy}$ and from $u_y = -v_x$ that $u_{xy} = -v_{xx}$ whence

$$v_{xx} + v_{yy} = u_{xy} - u_{yx} = 0.$$

So both u and v are harmonic.

Recalling the notation that for $F = (P, Q)$ then $F^* = (Q, -P)$, we have

$$\nabla u = (u_x, u_y) = (v_y, -v_x)$$

so

$$\nabla v = F^*$$

gives an orthogonal field.

Lastly the level curves are perpendicular to the gradient fields, F and F^*, so since these are orthogonal so are those families of curves.

□

In fact the converse also holds:

Proposition 36.39. *If u, v are $\mathcal{C}^2$ functions which are harmonic, such that the pair (u, v) satisfies the Cauchy-Riemann equations, then $f = u + iv$ is analytic.*

Proof. As above, the derivative of $F : \mathbb{R}^2 \to \mathbb{R}^2$ with $F = (u, v)$ is the matrix

$$DF = \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix}.$$

The Cauchy-Riemann equations imply that this equals

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

and so the map is given by multiplication by the complex number $w = re^{i\theta}$ as above. That the limit exists for DF implies that the limit exists for $f'(z)$ and equals w.

□

Definition 36.17. If if $f = u+iv$ is analytic then u, v are called *harmonic conjugates.*

Proposition 36.40.

(i) If $\mathcal{U}$ is a simply connected domain and $u : \mathcal{U} \to \mathbb{R}$ is harmonic, then there exists a unique $v : \mathcal{U} \to \mathbb{R}$ such that (u, v) are harmonic conjugates.
(ii) The ordered pair (u, v) are harmonic conjugates iff the pair $(v, -u)$ are (so order matters here!)

Proof. (i): By the previous proposition it is enough to find v harmonic such that (u, v) satisfies the Cauchy-Riemann equations, so such that $v_y = u_x$ and $v_x = -u_y$. But this is just like the problem of finding a potential for a curl zero vector field!

Thus, we consider the vector field $F = (P, Q) = (-u_y, u_x)$.

Then $\text{curl}(F) \cdot \mathbf{k} = Q_x - P_y = -u_{xx} - u_{yy} = 0$.

By Theorem 36.25, there is a potential for F; we call this v. Thus $\nabla(v) = (v_x, v_y) = (P, Q) = (-u_y, u_x)$ so $v_{xx} + v_{yy} = -u_{xy} + u_{yx} = 0$ by the equality of mixed partials, whence v is a harmonic function such that the pair (u, v) are indeed harmonic conjugates.

In this proof of (i) we have followed Churchill [CB14]. □

Remark 36.12. Lang [Lan99] and Marsden-Hoffman [MH87] have nice treatments of this. Following Marsden and Hoffman, note that since $if = v - iu$ is analytic, then v and $-u$ are harmonic conjugates (but that the order is important!) A second, purely complex analytic, proof of (i) is given by Marsden [MH87]. See also Ahlfors [Ahl66].

Fig. 93 shows the harmonic conjugates for the function $f(z) = z^2$.

Corollary 36.41. *Given a harmonic function $u : \mathcal{U} \to \mathbb{R}$, where $\mathcal{U}$ is a simply connected domain in $\mathbb{R}^2$, then there exists a unique $v : \mathcal{U} \to \mathbb{R}$, which is harmonic such that (u, v) satisfy the Cauchy-Riemann equations. Also there exists a unique analytic f on $\mathcal{U}$ thought of as a subset of $\mathbb{C}$ such that $u = \Re(f)$. Moreover $f = u + iv$. Writing $\widetilde{f}$ for the second analytic function defined from the harmonic function v, then $\widetilde{f} = v - iu$ has harmonic conjugate pair $(v, -u)$. Furthermore $\widetilde{f}(z) = -if(z)$.*

Harmonic functions are characterized by the important *mean value property*: for a proof see e.g. [MH87].

Theorem 36.42. *A $\mathcal{C}^2$ function u is harmonic iff the value at a point $\mathbf{p}$ is equal to the average of the values on any circle about $\mathbf{p}$.*

Definition 36.18. A flow τ_t on $\mathbb{R}^n$ is a *gradient flow* iff there is a function $\varphi : \mathbb{R}^n \to \mathbb{R}$ such that for the field $F = \nabla\varphi$, then the flow orbits are tangent to the gradient vector field. That is, the orbits $\gamma(t) = \tau_t(x)$ for some initial point x satisfy the differential equation

$$\gamma'(t) = F(\gamma(t))$$

.

We conclude:

Theorem 36.43. *Let u be a harmonic function on $\mathcal{U} \subseteq \mathbb{R}^2$. Let v be its harmonic conjugate. Write $F = (u, v)$ and $\widetilde{F} = (-v, u)$, so $F = \nabla u$ and $\widetilde{F} = \nabla v$. Then the gradient flow of u is the flow of F, and the gradient flow of v is the flow of $\widetilde{F}$. The flow lines of F are the level curves of v and the flow lines of $\widetilde{F}$ are the level curves of u. The orbits of F and $\widetilde{F}$ are mutually orthogonal.*

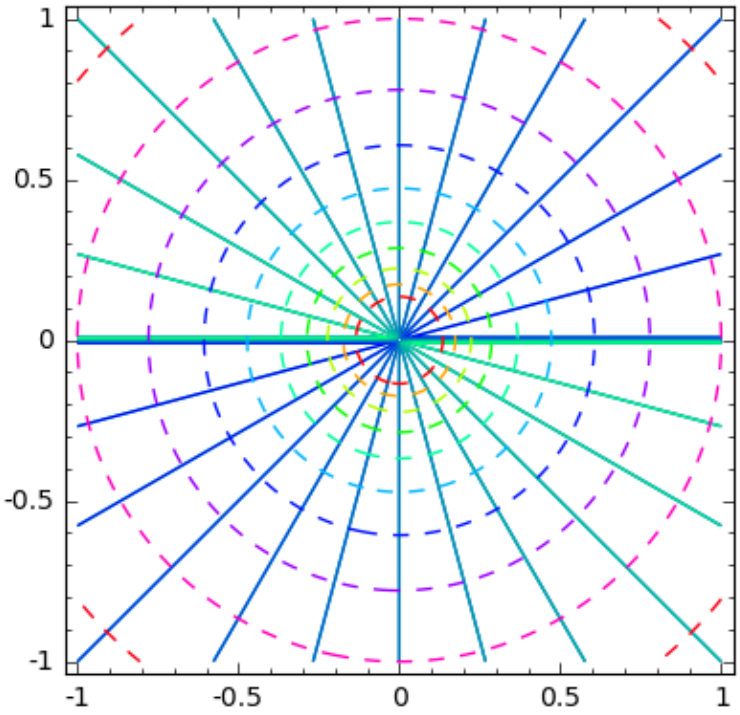


FIGURE 96. Equipotential curves and lines of force for the electrostatic field of a single charge in the plane. The equipotentials are level curves for the potential function φ and change color as the angle increases from 0 to π and again from π to 2π. This depends on the formula chosen for φ and the "color map" chosen for the graphics. In complex terms, the complex log function is $f(z) = \log(z)$ and for $z = re^{i\theta}$ with $\theta \in [0, 2\pi)$ then $f(z) = \log(re^{i\theta}) = \log(r) + \log(e^{i\theta}) = \log(r) + i\theta = u + iv$ with harmonic conjugates $u(x, y) = \log(r)$ and $v(x, y) = \theta$. We see the level curves in the Figure; they form a spiral staircase.

Example 51. Consider $f(z) = z^2 = u + iv$. The gradient fields are $F(\mathbf{v}) = A\mathbf{v}$ for

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

and

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

for the potentials u and v respectively.

36.11. **Electrostatic and gravitational fields in the plane and in $\mathbb{R}^3$.** The same geometry (with dual, orthogonal families of level curves) happens for electrostatic fields: one family is the *equipotentials* (curves or surfaces, depending on the dimension) while the other depicts the *lines of force*: flow lines tangent to the force vector field. See the Figures.

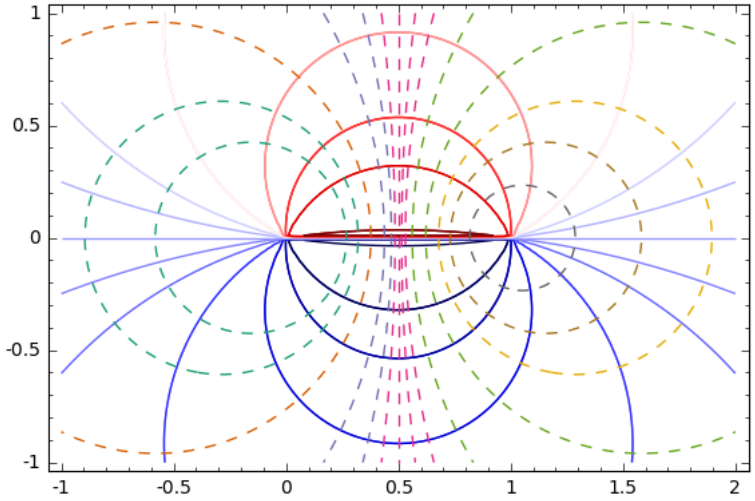


FIGURE 97. Equipotential curves and lines of force for the electrostatic field of two opposite charges in the plane. Colors indicate different levels of the potential and dual potential, where these are the harmonic conjugates coming from the associated complex function $g(z) = f(z) - f(z-1) = \log(z) - \log(z-1)$. These harmonic functions are $u(x,y) - u(x-1,y)$ and $v(x,y) - v(x-1,y)$.

When the opposite charges of Fig. 97 get closer and closer, the behavior approximates that of an *Electrostatic Dipole*; see Figs. 98, 102. The charges would cancel out, if we place one on top of the other, but if we take a limit of the fields as the distance d goes to 0 as charges c are balanced with this so that the product dc remains constant, then the limit of the fields (and potentials) exists. Note there is a limiting vector from plus to minus, along the x-axis. The picture is for the case of charges in the plane.

We note here that the *pictures* are unchanged by this sort of normalization, since:

Lemma 36.44.
(i) If F is a conservative field on $\mathbb{R}^n$ with potential function φ, then the collection of equipotential curves (or dimension $(n-1)$ submanifolds) is the same as for the field aF, $\mathbf{a} \neq 0$.
(ii) If γ is a line of force for F, then γ is orthogonal to each equipotential submanifold.

Proof. (i) We have: $\nabla\varphi = F$ iff $\nabla a\varphi = aF$, and the level curve of level c corresponds to that of level ac.
(ii) line of force for F is a curve γ with the property that $\mathcal{F}(\gamma(t) = \gamma'(t)$, i.e. γ is tangent to the field everywhere (is an orbit of the flow for the ODE). Then $\varphi(\gamma(t) = c$ so $\varphi(\gamma(t))' = 0$ but by the Chain Rule this is $\varphi(\gamma(t))' = \nabla\varphi(\gamma(t)) \cdot \gamma'(t) = F(\gamma(t)) \cdot \gamma'(t)$.

□

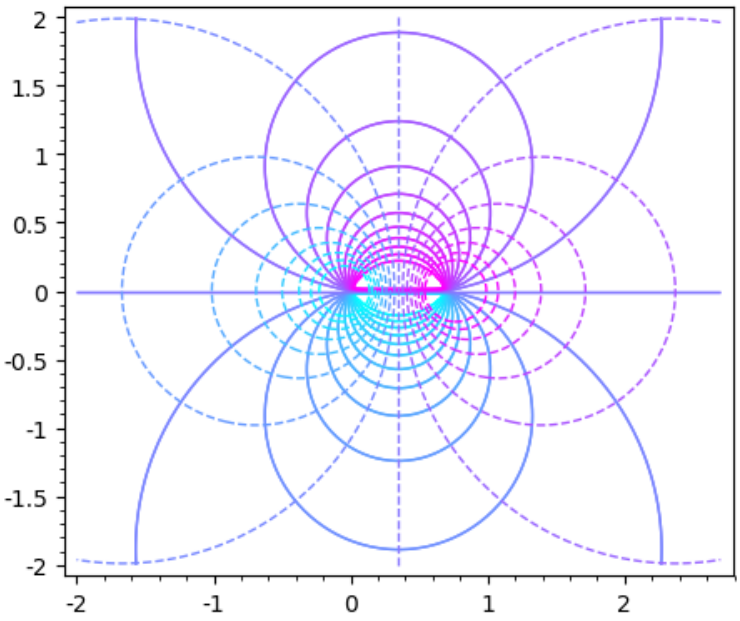


FIGURE 98. Equipotential curves and lines of force for the electrostatic field of two unlike charges, now closer together.

That the pictures converge (of both the equipotentials and field lines) looks clear from the figures, but to have the fields and potentials converge we need this normalization.

The potential function shown is

$$u(u,y) = \frac{1}{d}\log\frac{(x+d)^2+y^2}{(x-d)^2+y^2}$$

for $d = 1, .5, .05$.

Dipoles (both electric and magnetic) are useful in applications to electrical engineering and are itnriguing mathematically.

We mention that the geometry of fields in two-dimensional space has practical relevance: for example, the magnetic field generated by electric current passing throung a wire (in the form of a line) decreases like $1/r$, as we can think of the field as being in the plane perpendicular to the wire. For fascinating related material see the Wikipedia article on *Ampere's circuital law*.

Experiments show that the force between two charged particles with charges $q_1, q_2 \in \mathbb{R}$ with position difference given by a vector $\mathbf{v} \in \mathbb{R}^3$ is

$$\frac{q_1 q_2}{r^2} \cdot \frac{\mathbf{v}}{||\mathbf{v}||}, r = ||\mathbf{v}||$$

(so it is positive hence repulsive if the charges have the same sign).

An intuitive explanation for the factor of $1/r^2$ is this: suppose we have a light bulb at the origin and we want to calculate the light density at distance r; the light consists of photons, and the number emitted per second is the same as the number that pass

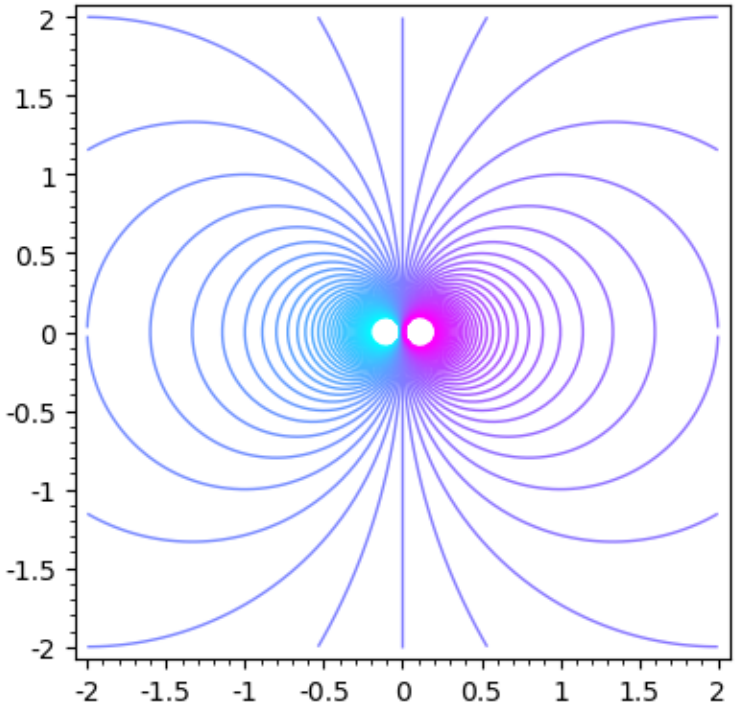


FIGURE 99. Equipotential curves for the electrostatic field of a planar dipole: two unlike charges close together.

through a sphere of radius r, which is proportional to the area $4\pi r^2$. Another way to say this that we are counting the number of field lines per unit area. Both the electrostatic field of a single charge and gravity (which is more simple as the is no negative gravity) are mediated by radiating particles and so should decrease in the same way.

We claim that the attractive potential φ of a single charge in $\mathbb{R}^3$ is

$$\varphi = 1/r = (x^2 + y^2 + z^2)^{-1/2}$$

Since the force field is then $F = \nabla\varphi$ we have $F = (P, Q, R)$ where

$$P = \frac{-x}{(x^2 + y^2 + z^2)^{3/2}}$$

and similarly for Q,R. The field strength at (x, y, z) is then

$$F(x, y, z) = ||(x, y, z)||/||(x, y, z)||^3 = 1/r^2$$

as we wanted.

We are thinking of a single large charge being tested by a small charge; we are not yet calculating the resulting field of two equal charges (or the gravitational field of two equal mass objects).

In two dimensions, the math is very different, as the field strength now should be proportional to $1/r$ as it is inversely proportional to the circumference of a circle, $2\pi r$.

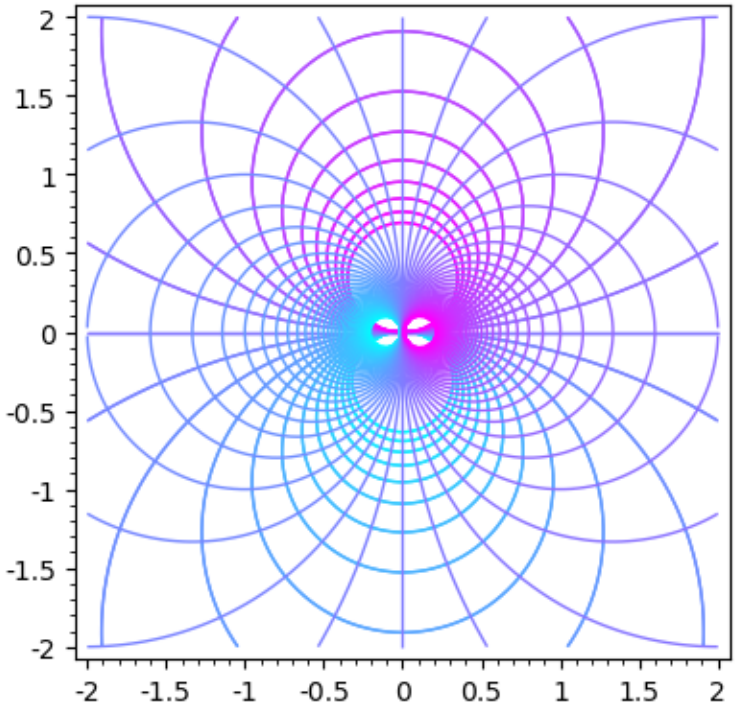


FIGURE 100. Equipotential curves and lines of force for the electrostatic field of a planar dipole: two unlike charges very close together. The potential is $1/d \log(((x-d)^2+(x+d)^2)$ for $d = 0.5$.

Thus in $\mathbb{R}^2$, for a single unit charge particle at the origin, we claim that the potential is

$$\varphi(x,y) = \frac{1}{2}\log(x^2+y^2)$$

for then the force field is

$$F = (P,Q) = \nabla\varphi = \left(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2}\right)$$

which has norm

$$||F|| = ||(x,y)||/||(x,y)||^2 = 1/r,$$

as we wished.

The dual field is

$$F^* = (-Q,P) = \left(\frac{-y}{x^2+y^2}, \frac{x}{x^2+y^2}\right)$$

which as we have seen in §36.5 has potential Θ, given by $\psi(x,y) = \arctan(y/x)$ or $\psi(x,y) = \operatorname{arccot}(x/y)$ depending on the location (since $\mathbb{R}^2\backslash\mathbf{0}$ is not simply connected).

The corresponding analytic function is

$$f(z) = \log(z)$$

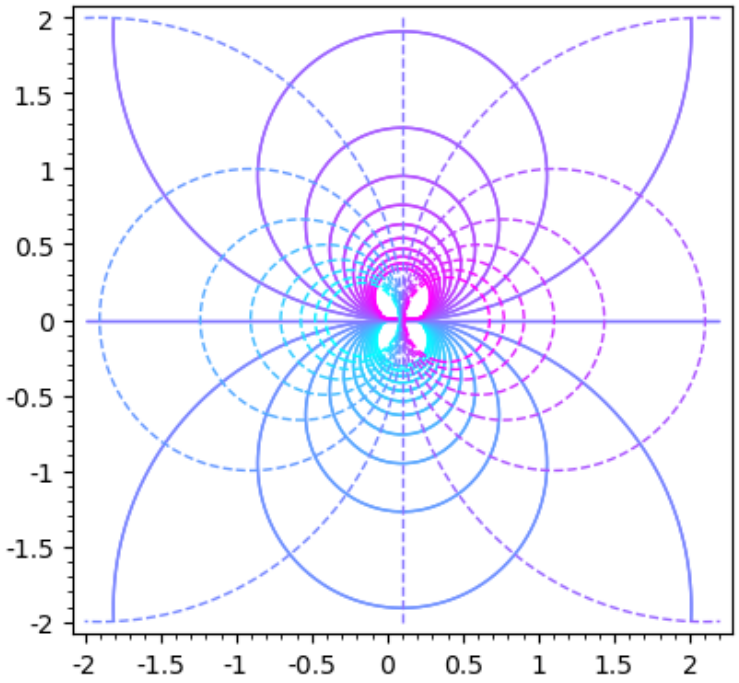


FIGURE 101. Equipotential curves and lines of force for the electrostatic field of an approximate planar dipole: two unlike charges close together.

and for $z = re^{i\theta}$ with $\theta \in [0, 2\pi)$ then $f(z) = \log(re^{i\theta}) = \log(r) + \log(e^{i\theta}) = \log(r) + i\theta = u + iv$ giving the harmonic conjugates $u(x, y) = \log(r)$ and $v(x, y) = \theta$, whose level curves we see in Fig. 96.

This is the case of a single charge. In fact, when combining objects all we have to do is add the two potentials, $\varphi = \varphi_1 + \varphi_2$, and then the gradient will give the field. See Figs. 97, 103 for the cases of two oppositely, and equally, charged particles.

That we sum the potentials means in two dimensions that we sum the associated complex functions as well; for opposite charges we change one of the signs.

In this figure, we have depicted two sets of curves: the level curves of the field φ (the equipotentials), and the flow lines of the gradient field $F = \nabla\varphi$ (the lines of force).

We can formulate this as a theorem; compare to Theorem 36.38 regarding analytic functions:

Theorem 36.45. *For an electrostatic field $F = (P, Q)$ on the plane, then P and Q are harmonic conjugates, whence*
(i)their gradient vector fields are orthogonal;
(ii) their families of level curves are orthogonal.

Further, the potential P and dual potential Q are (perhaps integral) linear combinations of the log *and argument (angle) functions on $\mathbb{R}^2$. The corresponding analytic functions are (integral) linear combinations of the complex log function.*

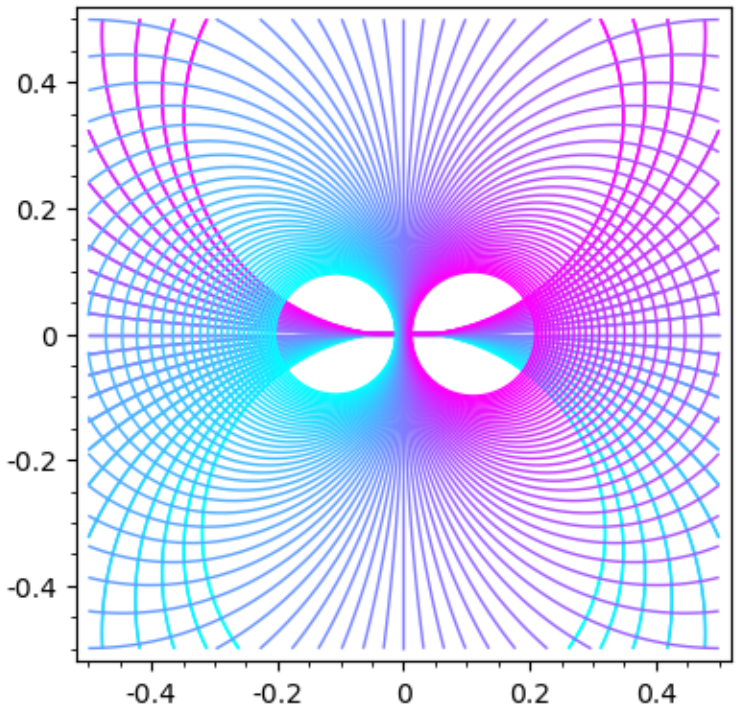


FIGURE 102. Equipotential curves and lines of force for the electrostatic field of an approximate planar dipole: two unlike charges close together.

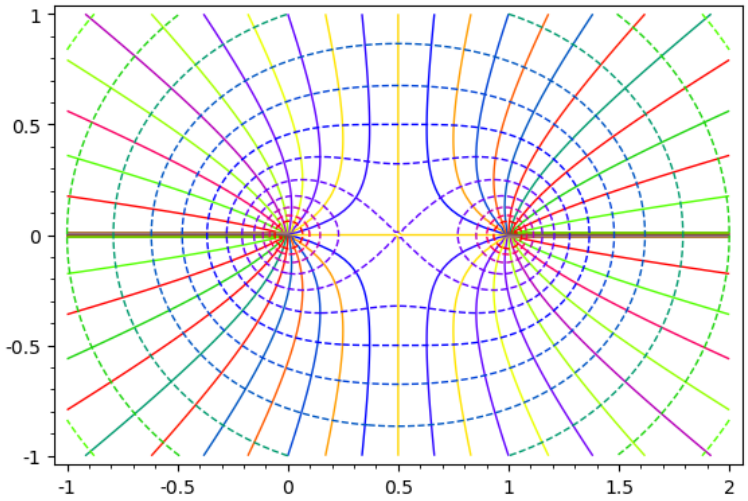


FIGURE 103. Equipotential curves and lines of force for the electrostatic field of two like charges in the plane. Since for one charge at $\mathbf{0}$ the associated complex function is $f(z) = \log(z) = u + iv$, here it is $g(z) = f(z) + f(z-1) = \log(z) + \log(z-1)$. The equipotentials and field lines are respectively the level curves for the harmonic conjugates $u(x,y) + u(x-1,y)$ and $v(x,y) + v(x-1,y)$.

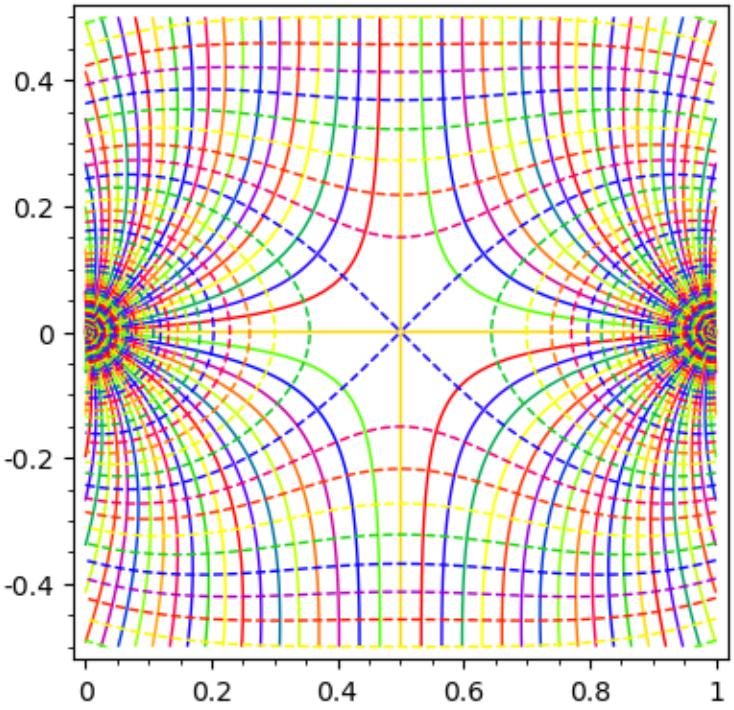


FIGURE 104. Equipotential curves and lines of force for the electrostatic (force) field of two like charges in the plane. Close to the center, $(1/2, 0)$, the potential and its dual start to approximate the dual hyperbolas of Fig. 93.

Proof. For a finite combination of point charges at points $\mathbf{p}_i \in \mathbb{R}^2$ with charges $q_i \in \mathbb{R}$, the associated analytic function on $\mathbb{C}$ is $f(z) = \sum q_i \log(z - z_i)$ where $\mathbf{p} = (x, y)$ corresponds to $z = x + iy$.

For a charge density given by a Riemann integrable real-valued function q, the associated analytic function is the vector-valued integral version of this (see §37.11):

$$f(z) = \int_{\mathbb{R}^2} q(w) \log(z - w) \mathrm{d}x\mathrm{d}y.$$

(The more general measure version of this also holds).

□

At first we may think that a potential such as the hyperbola shown in Fig. 93, cannot come from an electrostatic field. However as Feynman Vol II §7.3 [FLS64] points out, it can (in the limit): the field in the exact middle of two opposite charges of Fig. 97 looks just like this. See Figs. 104, 105.

Theorem 36.46. *For gravitational fields in $\mathbb{R}^2$, we have the same statement as Theorem 36.45 except that now only positive values of the density function q can occur.*

In fact, according to Feynman, any harmonic function and hence any complex analytic function can occur for a physical electrostatic field in $\mathbb{R}^2$. One can prove this as follows.

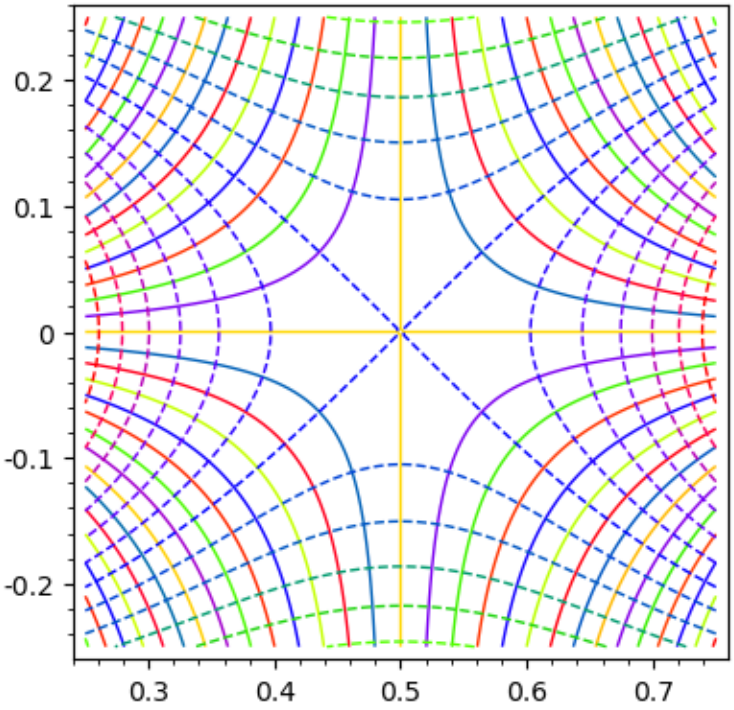


FIGURE 105. Equipotential curves and lines of force for the electrostatic field of two like charges in the plane. Close to the center, $(1/2, 0)$, the potential and its dual start to approximate the dual hyperbolas of Fig. 93.

From the mathematical point of view, there are two equivalent ways to characterize an electrostatic field (in $\mathbb{R}^2$ or in $\mathbb{R}^3$). The first is that the potential of the field is a solution of *Poisson's equation*,

$$\nabla^2(\varphi) = \rho$$

where ρ is a signed measure describing the distribution of charge. From this point of view one can then go about solving this linear partial differential equation. The second is to describe the *fundamental solution* , which is the single point charge, with its associated (gradient) field, and then define an electric field to be a (vector-valued integral) linear combination of such fundamental solutions, integrated with respect to the charge density.

From the first point of view what is fundamental is the PDE, from the second what is most basic is the fundamental solution (this is Coulomb's law!). What bridges the two is the *superposition principle*, which simply says the space of solutions is a vector space: we can take linear combinations.

In other words, for this linear equation knowing the fundamental solution characterizes the infinite-dimensional vector space of all solutions. And conversely, one of the methods for solving the PDE is to find its fundamental solution.

(For gravity the solution space is not all of the vector space but rather the positive cone inside of it).

Now $\nabla(\varphi) = F$ is the field, so Poisson's equation states that

$$\nabla^2(\varphi) = \text{div}(F) = \rho.$$

Thus from the field or the potential we can determine the charge distribution. Applying the operator ∇ is a type of derivative; the opposite procedure is a type of integration. Thus given the charge density ρ we find the field by solving the (partial) differential equation $\text{div}F = \rho$, and given the field we find the potential by solving the PDE

$$\nabla\varphi = F.$$

Combining these, given ρ we can find φ by solving the PDE

$$\nabla^2\varphi = \rho,$$

which is now a *second order* PDE as it involves second order partials.

The general operation of solving a DE is referred to as *integration*. As always, differention is automatic, while integration can be hard! Mathematically speaking, the first task is to prove that under certain circumstances a solution exists, and conversely trying to identify any *obstructions* to having a solution. Such obstructions are often especially interesting because they are topological; e.g. the equation $\nabla\varphi =$ only has a solution on a simply connected $\mathcal{U} \subseteq \mathbb{R}^2 \setminus \{\mathbf{0}\}$.

If there is no charge in a region $\mathcal{U}$, then from Poisson's equation

$$\nabla^2(\varphi) = \rho = 0$$

and the potential function φ is harmonic. Thus for Figs. 97, 97, the potential is 0 everywhere except exacly at those two points. At those points themselves the potential is infinite and the field is not only infinite but points in all directions, so neither is defined. When we have a continuous charge density, however, these are defined everywhere. In that case, by Poisson's equation the potential is not harmonic as $\nabla^2(\varphi) = \rho \neq 0$. When the charge density is continuous but nonzero, the field and potential make perfect sense mathematically being continuous functions, but the potential is no longer a harmonic function, so it certainly cannot (in
R^2) have a harmonic conjugate and does not extend to a complex analytic function. Hence the tools of Complex Analysis are not as applicable. Nevertheless, there is still a dual potential, whose level sets are orthogonal to those of φ, similar to the harmonic case.

To prove this, (I believe and would like to work this out!) we can again refer to the fundamental solution; since it hold there it must extend to all densities ρ.

But what "is" a point charge? From the mathematical point of view it is a point mass, simply a measure concentrated at a point. In physics this is called a *Dirac delta function*, which is the viewpoint of Riemann-Stieltjes integration. From the standpoint of Lebesgue integration, it is a measure and not a function at all.

Then we know how to rigorously treat two cases: point masses and continuous densities. Similarly, one can include densities given by any other Borel measures.

I say "density" rather than "distribution" here because that word will immediately get used in a very different way! That is the yet more sophisticated viewpoint of Laurent Schwartz' theory of distributions, see e.g. [Rud73]. Roughly speaking a Schwartz distribution is a continuous linear functional defined on a carefully chosen

space of *test functions* which are smooth and rapidly decreasing. This enables one to define derivatives, by duality. Thus the advantage of Schwartz distributions is that they can be differentiated and also can be convolved. Thus if one finds a fundamental solution to be a Schwartz distribution, the general solution is found by convolving this over the density. This is exactly what we have described above.

For the simplest case of the fields described above we can get away with point masses, but for more sophisticated examples we really do need Schwartz distributions. This is the case when we consider *dipoles*, but that is beyond the present scope.

For a clear overview of the physics, see the beginning of Jackson's text [Jac99]; this however goes quickly into much (much) deeper material, including boundary values, dipoles, Green's functions, and magnetism, dynamics and the connections with Special Relativity.

For a remarkable mathematical treatment see Arnold's book on PDEs: [Arn04].

Now to sketch a proof of Feynman's claim, given a harmonic function, we define a field to be the gradient of this potential. Given a field, we find such a potential. ...

36.12. **Parametrized surfaces.**

Definition 36.19. A *parametrized surface in* $\mathbb{R}^m$ is a $\mathcal{C}^1$ map $\sigma : \mathcal{U} \to \mathbb{R}^m$ where $\mathcal{U} \subseteq \mathbb{R}^2$. We write $S = \text{Im}(\sigma)$ for the image. This is the associated *un*parametrized surface, and is just a set of points.

Writing the coordinates as $\sigma(u, v) = (x, y, z)$ then we have the three coordinate functons $x, y, z : \mathcal{U} \to \mathbb{R}^m$. A key to understanding the surface is via its derivative; Recall that for $F : \mathbb{R}^n \to \mathbb{R}^m$ the $(m \times n)$ matrix of partial derivatives can be thought of, alternatively, as m gradients (the rows) or n tangent vectors (the columns). Explicitly, for $F = (F_1, \ldots, F_m)$ then the first row is ∇F_1, while for $\mathbf{p} = (x, x_2, x_3, \ldots x_n)$ then defining a curve by $F_{\mathbf{p}}(x) = F(x, x_2, x_3, \ldots x_n)$ the first column is the tangent vector

$$F'_{\mathbf{p}}(x) = \left[\tfrac{\partial F}{\partial x}\right]_{\mathbf{p}}$$

and so on.

So at a point $\mathbf{p} = (u_0, v_0)$ the derivative $D\sigma$ is

$$\begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} \end{bmatrix}_{\mathbf{p}} = \begin{bmatrix} \frac{\partial \sigma}{\partial u} & \frac{\partial \sigma}{\partial v} \end{bmatrix}_{\mathbf{p}} = \begin{bmatrix} \nabla x \\ \nabla y \\ \nabla z \end{bmatrix}_{\mathbf{p}} .$$

These columns $\frac{\partial \sigma}{\partial u}, \frac{\partial \sigma}{\partial v}$ will be important in three ways:
–they allow us to write the equation of the tangent plane as a paramterized plane;
–they allow us to write the equation of the tangent plane in general form;
–they allow us to calculate the surface area.

We define manifolds of dimension d in two main ways: *implicitly* i.e. as the set of solutions of some vector equation or *parametrically*, as described explicitly by the image of a function, from the space of parameters to the ambient space (i.e. the larger space in which our manifold is embedded).

In fact this dichotomy is motivated (as always!) by linear algebra. Recall that:

Proposition 36.47. *Let V, W be vector spaces and let $L : V \to W$ be a linear transformation. Then* $\ker(L) = \{\mathbf{v} : L(\mathbf{v}) = \mathbf{0}\}$ *and* $Im(L) = \{l(\mathbf{v}) : \mathbf{v} \in V\}$ *are subspaces. If L has maximal rank, then*

implicitly as the set of points where a function assumes a certain value. Examples are the circle $\{x^2 + y^2 = 1\} \subseteq \mathbb{R}^2$ or sphere $\{x^2 + y^2 + z^2 = 1\} \subseteq \mathbb{R}^3$; the first is a level curve of the function $F(x, y) = x^2 + y^2$ and the second a level surface of $G(x, y, z) = x^2 + y^2 + z^2$. In general, for $F : \mathbb{R}^{d+11} \to \mathbb{R}$, then the inverse image of a point has dimension d. When this will indeed give a smooth manifold is guaranteed by the *Implicit Function Theorem*, which is more general, allowing for the inverse image of submanifolds. See e.g. [HH15] §2.10 or [War71] for good treatments of this key result.
as a graph of a function;
as a parametrized manifold. For this
Finding a primitive.

36.13. **Least action.** $E_{\text{pot}} = -\varphi$.
work done (in any field) is

$$\int_\gamma F \cdot \mathrm{d}\gamma = E_{\text{kin}}(b) - E_{\text{kin}}(a).$$

in a conservative field, we also have a second expression for this: the work done is

$$\int_\gamma F \cdot \mathrm{d}\gamma = \varphi(B) - \varphi(A) = E_{\text{pot}}(A) - E_{\text{pot}}(B).$$

Grad Div curl. (1)Div 0: flow preserves volume; Pf in linear case (trace A) and flow (det 1); flux 0 for closed loop (Stokes Thm -Div Thm) d of what??).
(2) curl 0 iff locally Conservative iff circulation 0 iff flow is gradient flow; d of potential
Conservative implies circulation and curl zero; converse holds locally
exs of curl 0 rotations
potential for rotation
TO DO: curl, div independent; both equal for DF; pure gradient flow vs rotation flow; most are mixture; sum in LA= composition of flows; curl corresponds to what in LA for higher dim??? div= trace always...
harmonic conjgs;dual grad-rot flow.
The Div Curl of F on R^2 comme from $DF|_{\mathbf{p}}$ matrix; hence same as for linearisation at that point.
Linear flow pres vol iff Div= trace=0.
Linear flow: F is conserv iff curl =0 iff gradient flow iff perp to level curve flow of exact ODE ; linear + exists potential....can we find it? gradient flow means what???
Finding a potential
Next we see (by working out some examples) how to find the potential of a conservative vector field.
Divergence zero implies zero implies flux volume-preserving flow; linear case, refernce to general; converse holds locally

We know that given $F : \mathbb{R}^n \to \mathbb{R}$, the gradient vector field is orthogonal to the level hypersurfaces (submanifolds of dimension $n-1$) of F, so level surfaces in $\mathbb{R}^3$ and level curves in $\mathbb{R}^2$.

We know that a vector field is conservative iff it is the gradient of a function. There are two ways that this can fail to be the case: locally or globally.

We want to first examine the local problem: when is a vector field locally conservative?

Switching equivalently to the language of differential forms, the vector field V is conservative iff the associated $1-$form η is *exact*. We know that a necessary condition for this to occur is that the form be *closed*, i.e. that $d(\eta) = \mathbf{0}$. In $\mathbb{R}^2$ or $\mathbb{R}^3$ this is the same as curl $= \mathbf{0}$.

Poincaré's Lemma tells us that locally, the converse holds: any closed form is exact. A basic counterexample for the global exactness is the angle function on the plane: there is a local potential (the infinite spiral staircase) but this is a multivalued function, so not a potential in the usual sense.

Here is a method to try to find a potential for any vector field in the plane. Given a nowhere-$\mathbf{0}$ vector field V, we want to find a potential φ, that is a function $\varphi : \mathbb{R}^2 \to \mathbb{R}$ such that $\nabla\varphi = V$. In this case, its level curves are orthogonal to V. So, let us consider the orthogonal vector field W to V, say at angle $+\pi/2$. Then, using the Fundamental Theorem of ODEs, draw the integral curves. These are unique hence do not intersect. Globally, they might say be spirals, we can define a function φ with different values on each. Thus, φ is a candidate for a potential.

We can see an example in theillustrations of the electrostatic potentials Fig. 103, 97.

There are two families of curves: the equipotentials and the lines of force.

The lines of force are tangent to the gradient vector field. For opposite charges, we can picture the gradient flow as flowing from the positive to the negative charge. In fact, we can interpret this as a gravitational field, with a mountain at the positive and a valley at the negative charge. For like charges, we can picture two mountains.

It is important to remember that there are two quite different interpretations, as force fields or as velocity fields. The gradient flow refers to the velocity field, and a particle moves along the curve with that tangent vector. For the force field interpretation, the particle accelerates and may go off the curve because of the acceleration due to the curvature.

In any case, we can try to imagine switching roles, so the equipotential curves become the orbits of a gradient flow and vice-versa.

If this works, we will have succeeded in constructing a potential for our vector field.

But what can go wrong with this intuitive idea?

–integrating factor (always exists in plane)

—examples

–doesn't always exist in $\mathbb{R}^n$: Frobenius theorem

37. Minicourse on Ordinary Differential Equations: From flows to vector fields and back again.

TO DO: [Tay96] Taylor PDE Vol I

In this section we present an introduction to the classical theory of differential equations encountered in undergraduate math courses, in one and higher dimensions, based on the linear algebra just covered, especially the look at the exponential map and linear flows.

Though we focus here on $\mathbb{R}^n$, all this holds for differential equations on a smooth manifold M, where the vector field V defining the equation is a function from M to the tangent bundle TM. Equivalently, V is a *section* of the tangent bundle; the passage to manifolds is made in the usual way, via charts.

Given a smooth flow τ_t on a manifold M, our intuition is that the time derivative at $t = 0$ is a vector field V. The converse operation (given a vector field, can we find such a flow?) is called "solving an ODE". The flow orbit $\tau_t(\mathbf{v})$ of a point $\mathbf{v}$ is exactly the solution curve of the differential equation. But is this naive intuition correct?

In essence, *yes*! as we shall explain.

Now the orbit of a flow defines a curve by

$$\tau_t(\mathbf{v}) = \gamma(t).$$

Taking the time derivative gives the tangent vector $\gamma'(t)$. Evaluating at time 0 defines a vector field, by

$$V(\mathbf{p}) = \gamma'(0).$$

Conversely, from the vector field we derive the *vector differential equation*

$$\gamma(t) = V(\gamma(t)); \;\; \gamma(0) = \mathbf{v}.$$

We say this is a (vector) differential equation with *initial condition* $\gamma(0) = \mathbf{v}$. A *solution* of the DE is such a curve $\gamma_{\mathbf{v}}(t)$. The *Fundamental Theorem* of ODEs states that such a solution exists and is unique.

Vectors can have a variety of interpretations. The two most important are that $\mathbf{v}$ represents *displacement* i.e. movement, and that it represents *force*. A third important interpretation comes from magnetism, where the vector product is involved. All three of these are very different and must not be confused!

These interpretations pass over to vector fields. In the first case, V is interpreted as a *velocity vector field*, in the second as a *force field* like for gravity or electric charge (electrostatics). The third could be a magnetic field.

The point we wish to make is that in all cases, perhaps by augmenting the dimension, we can take the interpretation of a velocity field.

This is what the above equation

$$\gamma(t) = V(\gamma(t))$$

says: the vector field specifies what the *velocity*: the tangent vector $\mathbf{v}(t) = \gamma'(t)$, must be at a point $\gamma(t)$.

For a *force field* we mean the following. The DE is now a *second-order* ODE as it involves the second derivative: $F(\gamma(t)) = m\gamma''(t$. This expresses Newton's Law $F = m\mathbf{a}$, where $\gamma(t)$ is the position, $\gamma'(t) = \mathbf{v}(t)$ the velocity and $\mathbf{a}(t) = \gamma''(t)$ the acceleration, and $m \geq 0$ is the mass of the object. of the object. Here we need two initial conditions, $\gamma(0)$ and $\gamma'(0) = \mathbf{v}(0)$, and out Fundamental Theorem guarantees that we will again have a unique solution.

The first key point to mention is that a vector differential equation on $\mathbb{R}^d$ is equivalent to a *system* of d one-dimensional equations. This is just like a system of linear equations in Linear Algebra being equivalent to a single matrix equation. In particular, for a linear vector field, that is, one given by a matrix, this is what is called a *system of linear first-order homogenous equations.* A second point is that a *higher-order differential equation* can be rewritten as a system of first-order equations. A third point is how we can treat in a similar way a *nonstationary* vector field. This is a parametrized family $(V^t)_{\mathbb{R}}$ of vector fields on $\mathbb{R}^d$ which is continuously varying in time. This leads to the notion of a *nonstationary flow*, see below, and in terms of differential equations, a *nonstationary. nonautonomous* or *time-varying* ODE. In fact, the nonstationary case can be treated as stationary in one more dimension, adding the variable t, thus giving a vector field (and actual flow) on $\mathbb{R} \times E$. See below.

Thus all ODEs can be treated as stationary first-order DEs and so as velocity vector fields.

Now as described above, we can move seamlessly from a flow to a vector field and back again. The "back again" is known as *solving'* or *integrating* the differential equation; the solutions are put together to get the flow.

These are therefore another case of the complementary operations from Calculus of differentiation and integration. Indeed γ_v is known as an *integral curve* of the differential equation, and the flow is given by *integration of the vector field.*

Thus, the main theoretical theorem one needs is the existence and uniqueness of solutions for a vector DE of first order. A solution is the curve given an initial condition, and from a wider viewpoint, the flow which has those orbits. The method of proof involves iterating a linear operator on the space of paths (the Picard operator); choosing an initial value, this gives an eventual contraction, and the solution is the unique fixed point. Using this operator one can prove part of the above statement: these are the orbits of a continuous flow.

To nail down our original intuition, it remains to verify that the flow is not only continuous and differentiable, but that it is smooth in the space as well as time directions, and that the derivative natrix is the space derivative of the vector field.

There are three derivatives here: first, the time derivative along an orbit (the tangent vector) should be the vector field; this is just the above statement of the differential equation. Next, the time derivative of the flow should be a matrix; the spatial derivative of the vector field should be a matrix, called the *linearization* of the vector field. An intriguing question then is what is the relationship between these last two. A careful examination of this takes one into quite different areas: dynamical systems, differential geometry as well as classical ODE.

37.1. **Differential equations in one dimension.** See also Definition 37.2 below:

Definition 37.1. For $n \geq 0$, given a continuous function of $(n+1)$ variables, $F : \mathbb{R}^{n+1} \to \mathbb{R}$, so $F = F(x_0, \ldots, x_n) \in \mathbb{R}$, then an *autonomous* or *stationary differential equation in one dimension* is an equation of the form

$$(*) \qquad F(y, y', \cdots, y^{(n)}) = 0$$

where F is applied to a function $y : \mathbb{R} \to \mathbb{R}$ that we are trying to find together with its derivatives $y^{(n)}$. (Since $n + 1 \geq 1$, at least one derivative is involved.)

To introduce the time-varying case, we begin with a parameterized family $(F_t)_{t\in\mathbb{R}}$ of such functions, as that will correspond exactly to a parameterized family $(V^t)_{\mathbb{R}}$ of vector fields.

Now given a continuous family $(F_t)_{t\in\mathbb{R}}$, setting $\widetilde{F} : \mathbb{R}^{n+2} \to \mathbb{R}$ where $\widetilde{F}(t, x_0, \ldots, x_n) = F_t(x_0, \ldots, x_n)$ then these are equivalent:

$$F_t(x_0, \ldots, x_n) = 0$$

for all t and

$$\widetilde{F}(t, x_0, \ldots, x_n) = 0.$$

Taking this second as the definition, then by a *nonautonomous*, *nonstationary* or *time-varying* DE we mean:

$$(*) \qquad F(t, y, y', \cdots, y^{(n)}) = 0,$$

where F is now a continuous function of $(n + 2)$ variables.

The above DEs (*) are said to be **order** n.

Often a DE is in **explicit form** or **in normal form** of order n when the highest-order derivative occurs separately, i.e. if it is in the form

$$(*) \qquad y^{(n)}(t) = F(t, y, y', \cdots, y^{(n-1)})$$

where now F is a given continuous function of $(n + 1)$ variables. Otherwise it is an **implicit** DE.

The Implicit Function Theorem gives conditions when given an equation

$$F(x_1, \ldots, x_n) = 0$$

we can solve for one of the variables, so for instance $x^2 + y^2 = 1$ becomes $y = \pm\sqrt{1 - x^2}$. Given an implicit DE one tries to turn it into an explicit equation by solving for the highest order derivative $y^{(n-1)}$, though sometimes one can solve it in the implicit form (see the discussion below of *exact equations*. In that case the solution itself may be given in implicit form and we can try to solve for the function $y(t)$.

In any case in courses DEs are often encountered already in explicit form.

Here are some examples:

(1)For $a : \mathbb{R} \to \mathbb{R}$ continuous, $y' = a(t)$ is an explicit equation, nonautonomous unless $a(t) = a$ is constant. The solution is just the antiderivative of $a(t)$.

(2)$F(r, s) = s$, so $y' = F(t, y) = y$: $y' = y$. This is autonomous. It is one of the most useful DEs. In theoretical terms it can be used to define the exponential function (this definition relies on knowing the theorem on existence and uniqueness of solutions of DEs), and since a second definition of e^t or e^z goes by way of power series, we can turn that around and study complex and matrix generalizations of that equation. Of course its importance in applications is that this equation models exponential growth, leading to many modifications aimed at producing more accurate models.

(3)For $a, b \in \mathbb{R}$, $F(r, s) = a \cdot s + b$, so $y' = ay + b$. This is called an *autonomous linear* DE.

(4) For $a, b : \mathbb{R} \to \mathbb{R}$ continuous, $F(r, s) = a(t) \cdot s + b(t)$, so $y' = a(t)y + b(t)$. This is termed a *linear* first-order DE as it is given by an affine operator on function space, nonautonomous unless a, b are constant functions.

Systems of equations: geometric definition.

In a few words, a *system* of real DEs is, geometrically, the same thing as a vector field V on $\mathbb{R}^d$, with the solutions describing the integral curves. We begin here as the analytic definition via a formula can at first be hard to digest.

Given a vector field V on $\mathbb{R}^d$ (first in the stationary case) we describe the passage to a system of DEs in the analytic sense. After that we give the precise definition of the latter.

Writing $E \equiv \mathbb{R}^d$, we express $V : E \to E$ in coordinates: $V = (V_1, \ldots, V_d)$ and for each k, $V_k = V_k(x_1, \ldots x_d)$.

A curve γ is rewritten $\gamma = (y_1, \ldots, y_d)$, so $\gamma' = (y_1', \ldots, y_d')$ and $\gamma'(t) = (y_1'(t), \ldots, y_d'(t))$.

Then our stationary equation $\gamma'(t) = V(\gamma(t))$

becomes the d simultaneous equations

$$\begin{cases} y_1' = V_1(x_1, \ldots, x_d) \\ \vdots \\ y_d' = V_d(x_1, \ldots, x_d) \end{cases}$$

or more fully:

$$\begin{cases} y_1'(t) = V_1(x_1(t), \ldots, x_d(t)) \\ \vdots \\ y_d'(t) = V_d(x_1(t), \ldots, x_d(t)) \end{cases} \tag{137}$$

This is an example of an autonomous system of first-order DEs.

Systems of equations: analytic definition.

We have essentially arrived at the correct definition above. Thus, suppose we have unknown functions y_1, y_2 of t. If we have two equations

$$\begin{cases} y_1' = F_1(y_1, y_2) \\ y_2' = F_2(y_1, y_2) \end{cases}$$

this is a *system of differential equations*, in this case a system of two first-order autonomous equations. Similarly,

$$\begin{cases} y_1'' = F_1(y_1, y_2, y_1', y_2', t) \\ y_2{}'' = F_2(y_1, y_2, y_1', y_2', t)) \end{cases}$$

is a system of two nonautonomous second-order equations.

Note that the functions F_i involve both y_1 and y_2. Hence we should really think of these as given by a single vector function.

Defining in the above case $F = (F_1, F_2)$ then F is a continuous function of five variables, $F : \mathbb{R}^5 \to \mathbb{R}^2$.

Note that in the first-order case this takes us exactly to the system we defined from a vector field.

To complete the circle of these ideas it remains to see how any higher -order system can in fact be replaced by an equivalent system of first-order equations, which in turn can be represented as a vector field.

The idea is simple: $(y, y', y'', \ldots, y^{(n)})$ in an n^{th}-order equation is replaced by $(w_1, w_2, \ldots, w_n)$, thus $w_1 = y, w_2 = y'', \ldots, w_n = y^{(n)}$.

Basically to define higher-order systems we just include more variables; thus for order 2 we have d equations in the $2d$ variables w_k, w_k' and so on.

Exercise 37.1. (Harmonic oscillator)

For a simple but important example, which we return to in Exercise **??**, we have the harmonic oscillator equation

$$y'' = -y$$

The idea is that a spring is attached to a wall and the other end to an object; when this is pulled out to a distance y the force felt is approximately $-cy$, where $c > 0$ is a constant called the *spring constant*. The reason for the $-$ sign is that it is being pulled back toward its rest position 0: in the negative direction if if $y > 0$, in the positive direction if $y < 0$, changing as it oscillates.

Ir is clear that (taking for simplicity $c = 1$) $y(t) = \sin t$ and $\cos t$ are solutions.

We show how this second-order equation in one dimension can be recoded as a system of two one-dimensional first-order equations, and equivalently as a vector DE in $\mathbb{R}^2$, given by a linear vector field.

For this, we set $w_1 = -y, w_2 = y'$. We thus have the pair of equations $w_2' = y'' = -y = w_1,$, $w_1' = -y' = -w_2$ giving the system

$$\begin{cases} w_1' = -w_2 \\ w_2' = w_1 \end{cases}$$

This can be written in matrix form, where $\mathbf{w} = (w_1, w_2)$, as $\mathbf{w}' = A\mathbf{w}$ where $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

Note that A defines a linear vector field on $\mathbb{R}^2$. It important to note that the variable for time does not occur here, as this is an autonomous second-order equation and hence an autonomous system of two first-order equations; the variables $(w_1, w_2) = (y, y')$ represent, for the oscillator, position and velocity (or momentum). Thus e.g. the vector solution $\mathbf{w}(t) = (\cos t, \sin, t)$ gives the one-dimensional solution $y(t) = \cos t$ for the original equation; the graph of the curve $\mathbf{w}(t)$ is the spiral $(t, \cos t, \sin t)$ which projects to the position $y(t) = \cos t$ and velocity $y'(t) = \sin t$.

The matrix for the general n^{th}-order linear case

$$y^{(n)} = a_1 y + a_2 y' + \cdots + a_n y^{(n-1)}$$

then has the nice form

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & \\ \vdots & & & & \\ 0 & 0 & 0 & \dots & 1 \\ a_1 & a_2 & a_3 & \dots & a_n \end{bmatrix}.$$

37.2. **Ordinary differential equations: The classical one-dimensional case.** The review in the previous section of linear algebra and in particular of upper triangular and diagonal forms prepares us for the study of vector ordinary differential equations, equivalently sistems of one-dimensional ODEs, equivalenly flows in $\mathbb{R}^n$.

Here we review some basics regarding this one-dimensional situation i.e. differential equations on the real line, which will then be combined with the linear algebra in the following sections.

For this classical one-dimensional case there are many references, in books and online, which go into a wealth of examples and detail, as the study of such examples plays an important part in applications to physics, biology economics, and engineering. Our aim here is to get a first taste of this extensive body of knowledge, before moving on to our main interest, the higher-dimensional case.

Our notes for this section are based on course notes by Marina Talet, Université Aix-Marseille, 2021.

1. Introduction:

Consider an integration exercise from Calculus such as: given $f(x) = 1/x$, defined on $\mathbb{R} \setminus \{0\}$, find

$$F(x) = \int f(x)\mathrm{d}x$$

with the solution

$$F(x) = \log|x| + c$$

The function F has several names: the *antiderivative*, *integral*, or *primitive* or *indefinite integral* of f, and the operation of finding F is termed *integration* of the function f.

We can rewrite this as: find $y = y(x)$ where

$$y' = f$$

which can be considered as the simplest type of differential equation. If we specify that $y(1) = 0$, this *initial condition* fixes the solution on the interval $(0, +\infty)$, as then $c = 0$.

The explanation for the integral being *indefinite* in that it is defined up to a constant c is that the derivative map $D : \mathbb{C}^{k+1} \to \mathbb{C}^k$ is a linear transformation with one-dimensional kernel the constant functions.

The general concept of *integrating* or *finding a primitive* goes far beyond this.

We find further examples in what follows.

Exponential growth.

Let us recall the two main rules for exponents:

(i) $a^{b+c} = a^b a^c$

and

(ii) $a^{bc} = (a^b)^c$.

An *exponential function* is of the form $f(x) = a^x$ for $a > 0$ and $x \in \mathbb{R}$. The number a is termed the *base* and x the *exponent*. By the above rules, $1/a = a^{-1}$ and $(a^{\frac{1}{2}})^2 = a$ whence $\sqrt{a} = a^{\frac{1}{2}}$. This makes it easy to define a^x for exponent rational.

Thus exponentiation turns adddition into multiplication, multiplication into taking powers. Mathematically speaking, the fact that $a^{x+y} = a^x a^y$ and $a^0 = 1$ tells us that writing $\Phi_a(x) = a^x$, the map Φ is an isomorphism from the additive group of real numbers $(\mathbb{R}, +)$ to the multiplicative group of positive real numbers $(\mathbb{R}^> \cdot)$ where $\mathbb{R}^> \equiv (0, +\infty)$.

We write $\ln_a(x)$ for the inverse function of base a, that is, for $f(x) = a^x$ and $g = \ln_a(x)$ then $f \circ g(x) = x$ and $g \circ f(x) = x$ wherever these are defined: the first for $x > 0$ and the second for all $x \in \mathbb{R}$.

This function does the opposite of the exponential: it maps $\mathbb{R}^> \equiv (0, +\infty)$ to $(\mathbb{R}, +)$ and converts multiplication to addition and powers to products, thus

$$\ln_a(xy) = \ln_a(x) + \ln_a(y)$$

and

$$\ln_a(x^y) = y \ln_a(x).$$

The most practical base in many applications is 10 or 2, but in pure mathematics by far the most important base is the irrational number $e = 2.1714...$ The reason is that the formula for the derivative is much simpler for base e, as we shall see shortly.

But first, to define e^x for real, non-rational exponents, we note that there are several approaches one can take.

(1)First, we can use continuity: it can be proved that there is a unique continuous way to extend this function to the reals. That is, we can approximate x by rational numbers and take the limit.

We can also give more explicit definitions.

(2) As we see below, there exists a unique solution to the *differential equation* (or DE) $y' = y$ satisfying $y(0) = 1$ (this is called an *initial condition* for the DE). We define this function to be $\exp(t) = y(t)$, and then and define the number e to be $\exp(1)$. This is also the slope of e^x at $x = 0$.

(3) We define e^x to be the function with the following series expansion:

$$\exp(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots + \frac{x^n}{n!} + \ldots$$

where the factorial is defined by $0! = 1$, $k! = 1 \cdot 2 \cdot \cdots \cdot k$.

This expresses e^x as a limit of rational numbers. In particular, the number $e = e^1$ can be approximated as a decimal using this series.

One proves in Calculus:

Theorem 37.1. *This series converges for all $x \in \mathbb{R}$.*

Note that the derivative of the series, taken term-by-term, does satisfy the DE $y' = y$, $y(0) = 1$, so (3) yields (2). Conversely, knowing the derivative from (2) gives (3) as the Taylor series: recall that for an infinitely differentiable function f the Taylor series about 0 is

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(0)x^k}{k!}.$$

We let $\ln(x)$ denote the *natural logarithm*, the inverse function $g(x)$ of $f(x) = e^x = \exp(x)$, so $\ln = \ln_e$.

Another way to define ln is via integration: for $x > 0$,

$$\ln(x) = \int_1^x \frac{1}{x} \mathrm{d}x$$

Another possible definition for ln is via its Taylor series, calculated around the value $x = 1$, in other words, find the series for $\ln(x + 1)$ around 0.)

This leads to a third definition of exp:

(4) First we define ln, in one of these ways; then exp is defined to be its inverse function.

Next we define, for any base $a > 0$:

$$a^x = e^{(\ln a)x}.$$

Hence the derivative is $(a^x)' = \ln(a)a^x$. Note that indeed $(e^x)' = e^x$, and the number e is the only base such that a^x is its own derivative.

We then denote by $\ln_a$ its inverse function.

Exponential growth and doubling times; exponential decay and half-life

Suppose a quantity $f(n)$ doubles every day, starting at 1 at time $n = 0$. Then we have $f(n) = 2^n$. (You should draw the graph, for say $n = -3, \ldots 3$). Here the *doubling time* is 1.

When we first see this equation, we naturally wonder why not to use base 2 (or perhaps 10!) instead of the irrational number $e = 2.1714\ldots$. The reason is because base e has the simplest expressions for the derivative, hence also for the series. In fact, for both calculations and theory, for exponential and also for logs, it is generally easier to first change to base e.

Nevertheless the concept of doubling time is intuitively very useful.

When a^t for $a < 1$ we call this *exponential decay*, for example the decay of radioactivity of a substance.

When we know the doubling time of for instance a pandemic or a bank account, we can easily make rough estimates in our heads, and similarly for exponential decay of a radioactive substance.

These can vary considerably, ranging from 4.4 billion years for Uranium-238 to 10^{-24} seconds for Hydrogen-7. Plutonium-239 has a half-life $24,110$ years, indicating its danger when in radioactive waste, while Carbon-14 which is so useful in the radiocarbon dating process used by archeologists has a half-life of $5,730$ years.

Exercises:
(1) Show that $\ln_a(x) = \ln(x)/\ln(a)$.
(2) Find the Taylor's series for $\ln(x+1)$ about $x = 0$.
(3) Prove that the series for e^x converges for all $x \in \mathbb{R}$.
(4) Find a formula for the doubling time t_d for $f(t) = e^{at}$, for $a > 0$.
(5) Find the half-life t_h half-life for $f(t) = e^{at}$, for $a < 0$.

Convergence of exponential function.

(3) Prove that the series for e^x converges for all $x \in \mathbb{R}$.
Solution:

Proof. For x fixed, let $m > 2x$ so $x/m < 1/2$. Then for any $n > 0$,

$$\frac{x^{n+m}}{(n+m)!} \leq \frac{x^m}{m!} \cdot \left(\frac{1}{2}\right)^n$$

which gives a geometric series hence converges. Thus the sequence of partial sums is an increasing bounded sequence hence converges by the completeness property of the real numbers. □

Exercise 37.2. The series

$$\exp(z) = 1 + z + \frac{z^2}{2} + \frac{z^3}{6} + \cdots + \frac{z^n}{n!} + \ldots$$

comverges for all $z \in \mathbb{C}$.

Exercise 37.3. The series

$$\exp(A) = 1 + A + \frac{A^2}{2} + \frac{A^3}{6} + \cdots + \frac{A^n}{n!} + \ldots$$

comverges for all $A \in \mathcal{M}_{(d\times d)}$, the collection of square matrices (with entries in $\mathbb{R}$ or $\mathbb{C}$.)

Solving the equation $y' = ay$, $a \in \mathbb{R}$.

Exponential growth $y(t) = A^t$ for $A > 1$ grows at a rate proportional to the quantity at time t. Thus for example for $A = 2$, $2^{n+1} - 2^n = 2^n(2-1) = 2^n$, while for a bank account growing at 10% per year, $a = 1.10$ and $y(t+1) - y(t) = A^{n+1} - A^n = A^n(A-1) = c \cdot y(n)$ for the constant $c = A - 1$.

As noted, this includes both exponential growth or decay, and also the constant case $y' = 0$.

Method 1:

We simplify to the special case $a = 1$ and recall that $y(t) = e^t$ solves this. Then we see that $y(t) = Ke^t$ for $K \in \mathbb{R}$ also works. Lastly we note that $y(t) = Ke^{at}$ will provide a solution of the DE $y' = ay$, for any $a \in \mathbb{R}$.

This is valid for any $a, K \in \mathbb{R}$.

But are these all possible solutions? To answer this let $v(t) = e^{at}$ and suppose that u is another solution, so $u' = au$. Now since $v(t) = e^{at}$, $v^{-1} = e^{-at}$ whence $(v^{-1})' = -av^{-1}$.

We guess that $u = Kv$ is the only possibility, thus that u/v will be a constant. Equivalently, its derivative is 0. We compute:

$$(u/v)' = (u \cdot v^{-1})' = u'(v^{-1}) + u \cdot (v^{-1})' = auv^{-1} - auv^{-1} = 0$$

as we guessed, so $u/v = K$ and

$$u = Kv = Ke^{at}.$$

Method 2:

From the equation $y' = ay$,

we have that

$$\frac{y'}{y} = a.$$

We recognize this as a *logarithmic derivative*: that is, for a function f with positive values,

$$\ln(f)' = \frac{f'}{f}.$$

More generally, recall that $\ln|t|' = 1/t$ is valid for all $t \neq 0$. (Check this in two ways: from the formulas, and by drawing the graphs!)

Thus in fact for any f with no zero values, $(\ln|f|)' = \frac{f'}{f}$ is true.

So for $y = y(t)$ never zero we have

$$(\ln|y|)' = \frac{y'}{y} = a$$

.

We know our solution $y(t)$ is differentiable hence continuous. Therefore there are two cases, since $y(t) \neq 0$: either $y(t) > 0$ for all t, so $|y| = y$, or $y(t) < 0$ for all t, so $|y| = -y$,

Integrating in the first case,

$$\ln y + c_1 = \int \ln(y(t))' \mathrm{d}t = \int a \mathrm{d}t = at + c_2$$

so in summary

$$\ln y = at + c$$

and equivalently

$$y = e^{at+c} = Ke^{at}$$

where we define $K \equiv e^c > 0$.

In the second case,

$$(\ln|y|)' = \frac{y'}{y} = a$$

and $|y| = -y$ so

$$\ln(-y) + c_1 = \int \ln(|y(t)|)'\mathrm{d}t = \int a\mathrm{d}t = at + c_2$$

so

$$\ln(-y) = at + c$$

and equivalently

$$-y = e^{at+c}$$

$$y = -e^{at+c} = -Ke^{at}$$

where again $K \equiv e^c > 0$ so $-K < 0$.

Combining these we have for any y never 0,

$$y = e^{at+c} = Ke^{at}$$

This is valid for any $a \in \mathbb{R}$ and $K \neq 0$.

Finally we note that $K = 0$ also is a solution, and we again conclude that

$$y = e^{at+c} = Ke^{at}$$

for any $a \in \mathbb{R}$ and $K \in \mathbb{R}$.

This method is a bit more complicated only because of having to be careful with the absolute values, but it has the advantage of generalizing to other differential equations in an important way seen below when discussing separable equations.

Here is the Sagemath code for the figure. You can try to modify it for other DEs!

```
x=var('x')
y=var('y')
K=var('K')
ysol=(K*e^(x))
show(ysol)
p1=plot_slope_field(y,(x,-3,3.2),(y,-5,5))

for␣i␣in␣range(-16,16):
␣␣␣␣p1=p1+plot(ysol(K=i/8),x,-3,3.2,ymin=-5,ymax=5,axes=True,aspect_ratio=1/2)
p1
```

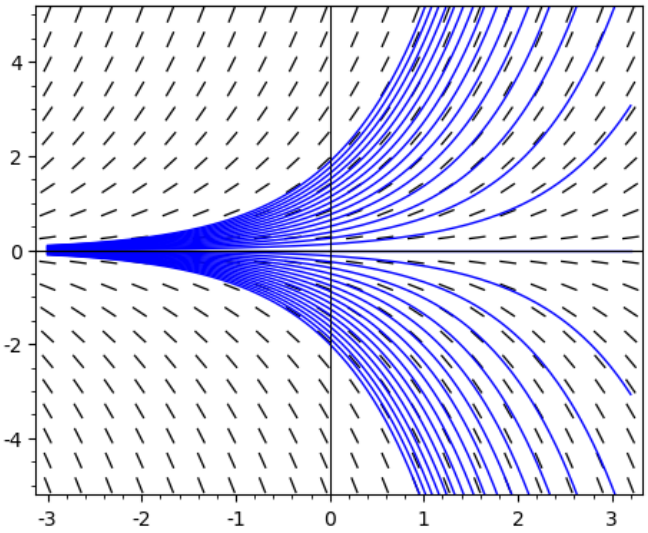


FIGURE 106. Slope field and solution curves for exponential growth $y' = cy$. The equation is in one dimension, and its flow is along the real line; these curves are the *graphs* of those solutions, so Including the time variable. This can also be viewed as solutions to a vector ODE in the plane, where the curves are tangent to the vector field $V(x, y) = (1, y)$. These solution curves are $\gamma(t) = (t, y(t))$ so $\gamma'(t) = (1, y'(t)) = V(\gamma(t)) = (1, y(t))$. The difference between a *slope field* and a vector field is this: segments in the slope field are parallel to the vector field but meet the curves in their midpoint. The picture of the slope field is often easier to understand, as it is much less cluttered since all the segments are all of the same maneagable length.

FIGURE 107. Curves tangent to vector field ($F(x, y) = (1, y)$. These are solution curves to the differential equation in one dimension for exponential growth or decay, $y' = ay$, in this case with $a = 1$. The solutions are $y(t) = Ke^{at}$ for $K \in \mathbb{R}$. If $a < 0$ it is exponential decay, and if $a = 0$ is constant. The graph of a solution $y(t)$ is the *image* of a solution curve for a DE in $\mathbb{R}^2$, for the curves $\gamma(t)$ tangent to the vector field. These solution curves are $\gamma(t) = (t, y(t))$ so $\gamma'(t) = (1, y'(t)) = V(\gamma(t)) = (1, y(t))$.

(Please share any nice figures you produce!)

1.a. Application to biology: more sophisticated models of growth.

Simple exponential growth gives a good model for population growth (of a species, whether plant, animal or bacteria or virus!) only in the beginning stages, before the population starts to run out of space or food. Here are some more realistic models.

Malthus model: the evolution of a yeast population over time.

We are led to solve the DE

$$y'(t) = by(t) - dy(t) = (b-d)y(t),$$

where b and d are real numbers (b is the birth rate and d the death rate) and $y(t)$ denotes the number of cells at time t.

This model is not very realistic (for example if $b > d$ above, there is a "population explosion" (!). To overcome this, one has the

Verhulst model:
Here we are led to solve the (more complicated) DE

$$y'(t) = (b-d)y(t)(M - y(t))$$

where M is the maximum possible number of cells.

We will learn to solve both of these. We shall see that the first equation above is linear homogeneous of first order and the second, the logistic equation, is a *separable* equation, see below.

1.b. Analytic definition of a differential equation:

A differential equation (denoted DE) is a relation between an unknown function y (to be determined) and a certain number of its derivatives. More precisely,

Definition 37.2. A *differential equation in one dimension* is an equation of the form

$$(*) \qquad F(t, y(t), y'(t), \cdots, y^{(n)}(t)) = 0,$$

where F is a given continuous function of $n+2$ variables (so at least one derivative is involved!) and $t \mapsto y(t)$ is a *unknown* function, that we are trying to find. We denote by $y^{(n)}$ its derivative of order n, a strictly positive integer.

The above DE (*) is then said to be **order** n.

The DE is said to be **explicit** of order n, or **in normal form**, if it can be solved for the highest-order derivative, in other words if can be written in the form

$$(*) \qquad y^{(n)}(t) = F(t, y(t), y'(t), \cdots, y^{(n-1)}(t))$$

where now F is a given continuous function of $n+1$ variables. Otherwise it is an **implicit** DE.

To motivate this terminology, recall that in Calculus the equation $x^2 + y^2 = 1$ is equivalent to the four equations $y = \pm\sqrt{1-x^2}$, $x = \pm\sqrt{1-y^2}$ and we can say that the first equation is an *implicit* equation in that it "implies" the other "explicit" equations where we have solved for one variable as a function of the other. For $\mathbb{R}^n$, the Implicit Function Theorem can help us determine when this can be done. In much the same way, we can have an implicit DE, for example $(y')^2 + y^2 = 1$ which implies the explicit equations $y' = \pm(1-y^2)$.

Examples:
$y'(t) + y(t) = 1$: is an DE of first order
$\sin(y'(t)) + y^3(t) = t$: an implicit DE of first order
$y^{(7)}(t) + y^9(t) = t + \sin(5t)$: an DE of order 7

$y(t) + y^2(t) = t$ is not an DE (as there is no derivative involved!).

A **solution** of (*) over an interval I of $\mathbb{R}$ is a function $t \mapsto y(t)$ which is n times derivable on I and which satisfies (*).

The interval I on which we solve a DE is very important, as changing the interval may allow for different solutions.

To **solve** (*) means to find **all** the solutions of (*).
There can be zero, one, several or an infinite number of solutions.

Examples:
- The DE $y'(t) = 0$ admits an infinite number of solutions. Indeed, $y(t) = c$ for any $c \in \mathbb{R}$ is a solution.
- The implicit DE $y'^2(t) + 1 = 0$ does not admit any real solution.

First-order equations

In this chapter, **we shall restrict ourselves to 1^{st} order differential equations**. We shall study, in particular, these two types of differential equation:
- linear DEs;
- separable DEs.

Definition: As above, a DE of first order is *explicit* or *in normal form* if it is of the form

$$y'(t) = f(t, y(t)),$$

where f is a given continuous (two-variable) function.

Note: it is, in general, easier to solve the DE in normal form.

Examples: The following are all first order DEs.
$y'(t) = \cos(y(t))$ is a DE in normal form.
$\sin(y'(t)) = y(t)$ is not in normal form.
$y'(t) = y(t) + 2te^t$ is in normal form.

2. First- order linear DE:

Definition: An DE is said to be 1^{st} order linear if it is of the form

$$y'(t) = a(t)y(t) + b(t) \qquad (E)$$

where $t \mapsto a(t)$ and $t \mapsto b(t)$ are two given continuous functions on I, and $t \mapsto y(t)$ is the unknown function to be determined.

- $a(t)$ is called the *coefficient* of (E) and $b(t)$ the *second coefficient* or *second term*.
- If $b \equiv 0$ then we say that the DE is **homogeneous** or *without second term*.

If $a(t) = a$ and $b(t) = b$ we say the equation is **autonomous** or **stationary**. Otherwise it is **non-autonomous**, **nonstationary**, **time-dependent** or **constant coefficient**.

Examples:
• $y'(t) + y(t) = 2$: linear constant coefficient DE of first order.
Indeed, we write $y'(t) = -y(t) + 2$. Here $a(t) = -1$ and $b(t) = 2$.

• $(1 + t^2)y'(t) = ty(t)$: linear non-autonomous DE of first order.
Indeed, as $1 + t^2$ never vanishes on $\mathbb{R}$ then we can rewrite this DE as
$y'(t) = \frac{t}{1+t^2}y(t)$.

Here $a(t) = t/(1 + t^2)$ and $b(t) = 0$. Thus, it is a homogeneous linear 1^{st} order non-autonomous DE.

• $y'(t) = y^2(t) + 1$: non-linear DE

Remark 37.1. The explanation for the term "linear equation" is as follows. Given two vector spaces V, W, recall that a transformation $T : V \to W$ is *linear* iff for any $\mathbf{u}, \mathbf{v} \in V$ we have $T(a\mathbf{u} + b\mathbf{v}) = aT(\mathbf{u}) + bT(\mathbf{v})$. A transformation A is *affine* iff there is a linear transformation T and $\mathbf{w} \in W$ such that $A(\mathbf{v}) = T(\mathbf{v}) + \mathbf{w}$. Note that in this case, $\mathbf{w} = A(\mathbf{0})$.

Exercise: We denote by $\mathcal{C} = \mathcal{C}^0$ the vector space of continuous functions and by $\mathcal{C}^k$ the $k-$times differentiable functions with $f^{(k)} \in \mathcal{C}^0$. Define the map $D : \mathbb{C}^1 \to \mathbb{C}$ by $D(f) = f'$. Show that this is a linear transformation.

Now given continuous functions $a(t), b(t)$ define $T : \mathbb{C}^1 \to \mathbb{C}$ by $T(f) = a(t)D(f) + b(t)$. Show that T is affine, and is linear iff $b \equiv 0$. (Often in Calculus or applied math an affine function is (incorrectly) called linear.)

2.a. Solution of a homogeneous linear DE of first order:

We are trying to solve

$$(H) \qquad y'(t) = a(t)y(t) \tag{138}$$

where a is a continuous function.

Remark: we note that $y(t) = 0$ (for all real t) is solution of (H). This is the *null solution*. But what are the other solutions?

Theorem: The general solution of (H) is

$$y_H(t) = Ce^{A(t)},$$

where C is a real constant and $A(t) = \int a(t)dt$ a primitive of a.

Notes:
- There are an infinite number of solutions of (H) (C is arbitrary in $\mathbb{R}$).
- For $C = 0$, we have the null solution.

Proof: We verify that y_H does verify (H). Indeed, by the Chain Rule,

$$y'_H(t) = CA'(t)e^{A(t)} = Ca(t)e^{A(t)} = a(t)y_H(t)$$

as $A' = a$.

How can we show that we have *all* the solutions? Here is the idea: we write the DE as

$$y'(t) - a(t)y(t) = 0.$$

Then we multiply the two members of the DE by $e^{-A(t)}$

$$e^{-A(t)}(y'(t) - a(t)y(t)) = 0 \quad \iff y'(t)e^{-A(t)} - a(t)e^{-A(t)}y(t) = 0.$$

We recognize the right-hand side as the derivative of $y(t)e^{-A(t)}$. So $(y(t)e^{-A(t)})' = 0$ hence $y(t)e^{-A(t)} = C$ for some constant C, as claimed.

Alternatively, we proceed exactly as for the constant coefficient case: for $v = e^{A(t)}$ note that $v^{-1} = e^{-A(t)}$ so $(v^{-1})' = -a(t)(v^{-1})$. (This is defined since $e^{A(t)} > 0$.) We assume u ia also a solution so $u' = a(t)u$. We then show that $(u/v)' = 0$ by the same calculation as before.

Examples:

• Solve, on $\mathbb{R}$, the homogeneous DE

$$y'(t) = \frac{t}{1+t^2}y(t) \quad (H)$$

Here $a(t) = t/(1+t^2)$, so

$$A(t) = \int \frac{t}{1+t^2}\,dt = \frac{1}{2}\int \frac{2t}{1+t^2}\,dt = \frac{1}{2}\int \frac{u'(t)}{u(t)}\,dt$$

where $u(t) = 1 + t^2$. Therefore

$$A(t) = \frac{1}{2}\ln|u(t)| = \frac{1}{2}\ln(u(t)) = \frac{1}{2}\ln(1+t^2) = \ln(\sqrt{1+t^2}).$$

The solutions of (H) are written as

$$y(t) = C\exp(\ln(\sqrt{1+t^2})) = C\sqrt{1+t^2}, \;\; t \in \mathbb{R}.$$

• Solve, on $\mathbb{R}$, the homogeneous DE

$$y'(t) = 2y(t) \quad (H)$$

Here $a(t) = 2$ so $A(t) = 2t$. The solutions of (H) are written as

$$y_H(t) = Ce^{2t}, t \in \mathbb{R}.$$

We have an infinite number of solutions, C ́being arbitrary. See Fig. 107, and note that we have only one solution passing through the point with coordinates $(0, 1)$.

Indeed, taking $t = 0$, $y_H(0) = Ce^0 = C$ which must be equal to 1. So $C = 1$ and the solution we seek is

$$\tilde{y}_H(t) = e^{2t}, t \in \mathbb{R}.$$

2.b. Solution of a nonhomogeneous linear DE of first order:

We recall the following from Linear Algebra:

Proposition 37.2. *Let V, W be vector spaces and $T : V \to W$ a linear transformation. Given some $\mathbf{w} \in W$, we wish to solve the* nonhomogeneous equation

$$T(\mathbf{v}) = \mathbf{w}.$$

We claim that all such solutions are a sum of one particular solution $\mathbf{v}_p$ such that $T(\mathbf{v}_p) = \mathbf{w}$ and any solution $\mathbf{v}_0$ of the corresponding homogeneous equation $T(\mathbf{v}_0) = \mathbf{0}$.

Proof. We have $T(\mathbf{v}_0 + \mathbf{v}_p) = T(\mathbf{v}_0) + T(\mathbf{v}_p) = \mathbf{0} + \mathbf{w} = \mathbf{w}$. Conversely, if $T(\mathbf{v}) = \mathbf{w}$ then $T(\mathbf{v} - \mathbf{v}_p) = \mathbf{0}$ so $\mathbf{v} - \mathbf{v}_p$ is a solution of the homogeneous equation. □

Remark 37.2. Geometrically, note that the solutions of the homogeneous equation are the kernel (nullity) of T, which is a subspace of V; adding on $\mathbf{v}_p$ translates this to an affine subspace which is the space of solutions of the nonhomogeneous equation.

We are trying to solve

$$(E) \qquad y'(t) = a(t)y(t) + b(t)$$

where a and b are two continuous functions (over an interval I).

We have seen above that $T(y) = y'(t) - a(t)y(t)$ is a linear transformation from $\mathcal{C}^1$ to $\mathcal{C}$, so we can apply the Proposition: we wish to solve the nonhomogeneous equation $T(y(t)) = b(t)$.

(I) Thus, we begin by solving the associated homogeneous DE:

$$(H) \qquad y'(t) = a(t)y(t).$$

The solutions of (H) have been given in 2.a. They are of the form

$$y_H(t) = Ce^{A(t)},$$

where C is a real constant and A a primitive of the function a.

(II) We are looking for one particular solution y_P of (E), i.e. a function y_P which satisfies:

$$y'_P(t) = a(t)y_P(t) + b(t).$$

From the Proposition we then have:

Theorem: The solutions of (E) are of the form

$$y(t) = y_H(t) + y_P(t).$$

Example: Solve on $\mathbb{R}$ the linear DE

$$(E) \quad y'(t) = -2y(t) + 1.$$

- We begin by solving the associated homogeneous DE

$$(H) \quad y'(t) = -2y(t).$$

Here $a(t) = -2$, so $A(t) = -2t$.

The solutions of (H) are of the form

$$y_H(t) = Ce^{-2t}, t \in \mathbb{R}.$$

- We try to guess a particular solution y_P of (E). We notice that $y_P(t) = 1/2$ is one because in this case $y_P'(t) = 0$ and then $0 = -2 \times \frac{1}{2} + 1$.

2.c. Particular solution

How to determine a particular solution y_P of (E)?

It is sometimes easy to "guess" y_P. For example, if the DE is linear with coefficients a and b constants then we look for $y_P(t) =$ constant to be determined.

Examples:

• $y'(t) + 2y(t) = 5$

As this is a DE with constant coefficients, and taking into account the above remark, we find $y_P(t) = 5/2$.

• $y'(t) + 2y(t) = e^t$

Since a is a constant and the second term b is te^t, we guess $y_P(t) = ce^t$ and then try to determine c.

If $y_P(t) = ce^t$ then $y_P'(t) = ce^t$. We write that $y_P'(t) + 2y_P(t) = e^t$ and then find that $c = 1/3$.

Therefore $y_P(t) = e^t/3$.

• $y'(t) + ty(t) = t$

Note that $y_P(t) = 1$ is a special solution, because $y_P'(t) = 0$ and $0 + t \cdot 1 = t$.

If we did not succeed to find a particular solution, we can then resort to the method of the variation of the constant (also known as Lagrange's method).

Method of varying the constant:

To introduce this idea, which we shall use to find a particular solution y_P, consider the following. We have seen that the solutions of the homogeneous equation (H)

$$y'(t) = a(t)y(t)$$

are

$$y(t) = Ce^{A(t)}$$

where

$$A' = a$$

.

But suppose we allow the constant factor C to be replaced by a differentiable function $C(t)$? What wider class of equations can we now solve?

We define

$$y(t) = C(t)e^{A(t)},$$

and calculate the derivative. As $A' = a$ we have

$$\begin{aligned} y'(t) &= C'(t)e^{A(t)} + C(t)e^{A(t)}A'(t) \\ &= C'(t)e^{A(t)} + a(t)y(t) \\ &= a(t)y(t) + b(t) \end{aligned}$$

where we are defining

$$b(t) = C'(t)e^{A(t)}.$$

Equivalently,

$$C'(t) = b(t)e^{-A(t)}.$$

or

$$C(t) = \int b(t)e^{-A(t)}dt.$$

Working backwards, given continuous function $b(t)$, we can define $C(t)$ as above, and conclude that we now have found a particular solution y_P for the equation we started with,

$$y'(t) = a(t)y(t) + b(t).$$

We denote this particular solution by $y_P(t) = C(t)e^{A(t)}$ where $C(t) = \int b(t)e^{-A(t)}dt$.

The general solution:

We wish now to find all solutions of the nonhomogeneous equation

$$y'(t) = a(t)y(t) + b(t)$$

We have already found all solutions y_H of the homogeneous equation, and one particular solution y_P of the nonhomogeneous equation. From Proposition 37.2, we know that adding these, $y = y_H + y_P$ gives the general solution.

Here we give a direct proof for this case.

We know that

$$y'_H(t) - a(t)y_H(t) = 0$$

$$y'_P(t) - a(t)y_P(t) = b(t).$$

Adding, $y = y_H + y_P$ solves $y'(t) - a(t)y(t) = b(t)$ i.e. $y'(t) = a(t)y(t) + b(t)$. We claim that we now have all solutions.

Thus, as $A' = a$ we have

$$\begin{aligned} y'_P(t) &= C'(t)e^{A(t)} + C(t)e^{A(t)}A'(t) \\ &= C'(t)e^{A(t)} + a(t)y_P(t) \\ &= a(t)y_P(t) + b(t). \end{aligned}$$

So

$$C'(t)e^{A(t)} = b(t) \Longleftrightarrow C'(t) = b(t)e^{-A(t)}.$$

from which

$$C(t) = \int b(t)e^{-A(t)}dt.$$

The particular solution given by the variation method of the constant is then

$$y_P(t) = \left(\int b(t)e^{-A(t)}dt\right)e^{A(t)}.$$

Therefore, the general solution of (E) is written as

$$y(t) = y_H(t) + y_P(t) = Ce^{A(t)} + \left(\int b(t)e^{-A(t)}dt \right) e^{A(t)}.$$

Example:
Solve for $t > -1$, and using the variation method of the constant, the DE

$$(E) \qquad y'(t) = \frac{1}{t+1}y(t) + \frac{1}{2}.$$

It is a 1^{st} order linear DE where u $a(t) = \frac{1}{t+1}$ and $b(t) = \frac{1}{2}$.

We have $A(t) = \int a(t)dt = \int \frac{1}{t+1} = \ln|t+1| = \ln(t+1)$ because $t > -1$ and therefore $t + 1 > 0$.

- The solutions of the associated homogeneous DE

$$(H) \qquad y'(t) = \frac{1}{t+1}y(t)$$

are

$$y_H(t) = Ce^{A(t)} = Ce^{\ln(t+1)} = C(t+1),$$

where C is an arbitrary real constant.

- We are looking for a particular solution (via the method of variation of the constant) in the form

$$y_P(t) = C(t)e^{A(t)} = C(t)(t+1),$$

with

$$C(t) = \int b(t)e^{-A(t)}\, dt.$$

We calculate

$$C(t) = \int \frac{1}{2}\, e^{-\ln(t+1)}\, dt = \frac{1}{2}\int \frac{1}{t+1}\, dt = \frac{1}{2}\ln(t+1).$$

So

$$y_P(t) = \frac{1}{2}(t+1)\ln(t+1).$$

Thus, the solutions of (E) for $t > -1$ are

$$y(t) = y_H(t) + y_P(t) = C(1+t) + \frac{1}{2}\,(t+1)\ln(t+1).$$

There are therefore an infinite number of solutions of (E), but we will see that there is only one only solution with the initial condition $y(0) = 1$.

Indeed, we set $t = 0$ in the formula for $y(t)$ above and then we write that $y(0) = 1$. This gives

$$y(0) = C(1+0) + \frac{1}{2}\,(0+1)\ln(0+1) = C = 1.$$

The solution is then

$$y(t) = (1+t) + \frac{1}{2}\,(t+1)\ln(t+1).$$

3. Separable Differential equations:

3.a. Definition: An DE is said to be *separable* if it is of the 1st order and of the form

$$(E) \qquad g(y(t))y'(t) = f(t)$$

where $f : I \to \mathbb{R}$ and $g : J \to \mathbb{R}$ are two given continuous functions.

The left-hand side of (E) depends only on the function $y(t)$ and the right-hand side only on t.

Examples:

- A linear and homogeneous DE (1st order) is separable. Indeed, we already know that the null function is a solution. If we are looking for the non-zero solutions of

$$(H) \qquad y'(t) = a(t)y(t),$$

(modulo a theoretical justification that we will not detail here) we are reduced to solving

$$\frac{y'(t)}{y(t)} = a(t).$$

This last DE is separable, with $g(x) = 1/x$ and $f(t) = a(t)$.

- The 1st order linear DE

$$y'(t) = y(t) + t$$

is NOT a separable DE. We cannot write it in the form of (E) above.

3.b. Method of solving (E):

Since f and g are two continuous functions over two intervals then they admit primitives F and G, respectively.

Note that the derivative of the composite function $G \circ y$ is

$$(G \circ y)'(t) = (G(y(t)))' = G'(y(t))y'(t) = g(y(t))y'(t).$$

We can therefore write

$$g(y(t))y'(t) = f(t) \iff \int g(y(t))y'(t)dt = \int f(t)dt + c$$

where c is an arbitrary constant.

So

$$G(y(t)) = F(t) + c,$$

where $y(t)$ is given implicitly. It remains to express $y(t)$ as a function of t.

Let's look next at a concrete example how solve a separable DE.

Examples:

• (E) $(1 + t^2)y'(t) = 2ty(t)$.

This is a linear DE and we already have one method of resolution.

It is also a separable DE. Let's try to solve this DE using the strategy described above.

We know that $y(t) = 0$ is a particular solution of (E). If we look for non-zero solutions ($y(t) \neq 0$), we rewrite the DE as:

$$\frac{y'(t)}{y(t)} = \frac{2t}{t^2+1} \Longleftrightarrow \int \frac{y'(t)}{y(t)} dt = \int \frac{2t}{t^2+1} dt + c$$

So

$$\ln|y(t)| = \ln(t^2+1) + c$$

because $1 + t^2 > 0$, and therefore

$$|y(t)| = e^{\ln(t^2+1)+c} = e^c(t^2+1) \Longleftrightarrow y(t) = \pm e^c(t^2+1).$$

So (counting the zero solution) the general solution of the DE is

$$y(t) = d(t^2+1), d \in \mathbb{R}.$$

• $y'(t) = y(t) - 1$.

Note that the constant function $y(t) = 1$ is a solution. Let's find the other solutions. Since $y(t) \neq 1$, then

$$\frac{y'(t)}{y(t)-1} = 1 \Longleftrightarrow \int \frac{y'(t)}{y(t)-1} dt = \int 1\, dt + c.$$

This gives

$$\ln|y(t)-1| = t + c \Longleftrightarrow |y(t)-1| = e^{t+c} = e^c\, e^t$$

or

$$y(t) - 1 = \pm e^c\, e^t \Longleftrightarrow y(t) = 1 + de^t$$

where $d = \pm e^c$ which is a non-zero constant.

The solutions of the DE (counting the solution $y(t) = 1$) are therefore of the form

$$y(t) = 1 + \lambda e^t, \lambda \in \mathbb{R}.$$

Exercise: *Look at the solution graphs* (for $\lambda < 0$, $\lambda = 0$) and $\lambda > 0$

Exercise: Solve on $\mathbb{R}$ the linear DE

$$(1+t^2)^2 y'(t) + 2ty(t) = 2t.$$

Hints:

$y_H(t) = Ce^{\frac{1}{1+t^2}}$

$y_P(t) = 1$.

Some solutions:

Here is how we can calculate that for any exponential function e^{at} for $a > 0$ then $1 = e^{a\cdot 0}$ and for $2 = e^{a\cdot t_d}$ we have $a \cdot t = \ln 2$, $t_d = \ln 2/a > 0$.

Calculating the half-life for exponential decrease, e^{at} for $a < 0$ we have $t_h = \ln \frac{1}{2}/a = -\ln 2/a > 0$.

37.3. **Exact differential equations.**
Implicit equations.
Exact differential equations.
Integrating factors.
Some solutions.

For example, the exterior derivative d operator sends $k+1$-forms on a manifold M to k-forms, and an equation $d\eta = \rho$ is a differential equation in this general sense. Again, a solution η can be termed a *primitive* and the procedure of solving this can be called *integration*. And again, the question of what is the kernel of the operator is important, though here the answer may be much more interesting, due to homology.

The simplest example here is where φ is a function on $\mathbb{R}^2$ and $d\varphi$ is (dual to) its gradient vector field.

(TO DO...)

37.4. **Integrating factors.**

(TO DO...)

(TO DO...)

37.5. **Flows, vector fields and systems of equations.** Consider a differentible manifold M with a $\mathcal{C}^1$ flow, $\tau_t : M \to M$. We let $\mathcal{F}(M)$ denote the collection of all such flows. $T(M)$ denotes the tangent bundle of M. A *(continuous) vector field* on M is a continuous map $V : M \to T(M)$. We write $\mathcal{V}(M)$ for the collection of all vector fields on M. We describe a map from flows to vector fields, that is, from $\mathcal{F}(M)$ to $\mathcal{V}(M)$. Given a flow τ on M and a point $p \in M$, we consider the curve $\gamma(t) = \tau_t(p)$ and define $d/dt(\tau_t(p))$ to be the tangent vector at time t, so at location $\mathbf{x} = \gamma(t)$, $\gamma'(t) \in T(M)_{\mathbf{x}}$.

We define a vector field V by

$$V(p) = \mathrm{d}/\mathrm{d}t(\tau_t|_{t=0}(p)).$$

Lemma 37.3. *This satisfies:*

$$V(\tau_t(p)) = d/dt(\tau_t(p)) \tag{139}$$

for all p, t.

Proof. From the flow property $\tau_{t+s} = \tau_s \circ \tau_t$, defining $q = \tau_t(p)$, then $\mathrm{d}/\mathrm{d}t(\tau_t(p)) = \lim_{h\to 0}(\tau_{h+t}(p) - \tau_t(p))/h = \lim_{h\to 0}(\tau_h(q) - \tau_0(q))/h = \mathrm{d}/\mathrm{d}t_t|_{t=0}(\tau_t(q)) = V(q) = V(\tau_t(p))$. Here we have done the calculation for a flow in $\mathbb{R}^n$; this is pushed to manifolds via charts, in the usual way. □

Equation (139) can be written as:

$$d/dt(\tau_t) = V(\tau_t)$$

or

$$d/dt(\tau) = V.$$

Thus *the time derivative of a flow is a vector field.*

The most general notion of differential equation is this: given some notion of derivative (usually but not always a time derivative), can we find an inverse operation, that

is, "integrate" it to find the antiderivative or primitive? This is called the *solution* of the differential equation.

In the present example, the question becomes: given a vector field, can we find a flow with that as its derivative? In other words, can we "integrate" the vector field to get a flow?

Remark 37.3. We have earlier encountered an ergodic theory example of a linear flow and its eigenvectors in Definition 14.5.

37.6. **Vector fields and flows in $\mathbb{R}^n$.** We consider the special case of $M = \mathbb{R}^n$, where the tangent bundle is the product $T(\mathbb{R}^n) = \mathbb{R}^n \times \mathbb{R}^n$, and identifying the base space, the first $\mathbb{R}^n$, with the fiber, the second, then a vector field is just a continuous map $V : \mathbb{R}^n \to \mathbb{R}^n$. This is visualized by drawing the vector $\mathbf{v}_p = V(p)$ at the location p.

Some good references are: [HS74] (we mostly prefer this original edition), [Arn12], [HK03], [Lan02], [Lan01]. When completing these notes we came across the new text of Viana and Espinar [VE21] developed in a masters level course at IMPA (publication date Feb 28, 2022). This contains a wealth of material for further learning, including interestig historical notes.

Remark 37.4. An interesting consequence of Theorem 35.35 is the following:

Corollary 37.4. *Given a linear flow $\tau_t = e^{tA}$ on $\mathbb{R}^n$, this flow is volume-preserving (i.e. the determinant of each map τ_t is one) iff the vector field $V(\mathbf{x}) = A\mathbf{x}$ has divergence* 0.

Proof. A calculation immediately shows that the divergence of V is the trace of A. □

This makes rigorous, for the linear case, the often-heard intuition of divergence 0: that material is not created, i.e. the flow of the vector field preserves volume.

This proof extends immediately to the nonlinear case once we have understood that the derivative of the flow is a vector field.

See [KH95], Proposition 5.1.9 regarding divergence and and volume preservation. and [HK03]. 6.1.5. The connection between trace, determinant and divergence is not however made there.

See Arnold p. 251 [Arn12] where this connection is made, Remark and Problem 3.

One-parameter subgroups and vector fields: examples.

We describe further (2×2) examples, via exploring the connection with ODEs: systems of DEs and vector-valued DEs.

Example 52. **(Hyperbolic Rotation flow 2)** We continue our study of Example 43; see also Example **??**. We have $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, and are studying the quadratic form $Q(x, y) = \mathbf{v}^t A\mathbf{v}$ for $\mathbf{v} = (x, y)$. This gives $Q(x, y) = 2xy$. We diagonalized the symmetric matrix A to $D = U^{-1}AU$ where $U = R_{\pi/4}$ is orthogonal. The corresponding quadratic form for D, which has diagonal elements $1, -1$ is $\widehat{Q}(x, y) = x^2 - y^2$.

Considering first the linear vector field given by D, the solutions of the vector DE $\mathbf{x}' = D\mathbf{x}$ are the right hyperbolas, and are invariant for the hyperbolic linear flow with

fixed point $\mathbf{0}$, $T_t = e^{tD} = \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}$. The $\widehat{Q}$-hyperbolas are invariant for a different hyperbolic flow. Recalling that

$$\cosh(t) = \frac{e^t + e^{-t}}{2},\ \sinh(t) = \frac{e^t - e^{-t}}{2}$$

and that, from part (a) of Theorem 35.19, $Ue^{tD}U^{-1} = e^{tUDU^{-1}} = e^{tA}$, this is

$$\begin{aligned} e^{tA} = UT_tU^{-1} =& \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} e^t + e^{-t} & e^t - e^{-t} \\ e^t - e^{-t} & e^t + e^{-t} \end{bmatrix} = \begin{bmatrix} \cosh(t) & \sinh(t) \\ \sinh(t) & \cosh(t) \end{bmatrix} \end{aligned} \tag{140}$$

We note that since

$$\cos(t) = \frac{e^{it} + e^{-it}}{2},\ \sin(t) = \frac{e^{it} - e^{-it}}{2i}$$

one has from the definitions, $\cosh, \sinh$ are:

$$\cosh(t) = \cos(it),\ \sinh(t) = -i\sin(it),$$

and we have the interesting Taylor series:

$$\cosh(t) = 1 + t^2/2 + t^4/4! + \dots \text{ and } \sinh(t) = t + t^3/3! + t^5/5! + \dots$$

We give a different, matrix proof of this below.

Each map e^{tA} is known as a *hyperbolic rotation*, and not just by analogy from the formulas!

Indeed, the usual rotations $\{R_t\}_{t\in\mathbb{R}} = SO(2)$, the special orthogonal group, preserves the Euclidean metric, the standard inner product, and the quadratic form $x^2 + y^2$, and hence it preserves the concentric circles. Each quadratic form is preserved by a Lie group. In the case of our form, the tilted hyperbolas are distance one from the origin with respect to the indefinite bilinear form, and are preserved by the group of orientation-preserving isometries which are the hyperbolic rotations.

The upper quadrant of the hyperbola provides a model for the Lorenz metric of Special Relativity, in one spatial and one time dimension. This metric is preserved by the special orthogonal group of signature $(+1, -1)$, in this case written $SO^+(+1, -1)$. (The physics convention is to switch x and y, giving signature $(-1, +1)$; the reason is because in the higher-dimensional case we have n spatial and one time dimension and write time first, giving signature $(-1, +1, \dots, +1)$).

Each branch of the hyperbola is a model for one-dimensional hyperbolic space. The hyperbola is preserved by the flow; we shall check that hyperbolic distance is indeed equal to time t. We know the hyperbolic metric in the upper cone for the lines $y = \pm x$ is the projective metric there, with $(d(a, b), (c, d)) = |\log(a/b) - \log(c/d)|$. See §23.5. We calculate this for the isometric flow T_t. Then $T_t(a, b) = (e^t a, e^{-t}b)$ so $e^t a/e^{-t}b = e^{2t}a/b$ and so for $\alpha = e^{2t}$, $\mathbf{v} = (a, b), \mathbf{w} = (c, d)$ then $(d(T_t(\mathbf{v})), T_t((\mathbf{w})) = |\log(\alpha(a/b)) - \log(\alpha(c/d))| = |\log(a/b) - log(c/d)| = d(\mathbf{v}, \mathbf{w})$. This can be seen geometrically using where straight lines from the origin (points in projective space) meet the hyperbola.

Exercises. Rewrite the following systems of equations in vector form $\mathbf{x}' = A\mathbf{x}$, for $\mathbf{x} = (x, y)$, $\mathbf{x} : \mathbb{R} \to \mathbb{R}^2$. Sketch the vector field and slope field by hand. Solve explicitly for initial condition $\mathbf{x}_0 = (a, b)$ and sketch the solutions.

(a)

$$\begin{cases} x' = -y \\ y' = x \end{cases}$$

(b)

$$\begin{cases} x' = x \\ y' = y \end{cases}$$

(c)

$$\begin{cases} x' = x \\ y' = -y \end{cases}$$

(d)

$$\begin{cases} x' = y \\ y' = x \end{cases}$$

(e)

$$\begin{cases} x' = x - y \\ y' = x + y \end{cases}$$

(f)

$$\begin{cases} x' = x \\ y' = 2y \end{cases}$$

(g)

$$\begin{cases} x' = x \\ y' = x + y \end{cases}$$

Solutions: In each case the solutions are of the form

$$\mathbf{x}(t) = e^{tA}\mathbf{x}_0$$

for the following matrices A. These vector fields and flows are already studied in Examples $(a) - (g)$ where the matrix A is: (a): Ex. 34: $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, (b): Ex. 35: $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (c): Ex. 36: $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, (d): Ex. **??**: $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, (e): Ex. 38: $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, (f): Ex. 39: $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$, (g): Ex. 41: $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.

Solving a second-order DE via a system of first-order DEs. We return to Exercise **??**:

Exercise 37.4. (Harmonic oscillator; hyperbolic rotation)
(i) Solve the second order linear equation $y'' = -y$ by the following strategy: we define $y' = x$ and $x' = -y$, giving a system of two equations of first order. Then we rewrite the system in vector form $\mathbf{x}' = A\mathbf{x}$, for $\mathbf{x} = (x, y)$ as above.

Now solve this vector DE explicitly for initial condition $\mathbf{x}_0 = (a, b)$ and sketch the solutions. Lastly, returning to the original equation $y'' = -y$, what are the solutions $y(t)$?
(ii) Do the same for the second order equation $y'' = y$.

Solution.

The matrices for (i), (ii) are $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, so we fall into Examples (a) and (d) above. For the first we have the solution
(i)

$$e^{tA}\mathbf{x}_0 = R_t\mathbf{x}_0 = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a\cos t - b\sin t \\ a\sin t + b\cos t \end{bmatrix} = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}$$

for the vector DE with initial condition $\mathbf{x}_0 = (a, b)$. The general solution for the one-dimensional second order equation $y'' = -y$ is therefore

$$y(t) = a\sin t + b\cos t. \tag{141}$$

Note that $x(0) = a = y'(0)$, so the initial condition is $y(0) = b, y'(0) = a$. Physically, this corresponds to a harmonic oscillator with mass and spring constant 1, and with initial position $y(0) = b$, initial velocity $y'(0) = a$.

Fixing $y(0) = b$, we see all the circles which meet the line $y = b$ in the plane, each corresponding to a different initial velocity.
(ii) For $y'' = y$ we have

$$e^{tA}\mathbf{x}_0 = R_t\mathbf{x}_0 = \begin{bmatrix} \cosh(t) & \sinh(t) \\ \sinh(t) & \cosh(t) \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a\cosh t + b\sinh t \\ a\sinh t + b\cosh t \end{bmatrix} = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}$$

whence the solution for the one-dimensional second order equation $y'' = y$ is:

$$y(t) = a\sinh t + b\cosh t \tag{142}$$

Here the (position, velocity) initial condition is $y(0) = b, y'(0) = x(0) = a$.

See Fig. 108 for the graphs of $\sinh t = (e^t - e^{-t})/2, \cosh t = (e^t + e^{-t})/2$. (Because of cancellation, the graphs of $a\sinh t + b\cosh t$ are not particularly interesting, but that also holds for $a\sin t + b\cos t$, which is just a translated sin function: note that this is physically clear as it is harmonic motion of the same period and amplitude but a different initial value).

37.7. **Systems of equations and vector fields.** We have so far seen some examples of a system of equations defining a vector field, but have not yet given the proper definitions nor considered wider context.

We saw above for example the "system"

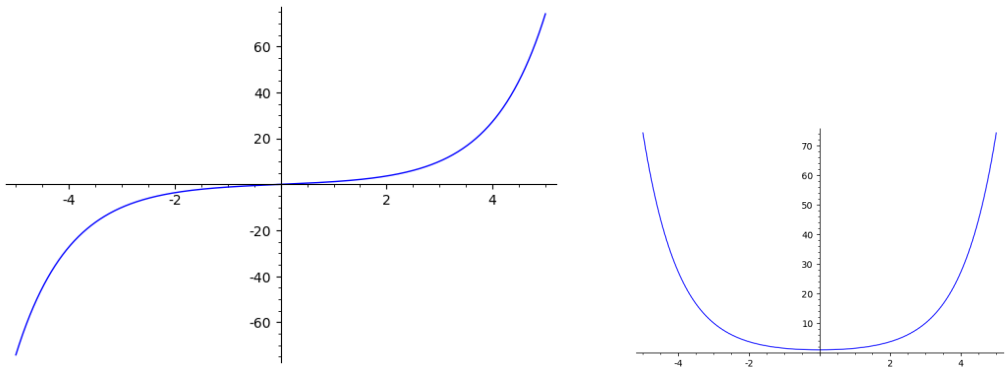


FIGURE 108. Graphs of sinh, cosh.

$$\begin{cases} x' = -y \\ y' = x \end{cases}$$

37.8. **Existence and uniqueness theorems.** Here we show that given a Lipschitz condition on the vector field, solutions for our differential equation always exist locally, and moreover exist globally and are unique. This holds both in the stationary and nonstationary cases. Moreover, one has information on the smoothness of the solutions: a smooth vector field guarantees smooth solutions. All of this can be phrased in terms of flows. First we prove the standard fixed point theorem for contraction mappings of complete metric spaces, with an extension showing continuous dependence of the fixed point for a parametrized family of maps.

Existence and continuity of fixed points for parametrized contractions.

Definition 37.3. Given two metric spaces (X, d) and $(\widehat{X}, \widehat{d})$, a map $f : X \to \widehat{X}$ is K-Lipschitz for some $K > 0$ iff for each pair of points $x, y \in X$ we have

$$d(f(x), f(y)) \leq Kd(x, y).$$

This is called a *contraction mapping* iff $0 \leq K < 1$.

Lemma 37.5. *A K-Lipschitz function is continuous.*

Proof. If $x_n \to x$ then $d(f(x_n, f(x)) < Kd(x_n, x) \to 0$. □

We write $\mathcal{C}^k = \mathcal{C}^k(\mathbb{R}, \mathbb{R})$ for the functions whose k^{th} derivative exists and is continuous. Thus $\mathcal{C}^0$ denotes the contiuous functions, $\mathcal{C}^1$ the functions f whose derivative $f' \in \mathcal{C}^0$ and inductively $\mathcal{C}^{k+1}$ those functions whose derivative is in $\mathcal{C}^k$.

A function is called $\mathcal{C}^{k+1}$ if its k^{th} derivative is not only continuous but Lipschitz for some $K > 0$. Thus

$$\mathcal{C}^0 \subseteq \mathcal{C}^{0+1} \subseteq \mathcal{C}^1 \subseteq C^{1+1} \subseteq \ldots \mathcal{C}^k \subseteq C^{k+1} \subseteq \ldots$$

Note that $\mathcal{C}^{0+1}$ denotes the space of functions which are K-Lipschitz for some $K \geq 0$.

The basic tool is:

Lemma 37.6. *Let (X, d) be a complete metric space and $f : X \to X$ a strict contraction, i.e. there exists $c \in [0, 1)$ such that $d(f(x), f(y)) \leq cd(x, y)$. Then f has a unique fixed point.*

Proof. Note that since f is a contraction it is continuous: if $x_n \to x$ then $d(f(x_n), f(x)) \leq cd(x_n, x) \to 0$.

We show that if there is a fixed point x, it must be unique. Suppose there is another fixed point $y \neq x$. But then $d(x, y) = d(f(x), f(y)) \leq cd(x, y) < d(x, y)$, a contradiction.

Now let $x \in X$. If $f(x) = x$ we are done. So suppose $x \neq f(x)$. Define $x_n = f^n(x)$; we shall prove this is a Cauchy sequence. We know $x_0 \neq x_1$. Applying the map f, $d(x_{n+1}, x_n) \leq c^n d(x_1, x_0)$. Let n so large that $d(x_{n+1}, x_n) < \delta$. Then for any $k \geq 1$, $d(x_{n+k+1}, x_{n+k}) < \delta c^k$. Thus by the triangle inequality, $d(x_n, x_{n+k}) < \delta \sum_1^k c^j < \delta \sum_1^\infty c^j = \delta M$. Thus for δ small this is $\delta M < \varepsilon$.

This proves $(x_n)_{n\geq 0}$ is Cauchy. Since X is a complete metric space, there exists a limit point x_∞, such that $x_n \to x_\infty$.

We claim this is a fixed point. But $x_n \to x_\infty$ and also $f(x_n) = x_{n+1} \to x_\infty$ yet since f is continuous, $x_\infty = \lim x_{n+1} = \lim f(x_n) = f(\lim x_n) = f(x_\infty)$.

□

Lemma 37.7. *Let (X, d) be a complete metric space and let (W, d) be a metric space. Let $f(x, w)$ be a continuous map from $X \times W$ to X. Write $f_w(x) = f(x, w)$. Let $f_w : X \to X$ be a contraction for each $w \in W$ with the constant, $c \in [0, 1)$; that is, $d(f_w(x), f_w(y)) \leq cd(x, y)$. Then the unique fixed point x_w of f_w varies continuously in w.*

Proof. We follow Proposition 2.6.14 on p. 68 of [HK03].

Applying f_w to x we set $x_n^w = f_w^n(x) \to x_w$; let N be so large that for $n > N$, $d(x_n^w, x_w) \leq \varepsilon$.

Note that: $d(x, x_w) \leq d(x, x_n^w) + d(x_n^w, x_w) \leq d(x, x_n^w) + \varepsilon \leq \sum_{k=0}^\infty d(x_k^w, x_{k+1}^w) + \varepsilon$ for all $\varepsilon > 0$, whence:

$$d(x, x_w) \leq \sum_{k=0}^\infty d(x_k^w, x_{k+1}^w) = \sum_{k=0}^\infty d(f_w^k x, f_w^{k+1} x) \leq d(x, f_w(x)) \sum_{k=0}^\infty c^k = \frac{1}{1-c} d(x, f_w(x)).$$

Applying this to $x = x_y = f_y(x_y)$ we have

$$d(x_y, x_w) \leq \frac{1}{1-c} d(x_y, f_w(x_y)) = \frac{1}{1-c} d(f_y(x_y), f_w(x_y)) \to 0$$

as $y \to w$ by continuity in terms of the second parameter of $f(x, w) = f_w(x)$.

In words, the distance between two fixed points x_y, x_w is bounded by a universal constant times the distance between one, x_y and the map f_y applied to the other, x_w. This means that one application of the map moves x_w very close to x_y. This equals the distance between the two different maps applied to the same point, which is uniformly controlled by continuity of the family of maps.

□

Definition 37.4. For an interval $J = [a, b]$, we write $\mathcal{C}_J = \mathcal{C}(J, \mathbb{R}^n)$. The sup norm on this space is defined for $f \in \mathcal{C}_J$, by

$$||f||_J^\infty \equiv \sup_{t \in J} ||f(t)||$$

where $||f(t)||$ is the Euclidean norm of that vector in $\mathbb{R}^3$.

Proposition 37.8. *Let $\mathcal{C}_I = \mathcal{C}(I, \mathbb{R})$ denote the vector space of continuous functions from $I = [0, 1]$ to $\mathbb{R}$. Give $\mathcal{C}_I$ the metric coming from the sup norm. This makes $\mathcal{C}_I$ a complete metric space.*

Proof. Let f_n be a Cauchy sequence. Then for each fixed x, $f_n(x)$ is a Cauchy sequence in $\mathbb{R}$, hence converges to a point y as $\mathbb{R}$ is complete. We do this for each x and define the function $f(x) = y$. Thus $f_n \to f$ pointwise. But this is uniform convergence, hence the limiting function f is uniformly continuous, by a triangle inequality argument. □

Definition 37.5. Let $\mathcal{C} = \mathcal{C}(\mathbb{R}, \mathbb{R}^n)$ denote the vector space of continuous functions from $\mathbb{R}$ to $\mathbb{R}^n$. By the topology of uniform convergence on compact subsets, we mean that $f_n \to f$ iff for each compact subinterval J, $||f_n = f||_J^\infty \to 0$.

Proposition 37.9. *Let $\mathcal{C} = \mathcal{C}(\mathbb{R}, \mathbb{R}^n)$, with the topology of uniform convergence on compact subsets. Suppose that f_n is Cauchy on each compact interval. Then there exists a unique $f \in \mathcal{C}$ such that $f_n \to f$.*

Proof. We apply the previous Proposition to the coordinates, whence f_n is a Cauchy sequence on each interval $J = [a, b]$. The result follows. □

Higher derivatives for maps of Euclidean space. We recall some basics from vector calculus. Now given a map $f : \mathbb{R}^m \to \mathbb{R}^n$, the derivative at a point is the best linear approximation at that point, given by an $n \times m$ matrix, the matrix of partial derivatives. For the next section, we need to recall the notion of higher derivatives for vector functions. See p. 5 of [Lan02]. Writing E for Euclidean space $\mathbb{R}^m$ and F for $\mathbb{R}^n$, and $L(E, F)$ for the collection of linear maps from E to F, then $Df : E \to L(E, F) \cong \mathcal{M}_{n \times m}(\mathbb{R})$. Thus $D^2(f) \equiv D(Df) : E \to L(E, L(E, F))$, $D^3 f : E \to L(E, L(E, L(E, F)))$ and so on.

To understand the meaning of these formulas, given $\mathbf{p} \in E$ and $\mathbf{v}$ a vector in E then $\mathbf{v_p}$ is the vector $\mathbf{v}$ based at $\mathbf{p}$, and $Df_\mathbf{p}$ is an $(n \times m)$ matrix, equivalently a linear transformation from E to F, with $Df_\mathbf{p}(\mathbf{v_p}) = \mathbf{w} \in F$.

Now what does $D^2(f)_\mathbf{p}$ tell us? It says how this matrix-valued function varies, infinitesimally in a given direction, specified by a tangent vector $\mathbf{v_p}$.

This gives the best linear approximation to Df at $\mathbf{p}$, in the direction $\mathbf{v_p}$.

In other words, $(Df \circ \gamma)(t)$ is a curve in the space of matrices, and we want to know the tangent vector to this curve in the direction $\mathbf{v_p} = \gamma'(0)$. This tangent vector is itself a matrix since the curve is matrix-valued.

This matrix (i.e. linear transformation) depends on $\mathbf{p}$ and $\mathbf{v_p}$. Thus $D^2(f)_\mathbf{p}$ is an element of $L(E, L(E, F))$.

Then for $\widehat{L} = D^2 f|_\mathbf{p} \in L(E, L(E, F))$, choosing $\mathbf{v_p}$ in the first E, $\widehat{L}(\mathbf{v_p}) \in L(E, F)$ is the matrix giving this change. To specify which matrix, we choose $\mathbf{u} \in E$, and

then $(\widehat{L}(\mathbf{v_p}))(\mathbf{u}) \in F$. Now the first L in $L(E, L(E, F))$ indicates the collection of linear maps, so if we have two vectors $\mathbf{v}_1, \mathbf{v}_2$ at $\mathbf{p}$, these matrices add, that is, $\widehat{L}(\mathbf{v}_1 + \mathbf{v}_2) = \widehat{L}(\mathbf{v}_1) + \widehat{L}(\mathbf{v}_2)$. Furthermore each matrix is a linear transformation, so given $\mathbf{u}_1, \mathbf{u}_2$ in the second E we have $(\widehat{L}(\mathbf{v}_1))(\mathbf{u}_1 + \mathbf{u}_2) = (\widehat{L}(\mathbf{v}_1))(\mathbf{u}_1) + (\widehat{L}(\mathbf{v}_1))(\mathbf{u}_2)$. The result is summarized by saying that $\widehat{L}$ is linear in both $\mathbf{u}$ and $\mathbf{v}$, in other words it is a bilinear map with values in F. Thus $\widehat{L} \in \mathcal{B}^2(E, F)$, the collection of such bilinear maps.

In the same way, then $D^3 f : E \to L(E, L(E, L(E, F))) \cong \mathcal{B}^3(E, F)$ and similarly, $D^k f : E \to \mathcal{B}^k(E, F)$, the k-multilinear maps from $E \cong \mathbb{R}^m$ to $F \cong \mathbb{R}^n$.

It is easy as can be to write this formula, but not so easy to understand what it means.

More generally, the derivative of a map from a manifold M to a manifold N, of dimensions m, n, is a linear transformation of the tangent bundles. That is, for $f : M \to N$, $Df|_{\mathbf{p}} : TM_{\mathbf{p}} \to TN_{\mathbf{q}}$ where $\mathbf{q} = f(\mathbf{p})$. Choosing charts, this is an $(n \times m)$ matrix, and we transport the above conclusions to the manifolds via these charts.

Thus we picture Df as a matrix-valued function, where the charts determine the basis in domain and range. Now, just as the function f can be thought of as a section of a fiber bundle over M with fibers N, now Df is a section of the *vector* bundle over M with fibers the vector space of matrices $L(E, F)$.

As before, we have $Df : E \to L(E, F)$ so for $\widehat{L} = D(Df)$ then $\widehat{L} \in L(E, L(E, F))$. Thus, $\widehat{L}(\mathbf{v_p})(\mathbf{u}) \in F$. For each $\mathbf{u_p} \in TM_{\mathbf{p}}$, there is a value of $\widehat{L}(\mathbf{v_p})$ on the vector $\mathbf{u_p}$. Again, this value is linear in both $\mathbf{u_p}$ and $\mathbf{v_p}$, so $\widehat{L}$ is bilinear with F-values. For $\widehat{L}(\mathbf{v_p}, \mathbf{u_p})$, the first coordinate $\mathbf{v}$ determines the direction of change, while the second $\mathbf{u_p}$ specifies this matrix by telling us how that matrix acts on $\mathbf{u_p}$.

We state the Chain Rule in this context.

Theorem 37.10. *Let $\gamma : \mathbb{R} \to E$ and $f : E \to F$. Then for $f \circ \gamma : \mathbb{R} \to F$ we have:*

$$D(f \circ \gamma)(t) = Df(\gamma(t)(\gamma'(t)).$$

Picard operator. As above, let $\mathcal{C} = \mathcal{C}(\mathbb{R}, \mathbb{R}^n)$ Now let $V : \mathbb{R}^n \to \mathbb{R}^n$ be K-Lipschitz.

We define a linear transformation of $\mathcal{C}$, the *Picard operator*, and use this to prove a number of results. First we show fixed points are solutions; then we demonstrate the spatial continuity of these solutions, assuming a fixed point exists for each initial condition. Next we use an iteration argument to prove existence, by showing the operator is a contraction mapping in an appropriate metric space. We show smoothness in the time parameter along orbits. After extending to the nonstationary situation, and following Arnold's proof, we show smoothness in the spatial parameter. This also makes an interesting link with the corresponding linear flows.

Given the vector field V we define $\mathcal{P} : \mathcal{C} \to \mathcal{C}$ by

$$(\mathcal{P}(\gamma))(t) = \gamma(0) + \int_0^t V(\gamma(s))\mathrm{d}s.$$

This is linear on $\mathcal{P}$. Thus by restriction for each fixed $\mathbf{v} \in \mathbb{R}^n$ we have an affine operator on the affine subspace $\mathcal{C}_{\mathbf{v}} \equiv \{\gamma \in \mathcal{C} : \gamma(0) = \mathbf{v}\}$:

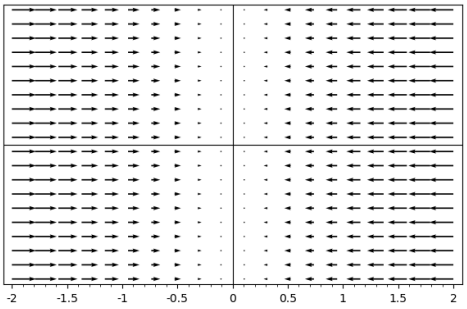


FIGURE 109. Picard operator: the fixed points form an attractor, a submanifold of $\mathcal{C}$ of dimension one.

$$(\mathcal{P}_{\mathbf{v}}(\gamma))(t) = \mathbf{v} + \int_0^t V(\gamma(s))\mathrm{d}s.$$

Remark 37.5. As we shall prove, for each chosen $\mathbf{v} \in \mathbb{R}^n$, the operator maps $\mathcal{C}$ to $\mathcal{C}_{\mathbf{v}}$, and is an eventual contraction.

Now each $\mathcal{P}_{\mathbf{v}}$ is an affine operator, the restriction of the linear Picard operator $\mathcal{P}$ on $\mathcal{C}$ to $\mathcal{C}_{\mathbf{v}}$. On the full space $\mathcal{C}$ itself, $\mathcal{P}$ is a contracton to an n-dimensonal submanifold of the infinite-dimensional space $\mathcal{C}$, the image of the map $\mathbf{v} \mapsto \gamma_{\mathbf{v}}$. See Fig. 109.

37.9. **Picard iteration: examples.**

Example 53. An already interesting example, presented in a number of texts, is Picard iteration for the most basic ODE: that describing exponential growth, the equation $y' = y$. Defining the vector field V on $\mathbb{R}$ by $V(y) = y$, then for $y = \gamma(t)$, this can be rewritten as

$$y' = \gamma'(t) = V(\gamma(t)) = \gamma(t) = y.$$

We iterate the Picard operator beginning with the constant path $\gamma_0 \equiv 1$. The initial condition will be $y(0) = 1$, so $\mathbf{v} = 1$. We have:

$$(\mathcal{P}_{\mathbf{v}}(\gamma_0))(t) = \mathbf{v} + \int_0^t V(\gamma(s))\ \mathrm{d}s = \gamma_1(t) = 1 + \int_0^t V(1)\ \mathrm{d}s = 1 + \int_0^t 1\,\mathrm{d}s = 1 + t.$$

The next iterate is

$$\gamma_2(t) = 1 + \int_0^t V(1+s)\ \mathrm{d}s = 1 + \int_0^t 1 + s\ \mathrm{d}s = 1 + t + t^2/2,$$

then

$$\gamma_3(t) = 1 + \int_0^t 1 + s + s^2/2\ \mathrm{d}s = 1 + t + t^2/2 + t^3/3!$$

and so on, with $\mathcal{P}^n_{\mathbf{v}}(\gamma_0) = p^{(n)}(t)$, the n^{th} Taylor polynomial for what we know from above is the unique solution, $y(t) = e^t$.

Picard iteration for autonomous linear (homogeneous) vector DE.

Next let A be an $(n \times n)$ matrix, and V the linear vector field on $\mathbb{R}^n$ defined by $V(\mathbf{x}) = A\mathbf{x}$. Then the DE $\mathbf{x}' = A\mathbf{x}$ can be rewritten as $\mathbf{x}' = \gamma'(t)$ and so $\gamma'(t) = V(\gamma(t))$.

We iterate the Picard operator beginning with the constant path $\gamma_0 \equiv 1$. The initial condition is $\mathbf{v} = \mathbf{x}_0$. We have:

$$\gamma_1(t) = (\mathcal{P}_{\mathbf{v}}(\gamma_0))(t) = \mathbf{v} + \int_0^t V(\gamma_0(s))\, \mathrm{d}s = \mathbf{v} + \int_0^t V(\mathbf{v})\, \mathrm{d}s = \mathbf{v} + \int_0^t A\mathbf{v}\, \mathrm{d}s = \mathbf{v} + tA\mathbf{v}.$$

Next we have

$$\gamma_2(t) = \mathbf{v} + \int_0^t V(\mathbf{v} + tA\mathbf{v})\, \mathrm{d}s = \mathbf{v} + \int_0^t A\mathbf{v} + tA^2\mathbf{v}\, \mathrm{d}s = \mathbf{v} + tA\mathbf{v} + t^2/2 \cdot A^2\mathbf{v} = (I + tA + t^2/2 \cdot A^2)\mathbf{v}$$

and then

$$\gamma_k(t) = (I + tA + \frac{t^2}{2}A^2 + \frac{t^3}{3!}A^3 + \cdots + \frac{t^k}{k!}A^k)\mathbf{v}$$

so

$$\gamma_k(t) \to e^{tA}\mathbf{v}.$$

This holds for all vectors $\mathbf{v}$.

Example 54. **(Rotation flow via Picard iteration)** In Proposition 35.15 we have seen that for $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ then its powers are $(A^0, A^1, A^2, A^3, \dots) = (I, A, -I, -A, \dots)$, whence

$$e^{tA} = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}$$

where $c = \cos t$, $s = \sin t$. We next see this in a different way, using Picard iteration.

Thus, as just calculated, for initial condition $\mathbf{v}$ we have:

$$\begin{aligned} \gamma_3(t) &= (I + tA + t^2/2A^2 + t^3/3!A^3)\mathbf{v} = \\ &= (1 - t^2/2)I\mathbf{v} + (t - t^3/3!)A\mathbf{v} = \\ &= \begin{bmatrix} 1 - t^2/2 & -t + t^3/3! \\ t - t^3/3! & 1 - t^2/2 \end{bmatrix} \mathbf{v} \end{aligned}$$

which as calculated in Example 55 converges as $k \to \infty$ to

$$e^{tA}\mathbf{v} = \begin{bmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{bmatrix} \mathbf{v} = R_t\mathbf{v}.$$

Example 55. **(Hyperbolic rotation flow via Picard iteration)** We return to our study in Examples 43 and 52. Here $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Now $A^k = I$ for k even, $A^k = A$ for

k odd. So applying Picard iteration, then as just calculated we have:

$$\gamma_3(t) = (I + tA + t^2/2A^2 + t^3/3!A^3)\mathbf{v} =$$
$$= (1 + t^2/2)I\mathbf{v} + +(t + t^3/3!)A\mathbf{v} =$$
$$= \begin{bmatrix} 1 + t^2/2 & t + t^3/3! \\ t + t^3/3! & 1 + t^2/2 \end{bmatrix} \mathbf{v}$$

which as calculated in Example 55 converges as $k \to \infty$ to

$$e^{tA}\mathbf{v} = \begin{bmatrix} \cosh(t) & \sinh(t) \\ \sinh(t) & \cosh(t) \end{bmatrix} \mathbf{v}$$

providing an alternative way of understanding the Taylor's series for these hyperbolic functions.

37.10. **Picard operator: Smoothness in time of solution curves; spatial continuity; contraction property.** Next we turn to the theory of the Picard operator, beginning with properties of a fixed point if one exists. Following that we show the existence (and uniqueness), as a consequence of the contraction property.

Lemma 37.11.
(i) γ is a fixed point for $\mathcal{P}_{\mathbf{v}}$ for all $\delta > 0$, iff γ is a solution for the vector differential equation: $\gamma'(t) = V(\gamma(t))$ for all t, with initial condition $\gamma(0) = \mathbf{v}$.
(ii) If V is a $\mathcal{C}^k$ vector field, then a fixed point γ has one more degree of smoothness: the curve $t \mapsto \gamma(t)$ is $\mathcal{C}^{k+1}$, for $k \geq 0$.

Proof. (i) If γ is a solution, thus $\gamma(0) = \mathbf{v}$ and $\gamma'(t) = V(\gamma(t))$, then

$$(\mathcal{P}_{\mathbf{v}}(\gamma))(t) = \mathbf{v} + \int_0^t V(\gamma(s))\, \mathrm{d}s = \mathbf{v} + \int_0^t \gamma'(s)\, \mathrm{d}s = \mathbf{v} + \gamma(t) - \gamma(0) = \gamma(t).$$

Conversely, if $\mathcal{P}_{\mathbf{v}}(\gamma) = \gamma$ then $\mathbf{v} + \int_0^t V(\gamma(s))\, \mathrm{d}s = \gamma(t)$ for all t, so differentiating, $\gamma'(t) = V(\gamma(t))$ and furthermore, $\gamma(0) = \mathbf{v} + \int_0^0 V(\gamma(s))\, \mathrm{d}s = \mathbf{v}$.

(ii) We consider $k = 0$, so V is continuous. Then since $\mathcal{P}_{\mathbf{v}}(\gamma) = \gamma$, $\gamma'(t) = \mathrm{d}/\mathrm{d}t(\mathbf{v} + \int_0^t V(\gamma(s))\, \mathrm{d}s) = V(\gamma(t))$ whence γ is differentiable hence continuous. Moreover $\gamma' = V \circ \gamma$ which we now know to be continuous, so γ is in fact $\mathcal{C}^1$.

Now if V is $\mathcal{C}^1$, thus $DV : \mathbb{R}^n \to \mathcal{M}_n(\mathbb{R})$ is continuous, then since as before $\gamma'(t) = V(\gamma(t))$ then by the Chain Rule, Theorem 37.10, $\gamma''(t) = DV|_{\gamma(t)}\gamma'(t)$ is continuous hence γ is $\mathcal{C}^2$.

We recall the statement of the Product Rule for matrix-values curves $A(t), B(t)$, part (ii) of Lemma 35.63: this states that $(AB)' = A'B + AB'$.

If V is $\mathcal{C}^2$, then $\gamma''(t) = DV|_{\gamma(t)}\gamma'(t)$ so $\gamma'''(t) = \left(DV|_{\gamma(t)}\gamma'\right)'(t)$. Now $A(t) = DV \circ \gamma(t) = DV|_{\gamma(t)}$ is a continuous curve in $\mathcal{M}_n$, as is $\gamma'(t)$ (these are $(n \times n)$ and the column vector $(n \times 1)$ matrices respectively) so $(AB)'(t) = A'B + AB'(t) = (DV|_{\gamma(t)})'(\gamma')(t) + DV|_{\gamma(t)}(\gamma'')(t)$. This exists and is continous. Thus γ is $\mathcal{C}^3$.

Now $\gamma'''(t) = (AB)'(t) = A'B + AB'(t) = (DV|_{\gamma(t)})'(\gamma')(t) + DV|_{\gamma(t)}(\gamma'')(t)$ so

$$\gamma^{(4)}(t) = (AB)'' = (A'B + AB')' = A''B + 2A'B' + AB''$$
$$= (DV''|_{\gamma(t)})\gamma'(t) + 2(DV'|_{\gamma(t)})\gamma''(t) + (DV'|_{\gamma(t)})\gamma'''(t),$$

whence $\gamma \in \mathcal{C}^5$.

Next we have $\gamma^{(5)}(t) = A'''B + 3A''B' + 3A'B'' + AB'''$ and so on, with the binomial coefficients of Pascal's triangle. By induction, the general statement is true.

□

Remark 37.6. See p. 61 of Lang's book, [Lan02]. (Lang leavs out the added degree of smoothness for γ, but note that this also makes sense conceptually because it is an "integral curve" of the vector field.) See e.g. p. 270 of [HK03].

Proposition 37.12. *(local fixed point theorem; continuous dependence on initial condition)*
(i) Let $\mathcal{C}_{\mathbf{v}} \subseteq \mathcal{C} = \mathcal{C}(\mathbb{R}, \mathbb{R}^n)$ be the subset of all continuous paths γ such that $\gamma(0) = \mathbf{v}$, and let $\mathcal{C}_{\mathbf{v},\delta}$ denote the paths in $\mathcal{C}_{\mathbf{v}}$ restricted to t in the interval $J_\delta = [-\delta, \delta]$. Then for δ sufficiently small ($< 1/K$, where as above, $V : \mathbb{R}^n \to \mathbb{R}^n$ is K-Lipschitz) $\mathcal{P}_{\mathbf{v}}$ is a contraction mapping on $\mathcal{C}_{\mathbf{v},\delta}$, with constant $c < 1$.
(ii) There exists a unique fixed point $\gamma \in \mathcal{C}_{\mathbf{v},\delta}$ for each chosen $\mathbf{v}$.
(iii) The local solution curves are spatially continuous, that is, $\gamma_{\mathbf{v}} : J_\delta \to \mathbb{R}^n$ depends continuously on the initial value $\mathbf{v}$. In fact, this dependence is Lipschitz.

Proof. (i): Let $\gamma, \widetilde{\gamma} \in \mathcal{C}_{\mathbf{v}}$. Then

$$\mathcal{P}_{\mathbf{v}}(\gamma)(t) - \mathcal{P}_{\mathbf{v}}(\widetilde{\gamma})(t) = \mathbf{v} + \int_0^t V(\gamma(s))\, \mathrm{d}s - \mathbf{v} - \int_0^t V(\gamma(s))\, \mathrm{d}s = \int_0^t V(\gamma(s)) - V(\widetilde{\gamma}(s))\, \mathrm{d}s.$$

We write $||\mathbf{v}||$ for the norm of a vector in $\mathbb{R}^n$ and $||f||_J^\infty$ for the sup norm of an $\mathbb{R}^n$-valued function f over the interval $J \equiv [-\delta, \delta]$.

Then since V is K- Lipschitz, we have for all $t \in \mathbb{R}$,

$$||\mathcal{P}_{\mathbf{v}}(\gamma)(t) - \mathcal{P}_{\mathbf{v}}(\widetilde{\gamma})(t)|| = \left|\left| \int_0^t V(\gamma(s)) - V(\widetilde{\gamma}(s))\, \mathrm{d}s \right|\right| \leq \left| \int_0^t ||V(\gamma) - V(\widetilde{\gamma})||(s)\, \mathrm{d}s \right| \leq$$
$$\left| \int_0^t K||\gamma - \widetilde{\gamma}||(s)\, \mathrm{d}s \right|. \tag{143}$$

For any interval $J \subseteq \mathbb{R}$ and all $t \in J$ this is

$$\leq K||\gamma - \widetilde{\gamma}||_J^\infty \left| \int_0^t \mathrm{d}s \right| \leq K|t| \cdot ||\gamma - \widetilde{\gamma}||_J^\infty.$$

And $K|t| \cdot ||\gamma - \widetilde{\gamma}||_J^\infty \leq K\delta \cdot ||\gamma - \widetilde{\gamma}||_J^\infty$ so choosing $\delta < 1/K$ then $c = \delta K < 1$ and the bound is $c||\gamma - \widetilde{\gamma}||_J^\infty$ so the map is a contraction.

(ii): Thus by the Contraction Mapping Principle, Lemma 37.7, $\mathcal{P}_{\mathbf{v}}$ restricted to the space $\mathcal{C}_{\mathbf{v},\delta}$ has a unique fixed point $\gamma : J \to \mathbb{R}^n$.

(iii): Our proof follows the argument in Lang, p. 63 and makes use of the Picard operator. (A general proof for a parametrized family of contraction mapppings is also possible, see p. 68 of [HK03] but this is a bit simpler, plus we need these estimates below in ???.) We consider the space of curves $\mathcal{C} = \mathcal{C}(\mathbb{R}, \mathbb{R}^n)$. We define, for each $\mathbf{v} \in \mathbb{R}^n$, the operator $\mathcal{P}_{\mathbf{v}}$ on $\mathcal{C}$ by $(\mathcal{P}_{\mathbf{v}}(\gamma))(t) = \mathbf{v} + \int_0^t V(\gamma(s))\, \mathrm{d}s$ for $\gamma \in \mathcal{C}$. Now

$\mathcal{P}_{\mathbf{v}}(\gamma)(0) = \mathbf{v}$. Let $\gamma_{\mathbf{v}}$ denote the fixed point for the map $\mathcal{P}_{\mathbf{v}}$ guaranteed by part (ii). Let also $\mathbf{w} \in \mathbb{R}^n$. Then for J an interval in $\mathbb{R}$,

$$||\gamma_{\mathbf{v}} - \mathcal{P}_{\mathbf{w}}(\gamma_{\mathbf{v}})||_J^\infty = ||\mathcal{P}_{\mathbf{v}}(\gamma_{\mathbf{v}}) - \mathcal{P}_{\mathbf{w}}(\gamma_{\mathbf{v}})||_J^\infty = ||\mathbf{v} - \mathbf{w}||$$

as the parts under the integral are identical so cancel. (This also holds for $J = \mathbb{R}$). Now we take J to be the interval $[-\delta, \delta]$ above which gives a contraction with constant $c \in [0, 1)$. Iterating and applying the triangle inequality,

$$\begin{aligned}
&||\gamma_{\mathbf{v}} - \mathcal{P}_{\mathbf{w}}^n\gamma_{\mathbf{v}}||_J^\infty = ||\gamma_{\mathbf{v}} - \mathcal{P}_{\mathbf{w}}\gamma_{\mathbf{v}} + \mathcal{P}_{\mathbf{w}}\gamma_{\mathbf{v}} - \mathcal{P}_{\mathbf{w}}^2\gamma_{\mathbf{v}} + \cdots + \mathcal{P}_{\mathbf{w}}^{n-1}\gamma_{\mathbf{v}} - \mathcal{P}_{\mathbf{w}}^n\gamma_{\mathbf{v}}||_J^\infty \\
&\leq |\gamma_{\mathbf{v}} - \mathcal{P}_{\mathbf{w}}\gamma_{\mathbf{v}}||_J^\infty + ||\mathcal{P}_{\mathbf{w}}\gamma_{\mathbf{v}} - \mathcal{P}_{\mathbf{w}}^2\gamma_{\mathbf{v}}||_J^\infty + \cdots + ||\mathcal{P}_{\mathbf{w}}^{n-1}\gamma_{\mathbf{v}} - \mathcal{P}_{\mathbf{w}}^n\gamma_{\mathbf{v}}||_J^\infty \\
&= ||\gamma_{\mathbf{v}} - \mathcal{P}_{\mathbf{w}}\gamma_{\mathbf{v}}||_J^\infty + \cdots + ||\mathcal{P}_{\mathbf{w}}^{n-1}(\gamma_{\mathbf{v}} - \mathcal{P}_{\mathbf{w}}\gamma_{\mathbf{v}})||_J^\infty \\
&\leq (1 + c + \cdots + c^{n-1})||\gamma_{\mathbf{v}} - \mathcal{P}_{\mathbf{w}}(\gamma_{\mathbf{v}})||_J^\infty \leq \sum_0^\infty c^n||\mathbf{v} - \mathbf{w}|| = \frac{1}{1-c}||\mathbf{v} - \mathbf{w}||
\end{aligned}$$

since $0 \leq c < 1$.

So $||\gamma_{\mathbf{v}} - \gamma_{\mathbf{w}}||_J^\infty = \lim_{n\to\infty} ||\gamma_{\mathbf{v}} - \mathcal{P}_{\mathbf{w}}^n\gamma_{\mathbf{v}}||_J^\infty \leq \frac{1}{1-c}||\mathbf{v} - \mathbf{w}||$, proving continuity of the map $\mathbf{v} \mapsto \gamma_{\mathbf{v}}$. Moreover, this last line shows the map $\mathbf{v} \mapsto \gamma_{\mathbf{v}}$ on $\mathcal{C}_J$, from initial condition to the unique local fixed point, is Lipschitz with constant $1/(1-c)$. This constant is ≥ 1 but can be chosen as close to 1 as we wish by using a higher iterate and hence a lesser c. □

Definition 37.6. We define a topology on the space of curves $\mathcal{C}(\mathbb{R}, \mathbb{R}^n)$ by declaring that $\gamma_n \to \gamma$ iff for every compact interval $I \subseteq \mathbb{R}$, $||\gamma_n - \gamma||_I^\infty \to 0$. (It is easy to define a subbase for a topology which gives this notion of sequential convergence). This is called the topology of uniform convergence on compact subsets of $\mathbb{R}$.

We recall the following little result: Lemma 16.4.

Lemma 37.13. *Let $f : X \to X$ be a function on a set such that for some $m > 1$, f^m has a unique fixed point. Then the same is true for f.*

Proof. Let x be the unique fixed point for f^m. Then x is a periodic point for f, of (least) period k which divides m. We want to show that $k = 1$. But this is true, since the orbit of x provides k distinct fixed points for f^m. Lastly, uniqueness for f follows from uniqueness for f^m. □

Theorem 37.14. *(Global solutions: existence and uniqueness; fixed point theorem; continuous dependence on initial condition in topology of uniform convergence on compacts)*
(i)For each chosen $\mathbf{v}$, there exists a unique fixed point $\gamma(t)$, for all $t \in \mathbb{R}$.
(ii) Given any compact interval $J \subseteq \mathbb{R}$, the operator $\mathcal{P}_{\mathbf{v}}$ is eventually a contraction on $\mathcal{C}(J, \mathbb{R}^n)$. That is, there exists $c \in [0, 1)$ and $N \geq 0$, depending on J, such that for any $n \geq N$, $d(\mathcal{P}_{\mathbf{v}}^n(\gamma), \mathcal{P}_{\mathbf{v}}^n(\widetilde{\gamma})) < c^n$. (By Lemma 37.13, this will give a second proof of (i).)
(iii) The global solution curves $\gamma_{\mathbf{v}} : \mathbb{R} \to \mathbb{R}^n$ depend continuously on the initial value $\mathbf{v}$, in the topology of uniform convergence on compact subsets of $\mathbb{R}$, and moreover are Lipschitz for any fixed interval length, with Lipschitz constant arbitarily close to 1.

Proof. (i) The proof is by compatibility of the definition on neighboring intervals. Given $\mathbf{v}_0 = \mathbf{v}$, and $\delta < 1/(2K)$, we have proved there is a unique solution γ on $[-\delta, \delta]$. Define $\mathbf{v}_1 = \gamma(\delta)$. Applying the theorem to the initial condition $\mathbf{v}_1$, there exists a unique curve γ_1 with $\gamma_1(0) = \mathbf{v}_1$ satisfying the equation and defined on $[0, 2\delta]$. Now define γ on $[-\delta, 2\delta]$ to be $\gamma(t) = \gamma_1(t)$ on $[-\delta, \delta]$, and $\gamma(t) = \gamma_2(t - \delta)$ on $[0, 2\delta]$. There are two definitions on the interval $[0, \delta]$ but these agree by uniqueness, as both give a solution on that interval with initial value $\mathbf{v}_0$. Continuing in this way for negative and positive multiples of δ completes the proof.

(ii) To prove $\mathcal{P}_{\mathbf{v}}$ is an eventual contraction, we replace the curve γ by the curve $\mathcal{P}_{\mathbf{v}}(\gamma)$ in Equation (149) we have for any $t \in \mathbb{R}$,

$$\begin{aligned} &||\mathcal{P}^2_{\mathbf{v}}(\gamma)(t) - \mathcal{P}^2_{\mathbf{v}}(\widetilde{\gamma})(t)|| = \left|\left|\int_0^t V(\mathcal{P}_{\mathbf{v}}(\gamma)(s)) - V(\mathcal{P}_{\mathbf{v}}(\widetilde{\gamma})(s))\, \mathrm{d}s\right|\right| \leq \\ &\left|\int_0^t ||V(\mathcal{P}_{\mathbf{v}}(\gamma)) - V(\mathcal{P}_{\mathbf{v}}(\widetilde{\gamma}))||(s)\, \mathrm{d}s\right| \leq \left|\int_0^t K||\mathcal{P}_{\mathbf{v}}(\gamma) - \mathcal{P}_{\mathbf{v}}(\widetilde{\gamma})||(s)\, \mathrm{d}s\right| \leq \\ &K\left|\int_0^t \left|\int_0^s K||\gamma - \widetilde{\gamma}||(x)\, \mathrm{d}x\right| \mathrm{d}s\right| \leq\ K^2||\gamma - \widetilde{\gamma}||^\infty_J \left|\int_0^t \left|\int_0^s \mathrm{d}x\right| \mathrm{d}s\right| \end{aligned} \tag{144}$$

Now

$$\left|\int_0^t \left|\int_0^s \mathrm{d}x\right| \mathrm{d}s\right| \leq \left|\int_0^t s\, \mathrm{d}s\right| \leq t^2/2$$

Thus our bound is

$$\leq\ K^2||\gamma - \widetilde{\gamma}||^\infty_J t^2/2.$$

This doesn't look like much progress, but iterating one more time we have for any $t \in \mathbb{R}$,

(149)

$$\begin{aligned} &||\mathcal{P}^3_{\mathbf{v}}(\gamma)(t) - \mathcal{P}^3_{\mathbf{v}}(\widetilde{\gamma})(t)|| = \left|\left|\int_0^t V(\mathcal{P}^2_{\mathbf{v}}(\gamma)(s)) - V(\mathcal{P}^2_{\mathbf{v}}(\widetilde{\gamma})(s))\, \mathrm{d}s\right|\right| \leq \\ &K^3||\gamma - \widetilde{\gamma}||^\infty_J \left|\int_0^t s^2/2\, \mathrm{d}s\right| =\ K^3||\gamma - \widetilde{\gamma}||^\infty_J t^3/3! \end{aligned} \tag{145}$$

We claim that for n iterates we get the bound of $K^n||\gamma - \widetilde{\gamma}||^\infty_J t^n/n!$.

To prove this by induction, we assume that we have this for n and prove for $(n+1)$. Thus, assuming that

$$||\mathcal{P}^n_{\mathbf{v}}(\gamma)(t) - \mathcal{P}^n_{\mathbf{v}}(\widetilde{\gamma})(t)|| \leq K^n \cdot t^n/n!$$

then

$$\begin{aligned} &||\mathcal{P}^{n+1}_{\mathbf{v}}(\gamma)(t) - \mathcal{P}^{n+1}_{\mathbf{v}}(\widetilde{\gamma})(t)|| = \left|\left|\int_0^t V(\mathcal{P}_{\mathbf{v}}(\mathcal{P}^n_{\mathbf{v}}\gamma)(s)) - V(\mathcal{P}_{\mathbf{v}}(\mathcal{P}^n_{\mathbf{v}}\gamma)(s))\, \mathrm{d}s\right|\right| \leq \\ &K\left|\int_0^t ||\mathcal{P}^n_{\mathbf{v}}(\gamma) - \mathcal{P}^n_{\mathbf{v}}(\widetilde{\gamma})||(s)\, \mathrm{d}s\right| \leq \\ &K^{n+1}\left|\int_0^t s^n/n!\, \mathrm{d}s\right| \leq K^{n+1} \cdot t^{n+1}/(n+1)! \end{aligned} \tag{146}$$

proving our claim. For $t \in J = [-R, R]$ then the bound is

$$\leq ||\gamma - \widetilde{\gamma}||_J^\infty \cdot K^n R^n / n!$$

which is exponentially decreasing as soon as $n > KR$, showing the eventual contraction on $J = [-R, R]$.

From the Lemma, the eventual contraction proves that not only does $\mathcal{P}_{\mathbf{v}}^n$ therefore have a unique fixed point, but so does $\mathcal{P}_{\mathbf{v}}$.

We learned the above strengthened estimate from ??? of (Fischman-Salam). See also p. 14 of [Sot79].

To prove (*iii*), we start with the estimate from (*iii*) of Proposition 37.10. There we concluded for a contraction$P_{\mathbf{w}}$ with constant $c < 1$ on $\mathcal{C}(J, \mathbb{R}^d)$ that $||\gamma_{\mathbf{v}} - \gamma_{\mathbf{w}}||_J^\infty = \lim_{n\to\infty} ||\gamma_{\mathbf{v}} - \mathcal{P}_{\mathbf{w}}^n \gamma_{\mathbf{v}}||_J^\infty \leq \frac{1}{1-c}||\mathbf{v} - \mathbf{w}||$, thereby proving continuity of the map $\mathbf{v} \mapsto \gamma_{\mathbf{v}}$, and further, that the map $\mathbf{v} \mapsto \gamma_{\mathbf{v}}$ is Lipschitz with constant $1/(1-c)$. Here we have shown that some power $P_{\mathbf{w}}^m$ is a contraction with constant c, so we simply apply the above to $P_{\mathbf{w}}^{nm}$ to reach the same conclusion for any given interval J. As before, the Lipschitz constant is ≥ 1 and can be taken as close to 1 as we wish. □

Definition 37.7. We define for each $s \in \mathbb{R}$ the shift map σ_s on *path space* $\mathcal{C}(\mathbb{R}, \mathbb{R}^d)$ by $\sigma_s(\gamma)(t) = \gamma(t + s)$. (This is a flow). Given a vector field V on $\mathbb{R}^n$, we define $\mathcal{S} \subseteq \mathcal{C}(\mathbb{R}, \mathbb{R}^d)$ to be the collection of all solution curves: those γ such that $\gamma'(t) = V(\gamma(t))$.

Proposition 37.15. *(Time-independence of paths; existence of flow.)*
(i) The solutions $\gamma_{\mathbf{v}}(t)$ with initial condition $\mathbf{v}$ satisfy the following: we have $\mathbf{v} = \gamma_{\mathbf{v}}(0)$; defining $\mathbf{w} = \gamma_{\mathbf{v}}(t)$ and $\mathbf{u} = \gamma_{\mathbf{v}}(t + s)$ then $\mathbf{u} = \gamma_{\mathbf{w}}(s)$. That is,

$$\gamma_{\mathbf{v}}(t + s) = \gamma_{\mathbf{w}}(s).$$

(ii) Equivalently, for a solution $\gamma_{\mathbf{v}}$, $\sigma_s(\gamma_{\mathbf{v}}) = \gamma_{\mathbf{w}}$ where $\mathbf{w} = \gamma_{\mathbf{v}}(s)$. Thus the subset $\mathcal{S}$ of solution curves is invariant for the shift flow $(\sigma_s)_{s\in\mathbb{R}}$.
(iii) Defining a collection $(\tau_t)_{t\in\mathbb{R}}$ of maps of $\mathbb{R}^n$ by the equation $\tau_t(\mathbf{v}) = \gamma_{\mathbf{v}}(t)$ where $\gamma_{\mathbf{v}}$ is the unique solution with initial condition $\mathbf{v}$, then τ is a flow. Moreover, it is a continuous flow on $\mathbb{R}^n$, and for each $\mathbf{v}$, the time derivative along an orbit curve, $d/dt(\tau_t(\mathbf{v}))$ exists.
(iv) There exists a unique continuous flow τ_t on $\mathbb{R}^n$ such that for any $\mathbf{v}$, the path $\gamma_{\mathbf{v}}$ defined by $\gamma_{\mathbf{v}}(t) = \tau_t(\mathbf{v})$ is the unique solution with initial value $\mathbf{v}$. If V is $\mathcal{C}^k$, then the orbits are $\mathcal{C}^{k+1}$, and the flow is $\mathcal{C}^k$.
(v) The map $\mathbf{v} \mapsto \gamma_{\mathbf{v}}$ is a continuous isomorphism from the flow $(\mathbb{R}^n, \tau_t)$ to the shift flow $(\mathcal{S}, \sigma_t)$.

Proof. (*i*) This is just the property of additivity with respect to time of the integral. That is, $\mathbf{w} = \gamma_{\mathbf{v}}(t) = \mathbf{v} + \int_0^t V(\gamma_{\mathbf{v}}(r))\mathrm{d}r$, so:

$$\gamma_{\mathbf{v}}(t + s) \equiv \mathbf{u} = \mathbf{v} + \int_0^{t+s} V(\gamma_{\mathbf{v}}(r))\mathrm{d}r = \mathbf{v} + \int_0^t V(\gamma_{\mathbf{v}}(r))\mathrm{d}r + \int_t^{t+s} V(\gamma_{\mathbf{v}}(r))\mathrm{d}r =$$

$$\mathbf{w} + \int_t^{t+s} V(\gamma_{\mathbf{v}}(r))\mathrm{d}r = \mathbf{w} + \int_0^s V(\gamma_{\mathbf{w}}(r))\mathrm{d}r = \gamma_{\mathbf{w}}(s).$$

(*ii*) follows from this.

(*iii*): Since we know that for any given $\mathbf{v} \in \mathbb{R}^n$, there exists a unique solution $\gamma_{\mathbf{v}}$ with that initial value, the equation $\tau_t(\mathbf{v}) = \gamma_{\mathbf{v}}(t)$ gives us a well-defined function

$\tau_t : \mathbb{R}^n \to \mathbb{R}^n$, using both properties: existence to get a value, uniqueness to know it is a single value. Defining $\mathbf{w} \equiv \gamma_{\mathbf{v}}(t)$, then from (i), we get the flow property: $\tau_{s+t}(\mathbf{v}) = \gamma_{\mathbf{v}}(t+s) = \gamma_{\mathbf{w}}(s) = \tau_s(\mathbf{w}) = \tau_s \circ \tau_t(\mathbf{v})$. Note also that $\tau_0(\mathbf{v}) = \mathbf{v}$, by existence. These two properties tell us it is a flow; note that each map τ_t is indeed a bijection, with inverse τ_{-t}.

(iv) We have just shown that for V a Lipschitz vector field, by the existence and uniqueness theorem, the equation defines a flow τ_t. Conversely, if there is a unique flow τ_t such that for each $\mathbf{v} \in \mathbb{R}^n$, $\gamma_{\mathbf{v}}$ defined by $\gamma_{\mathbf{v}}(t) = \tau_t(\mathbf{v})$ is a solution of the equation $\gamma'(t) = V(\gamma(t))$ with initial condition $\mathbf{v}$, i.e. $\gamma_{\mathbf{v}}(0) = \mathbf{v}$, then

Defining $\widetilde{\gamma}_{\mathbf{w}}(t) = \gamma_{\mathbf{v}}(t+s) = \gamma_{\mathbf{w}}(s)$....??? We know that given $\mathbf{v} \in \mathbb{R}^n$, there is a unique path $\gamma_{\mathbf{v}}(t)$ solving $\gamma'(t) = V(\gamma(t))$. We define $\tau_t : \mathbb{R}^n \to \mathbb{R}^n$ as follows:

$$\tau_t(\mathbf{v}) = \gamma_{\mathbf{v}}(t).$$

We claim this satisfies the flow property: $\tau_{t+s} = \tau_s \circ \tau_t$, that is, for $\mathbf{w} = \gamma_{\mathbf{v}}(t)$. we have $\gamma_{\mathbf{w}}(s) = \gamma_{\mathbf{v}}(t+s)$.

(v) ???

□

37.11. **Vector fields on Banach spaces; Nonstationary systems of ordinary differential equations.** We write $E = \mathbb{R}^d$, with the standard basis $\mathcal{B} = (\mathbf{e}_1, \ldots, \mathbf{e}_d)$, inner product and norm. Generalizing from this setting, and following Lang [Lan02], [Lan01], we now allow E to be a *Banach space*: a complete normed vector space. We recall some essential differences between finite and infinite dimensional topological vector spaces: in finite dimensions all norms are equivalent, and so convergence of a sequence of vectors $\mathbf{v}_n \to \mathbf{v}$ in any choice of norm, $||\mathbf{v}_n - \mathbf{v}|| \to 0$, is equivalent to convergence of coordinates, thus writing $\mathbf{v} = (v_1, \ldots v_d)$ then $(v_n)_k \to v_k$ for all $1 \leq k \leq d$. This holds since one of the possible norms is the sup norm (or L^∞ norm) $||\mathbf{v}|| = \sup_{i=1}^d |v_i|$.

These coordinates can be with respect to any choice of basis $\widetilde{\mathcal{B}} = (\mathbf{u}_1, \ldots, \mathbf{u}_d)$, and all such choices give an equivalent notion of convergence, again by equivalence of norms. The map $\lambda_i : \mathbf{v} \mapsto v_i$ where $\mathbf{v} = \sum_{i=1}^d v_i \mathbf{u}_i$ is an element of V^*, and any linear functional is a linear combination of these λ_i. Hence for finite dimensions, any linear function is continuous. This fails for infinite dimensions, where the dual space V^* of V is defined to be the space of all *continuous* lnear functionals, as those are nicely related to convergence. Indeed, we replace the notion of coordinates with respect to a basis with the coordinates given by these linear functionals. Thus, for $\lambda \in V^*$, $\lambda(\mathbf{v})$ is the "λ^{th}-coordinate" of $\mathbf{v}$, and we say $\mathbf{v}_n \to \mathbf{v}$ iff for all $\lambda \in V^*$, $\lambda(\mathbf{v}_n) \to \lambda(\mathbf{v})$. This defines what is called the *weak topology* of V, see [Rud73] p.65; the fact that there are different, natural topologies in infinite dimensions is much of what makes Functional Analysis (yes, the application of analysis ideas via linear functionals!) so fascinating and powerful.

Weak convergence works particularly well if V^* *separates points* on V, i.e. given $\mathbf{v} \neq \mathbf{w}$ then there exists $\lambda \in V^*$ such that $\lambda(\mathbf{v}) \neq \lambda(\mathbf{w})$. A locally convex Banach space always has this property, as a consequence of the Hahn-Banach Theorem, see [Rud73] p.60.

These coordinates given by V^* directly generailze coordinates with respect to a basis. The idea of linear combination then is extended to infinite series, with convergence in the norm or weak topologies. However for infinite dimensions there does not always exist a countable basis, which is why we use all of V^* to define coordinates, rather than a subset.

We also use functionals to define vector-valued integration. Thus e.g. for $F : \mathbb{R} \to V$ and measure μ on $\mathbb{R}$, $\int_{\mathbb{R}} F(t)\mathrm{d}\mu(t) = \mathbf{w}$ iff for each $\lambda \in V^*$, $\int_{\mathbb{R}} \lambda(F(t))\mathrm{d}\mu(t) = \lambda(\mathbf{w})$. See p.74 of [Rud73]. This notion will allow us for example to define the Picard operator for a vector field V on a Banach space E, by the same formula:

$$(\mathcal{P}_{\mathbf{v}}(\gamma))(t) = \mathbf{v} + \int_0^t V(\gamma(s))\ \mathrm{d}s.$$

In the next section we introduce nonstationary vector fields. For this purpose it is important to review some background, making the definitions precise.

Definition 37.8. We recall some definitions from Set Theory [Hal74]. Given sets X and Y, a *relation* R from X to Y is any subset $R \subseteq X \times Y$. We write xRy, read "x is related to y" when $(x, y) \in R$.

A *function* f from X to Y is a special type of relation, one such that each $x \in X$ is related to some $y \in Y$, but only one. That is, for each $x \in X \exists y :\ xRy$, and $(xRy) \wedge (xRw) \implies (y = w)$. We then write $f(x) = y$ if y is this unique element of Y. Given a function f let $R_f \subseteq X \times Y$ denote its relation. The *graph* of the function is the set of points $R_f = \{(x, f(x)) : x \in X\}$. In other words, from the set theory viewpoint, a function *is* its graph. The *image* of a function $f : X \to Y$ is $\mathrm{Im}(f) = \{y = f(x) :\ x \in X\}$.

By a *curve* γ in a Banach space E we mean a continuous function $\gamma : \mathbb{R} \to E$. The graph of the map γ is $\{(t, \gamma(t)) :\ t \in \mathbb{R}\}$. Thus $\gamma \in \mathcal{C}(\mathbb{R}, E)$ and also equals its graph, so $\gamma \subseteq \mathbb{R} \times E$.

Remark 37.7. This gives us two quite different interpretations, one static and one dynamical. The graph of f is is a set of points in the product space $\mathbb{R} \times E$, a static view. The dynamical interpretation is that the function f is a map-sending one point to another- which parametrizes its image.

This change of perspective involves viewing time as a spatial parameter, and is familiar from Calculus where a derivative of $f : \mathbb{R} \to \mathbb{R}$ is, from the dynamical point of view, a rate of change, while from the static perspective is the slope of the tangent line. Much of the power of the Calculus comes from the interplay of these very different points of view.

For example, consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ with $\gamma(t) = (\cos t, \sin t)$. This map parametrizes the unit circle; note that including the parameter serves to distinguish

the points $\gamma(t) = \gamma(s)$ if $s \neq t$. On the other hand, the graph is a set of points (a helix) in $\mathbb{R} \times \mathbb{R}^2$.

Definition 37.9. To make this distinction clearer we define: the *graph curve* of γ is the curve $\widehat{\gamma}$ where $\widehat{\gamma} : t \mapsto (t, \gamma(t))$. Thus the image of $\widehat{\gamma}$ is the graph of γ; $\widehat{\gamma}$ parametrizes the graph. It projects to the curve, via the identity map on the second coordinate, as $\pi : \mathbb{R} \times E \to E$; thus $\pi \circ \widehat{\gamma} = \gamma$ and $(t, \gamma(t) \mapsto \gamma(t)$. In the case of the above example, $\widehat{\gamma}(t) = (t, \gamma(t)) = (t, \cos t, \sin t) \in \mathbb{R} \times E$. Thus $\widehat{\gamma} \in \mathcal{C}(\mathbb{R}, \mathcal{C}(\mathbb{R}, E))$ where we topologize the space of paths $\mathcal{C}(\mathbb{R}, E))$ by uniform convergence on compact subsets.

We define $\widehat{E} = \mathbb{R} \times E$ and $E^t = \{t\} \times E \subseteq \widehat{E}$. This can be viewed as a vector fiber bundle over $\mathbb{R}$, with fiber E^t over the point t. It is a trivial bundle, indeed is a (global) product. This projects to E, sending E^t to E via the identity on the second coordinate, just as for the graph curve. Indeed, the graph curve $\widehat{\gamma}$ is a *section* of this fiber bundle: by definition, a choice of one point in the fiber for each base point t.

Now let $L(E) = L(E, E)$ denotes the linear transformations on E. For R^d via the choice of basis these are identified with $\mathcal{M}_{d\times d}$, the square matrices; for a Banach space, *linear* will mean here *continuous* linear (which is automatic for finite dimensions, as noted above).

A *vector field* V on E is an element of $\mathcal{V} \equiv \mathcal{C}(E) = \mathcal{C}(E, E)$. A *linear* vector field is $V \in L(E)$. A *nonstationary vector field* on E is a continuous curve in $\mathcal{V}$. Thus, it is a parametrized collection of vector fields, $(V^t)_{\mathbb{R}} = \{V^t : t \in \mathbb{R}\}$, with each vector field V^t defined on the same space E. It is *Lipschitz* if there is a uniform Lipschitz constant for each t.

The graph of this curve is

$$\{(t, V^t) : t \in \mathbb{R}\} \subseteq \mathbb{R} \times \mathcal{V}.$$

For each t, V^t is a vector field on $E^t = \{t\} \times E$. Its graph curve is $\widehat{V}$. This is a curve of vector fields on the fibers E^t of the fiber bundle $\widehat{E}$. The value at t is $\widehat{V}(t) = (t, V^t) \in \mathcal{C}(\mathbb{R}, \mathcal{V}(E^t))$.

This projects to the nonstationary vector field via the map π to the second coordinate, sending E^t to E.

An essential difference between the graph curve and the curve is that the spaces E^t, although projected to E via an isomorphism, are distinct spaces because the parameters are different. Thus we cannot add $\mathbf{v} \in E^t$ to $\mathbf{u} \in E^s$ unless $t = s$.

On the other hand, in E these vectors can be added, which allows us also to integrate and thereby define the Picard operator.

The nonstationary vector field is a a parametrized collection of vector fields $(V^t)_{\mathbb{R}}$ on the same space E whereas in $\widehat{V}$ we have $\widehat{V^t}$ defined on the distinct vector spaces E^t.

The way we picture $(V^t)_{\mathbb{R}}$ and its graph $\widehat{V}$ are quite different: the first we visualize as a fixed space E with a single changing vector field, the second as distinct vector fields on distinct spaces, all stacked up as fibers of a fiber bundle along the base space $\mathbb{R}$.

Definition 37.10. We define the *lift* of a nonstationary vector field $(V^t)_{\mathbb{R}}$: setting $\widehat{E} = \mathbb{R} \times E$, this is $\widehat{V} : \widehat{E} \to \widehat{E}$ where $\widehat{V}(t, \mathbf{v}) = (1, V^t(\mathbf{v}))$. This is a stationary vector field in one higher dimension, where time is treated like one more spatial dimension.

The *time derivative* γ' of a $\mathcal{C}^1$ curve γ is γ', the continuous map $t \mapsto \gamma'(t) \in E$ where the *tangent vector* to the curve at time t is defined to be

$$\gamma'(t) = \lim_{h \to \mathbf{0}} \frac{\gamma(t+h) - \gamma(t)}{||h||}$$

Equivalently, it is the unique vector such that, given $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\frac{||\gamma(t+h) - \gamma(t) - \gamma'(t)||}{||h||} < \varepsilon$$

for $||h|| < \delta$.

Thus γ' is a continuous curve in E.
We take the time derivative of the graph curve, its tangent vector. $\widehat{\gamma}(t) = (t, \gamma(t))$ so $\widehat{\gamma}'(t) = (t, \gamma(t))' = (1, \gamma'(t))$.

Example 56. For our example of $\gamma(t) = (\cos t, \sin t)$ so $\widehat{\gamma}(t) = (t, \gamma(t)) = (t, \cos t, \sin t)$ then the derivative of the first is the curve of tangent vectors $(-\sin t, \cos t)$, while of the second is $(1, -\sin t, \cos t)$, the tangent vector to the helix curve (the graph curve).

Definition 37.11. Summarizing the above, given a Banach space E, a *nonautonomous*, *nonstationary (n.s.)* or *time-dependent vector field* $(V^t)_{\mathbb{R}}$ is a parametrized family of continuously varying vector fields V^t on E, each of which is K-Lipschitz for some fixed K.

A *curve* is a continuous map $\gamma : \mathbb{R} \to E$. Its *graph curve* is $\widehat{\gamma}(t) = (t, \gamma(t))$ with values in $\mathbb{R} \times E$; this has as its image the graph of γ, and so parametrizes that graph.

The *nonstationary (vector) ordinary differential equation* defined from our nonstationary vector field is

$$\gamma'(t) = V^t(\gamma(t)). \tag{147}$$

An *initial condition* for the differential equation is a choice of $(s, \mathbf{v}) \in \widehat{E}$. A *solution* of the equation with this initial condition is $\gamma \in \widehat{\mathcal{C}}$ satisfying (147) with $\gamma(s) = \mathbf{v}$.

This defintion works for Banach spaces. If $E = \mathbb{R}^d$, this vector differential equation is equivalent to a *nonstationary system of differential equations in one dimension* by writing, as for the stationary case in (148), but now simply adding the variable t at the beginning of each line:

$$\begin{cases} y_1'(t) = V_1(t, x_1(t), \ldots, x_d(t)) \\ \vdots \\ y_d'(t) = V_d(t, x_1(t), \ldots, x_d(t)) \end{cases} \tag{148}$$

with initial conditions $y_k(s) = v_k$ for all $1 \leq k \leq d$.

Proposition 37.16. *The curve γ is a solution for (147) iff its graph curve $\widehat{\gamma}$ is a solution for the stationary differential equation*

$$\widehat{\gamma}'(t) = \widehat{V}(\widehat{\gamma}(t))$$

where $\widehat{V}$ is the lift of the nonstationary vector field $(V^t)_{\mathbb{R}}$ to $\widehat{E} = \mathbb{R} \times E$.

Proof. We have

$$\widehat{\gamma}'(t) = (t, \gamma(t))' = (1, \gamma'(t)) = \widehat{V}(\widehat{\gamma}(t))$$

since by definition, this is the stationary vector field $\widehat{V} : \widehat{E} \to \widehat{E}$ defined by $\widehat{V}(t, \mathbf{v}) = (1, V^t(\mathbf{v}))$, where $\mathbf{v} \in E$.

Note that the lift of γ' equals $\widehat{\gamma}'$.

□

Definition 37.12. Given a nonstationary vector field $(V^t)_{\mathbb{R}}$ on E, we also say a curve γ in E is *tangent* to the nonstationary vector field iff it is a solution, that is:

$$\gamma'(t) = V^t(\gamma(t)) = V^t_{\gamma(t)}.$$

From the Proposition, this holds iff the graph curve is tangent to the lift $\widehat{V}$.

A *path* is $\gamma \in \widehat{\mathcal{C}} \equiv \mathcal{C}(\mathbb{R}, \widehat{E}) \subseteq \Pi_{\mathbb{R}} E$, so γ is continuous with value $\gamma(t) \in E_t$ for each t. We give $\widehat{\mathcal{C}}$ the topology of uniform convergence on compact subsets of $\mathbb{R}$.

We shall prove the existence, uniqueness and continuity theorems for nonstationary vector fields in two ways. The simplest is to simply apply our previous results to the lift $\widehat{V}$ as this is both stationary and Lipschitz; it is Lipschitz since that holds in the spatial parameter $\mathbf{v} \in E$ and since the time parameter is the constant 1. We give first however a direct proof using the nonstationary version of the Picard operator, as it is important to see how everything goes through.

For this, given our nonstationary vector field $(V^t)_{\mathbb{R}}$ on E, we define the nonstationary version of the Picard operator $\mathcal{P}_{s,\mathbf{v}} : \widehat{\mathcal{C}} \to \widehat{\mathcal{C}}$ simply to be:

$$(\mathcal{P}_{s,\mathbf{v}}(\gamma))(t) = \mathbf{v} + \int_s^t V^r(\gamma(r))\mathrm{d}r.$$

We take first $s = 0$.

Lemma 37.17. *γ is a fixed point for $\mathcal{P}_{0,\mathbf{v}}$ iff γ is a solution for the vector differential equation $\gamma'(t) = V^t(\gamma(t))$ with initial condition $\gamma(0) = \mathbf{v}$.*

Proof. If γ is a solution, then

$$(\mathcal{P}_{0,\mathbf{v}}(\gamma))(t) = \mathbf{v} + \int_0^t V^r(\gamma(r))\mathrm{d}r = \mathbf{v} + \int_0^t \gamma'(r)\mathrm{d}r = \mathbf{v} + \gamma(t) - \gamma(0) = \gamma(t).$$

Conversely, if $\mathcal{P}_{0,\mathbf{v}}(\gamma) = \gamma$ then $\mathbf{v} + \int_0^t V^r(\gamma(r))\mathrm{d}r = \gamma(t)$ for all t so differentiating, $\gamma'(t) = V^t(\gamma(t))$ and furthermore, $\gamma(0) = \mathbf{v} + \int_0^t V^r(\gamma(r))\mathrm{d}r = \mathbf{v}$. □

Next we connect the Picard operators $\mathcal{P}_{s,\mathbf{v}}$ for different initial times s:

Lemma 37.18. *Given a nonstationary vector field $(V^t)_{\mathbb{R}}$ on E, γ is a fixed point for $\mathcal{P}_{0,\mathbf{v}}$ iff it is a fixed point for $\mathcal{P}_{s,\mathbf{w}}$ for $s \in \mathbb{R}$, where $\mathbf{w} = \gamma(s)$.*

Proof. Let $\mathcal{P}_{0,\mathbf{v}}(\gamma) = \gamma$. Equivalently, $\gamma(t) = \gamma(0) + \int_0^t V^r(\gamma(r))\mathrm{d}r$ for all t.

Now for $\mathbf{w} = \gamma(s)$,

$$(\mathcal{P}_{s,\mathbf{w}}(\gamma))(t) = \mathbf{w} + \int_s^t V^r(\gamma(r))\mathrm{d}r = \gamma(s) + \int_s^t V^r(\gamma(r))\mathrm{d}r =$$

$$\gamma(0) + \int_0^s V^r(\gamma(r))\mathrm{d}r + \int_s^t V^r(\gamma(r))\mathrm{d}r = \gamma(0) + \int_0^t V^r(\gamma(r))\mathrm{d}r = \gamma(t).$$

□

Corollary 37.19. *Given a nonstationary vector field $(V^t)_{\mathbb{R}}$ on E, γ is a fixed point for $\mathcal{P}_{s,\mathbf{v}}$ iff γ is a solution for the vector differential equation: $\gamma'(t) = V^s(\gamma(t))$ with initial condition $\gamma(s) = \mathbf{v}$.* □

Proposition 37.20. *(local fixed point theorem) Given a nonstationary vector field $(V^t)_{\mathbb{R}}$ on E,*
(i) Let $\widehat{\mathcal{C}}_{s,\mathbf{v}} \subseteq \widehat{\mathcal{C}} = \mathcal{C}(\mathbb{R}, E)$ be the subset of all continuous paths γ such that $\gamma(s) = \mathbf{v}$, and let $\widehat{\mathcal{C}}_{s,\mathbf{v},\delta}$ denote the paths in $\widehat{\mathcal{C}}_{s,\mathbf{v}}$ restricted to t in the interval $[s-\delta, s+\delta]$. Then for $\delta < 1/(2K)$, $\mathcal{P}_{s,\mathbf{v}}$ is a contraction mapping on $\widehat{\mathcal{C}}_{s,\mathbf{v},\delta}$, with constant $c < 1$.
(ii) There exists a unique fixed point $\gamma \in \widehat{\mathcal{C}}_{s,\mathbf{v},\delta}$ for each chosen $s, \mathbf{v}$.

Proof. For the stationary case this reduces to our previous proof, by taking $V^t = V$ on E and starting at $t = 0$. We include this nearly identical proof to highlight these notational differences.

Let $\gamma, \widetilde{\gamma} \in \widehat{\mathcal{C}}_{s,\mathbf{v}}$. Then

$$\mathcal{P}_{s,\mathbf{v}}(\gamma)(t) - \mathcal{P}_{s,\mathbf{v}}(\widetilde{\gamma})(t) =$$
$$= \mathbf{v} + \int_s^t V^r(\gamma(r))\mathrm{d}r - \mathbf{v} - \int_s^t V^r(\gamma(r))\mathrm{d}r = \int_s^t V^r(\gamma(r)) - V^r(\widetilde{\gamma}(r))\mathrm{d}r.$$

As above, we write $||f||_J^\infty$ for the sup norm of a function f over the interval $J \equiv [-\delta, \delta]$.

Then for $t \in J = [s-\delta, s+\delta]$,

$$\begin{aligned} &|\mathcal{P}_{s,\mathbf{v}}(\gamma)(t) - \mathcal{P}_{s,\mathbf{v}}(\widetilde{\gamma})(t)| = |\int_s^t V^r(\gamma(r)) - V^r(\widetilde{\gamma}(r))\mathrm{d}r| \\ &\leq \int_s^t ||V^r(\gamma(r)) - V^r(\widetilde{\gamma}(r))||_J^\infty \mathrm{d}r = 2\delta ||V^r(\gamma) - V^r(\widetilde{\gamma})||_J^\infty \\ &\leq 2\delta K ||\gamma - \widetilde{\gamma}||_J^\infty \equiv c||\gamma - \widetilde{\gamma}||_J^\infty \end{aligned} \tag{149}$$

since each V^s is K- Lipschitz.

Now choose δ so small that $c = 2\delta K < 1$. That is, $\delta < 1/(2K)$.

We have shown $\mathcal{P}_{s,\mathbf{v}}$ restricted to the space $\widehat{\mathcal{C}}_{s,\mathbf{v},\delta}$ is a contraction, hence it has a unique fixed point $\gamma : J \to E$.

□

Then we have:

Theorem 37.21. *Given a Lipschitz nonstationary vector field $(V^t)_{\mathbb{R}}$ on E, there exists a unique solution to the nonstationary the vector differential equation: $\gamma'(t) = V^s(\gamma(t))$ with initial condition $\gamma(s) = \mathbf{v}$. The solution curve is in $\mathcal{C}^k$ in the time direction, and is continuous in $(s, \mathbf{v})$, in the topopogy of uniform convergence on compact subsets of $\mathbb{R}$.*

Proof. We have given two proofs: for the first we apply Theorem 37.14 to the lifted (stationary) vector field $\widehat{V}$ on a space $\widehat{E} = \mathbb{R} \times E$ of one higher dimension. For the second, using the nonstationary Picard operator, we have seen that the formulas and proofs are nearly the same. □

Remark 37.8. Thus the stationary case can be used to prove the nonstationary; conversely, of course, the case of a stationary vector field V is a special case of

37.12. **Smoothness in space.** Given a nonstationary Lipschitz vector field $(V^t)_{\mathbb{R}}$ on E, and associated nonstationary differential equation

$$\gamma'(t) = V^t(\gamma(t))$$

with initial condition $\widehat{\mathbf{v}} = (s, \mathbf{v}) \in \widehat{E}$, we know the following so far:
–Existence and uniqueness: a global solution $\gamma(t)$ exists and is unique;

–Continuity in space: This solution $\gamma_{\widehat{v}} \in \widehat{\mathcal{C}}$ varies continuously with respect to this initial condition, in the topology of uniform convergence on compact subsets of $\mathbb{R}$.

–Smoothness in time: The time derivative $\gamma'(t)$ of a solution exists and is continuous. Further, if the vector fields V^t are in $\widehat{\mathcal{C}}^k(E)$ then γ is in $\widehat{\mathcal{C}}^{k+1}(\mathbb{R}, E)$ for all $k \geq 0$.

–Existence of a continuous flow: we know this so far for the stationary case, where we have shown the differential equation defines a continuous flow τ_t, with orbits differentiable in the time direction. We address nonstationary flows below.

Our next job is to prove differentiability in the space parameter.

For this we mostly follow Arnold [Arn12], p.280 ff. as we like this treatment especially: it is a beautiful and clear proof, once some details are filled in and motivated. Filling in these parts leads us to develop important related concepts: the system of equations of variations; the linearization of vector fields and of flows; extensions of flows and vector fields to fiber bundles; parametrized vector fields. A fascinating conclusion is that extensions of vector fields to fiber bundles give an infinitesimal vesion (for continuous time) of skew product transformations in ergodic theory, and equivalently, of random dynamical systems. This result is a connection between random flows and random vector fields, with the system of equations of variations being an especially important special case.

Linearization.

Given a vector field $V : E \to E$, we write $V^* = DV$ for its derivative. Its value $V^*(\mathbf{p})$ is a matrix $V^*(\mathbf{p})$ which best approximates the function V at the point $\mathbf{p}$, which

we call the *linearization* of the vector field at the point. Writing $E^* = L(E, E)$, then this defines a map $V^* : E \to E^*$.

Now E^* is a vector bundle over E. Using the standard charts, this is just the product vector space $E^* = E \times L(E, E)$, but it useful to think of it as a fiber bundle.

Let $\gamma(t)$ be a solution for the differential equation of V on E:

$$\gamma'(t) = V(\gamma(t)).$$

We lift this to the fiber bundle E^*, defining

$$\gamma^*(t) = V^*(\gamma(t))$$

and

$$\widehat{\gamma}^*(t) = (\gamma(t), \gamma^*(t))$$

This is now a curve $\widehat{\gamma}$ in the fiber bundle which projects to γ in the base space E.

For each $\mathbf{p}$, the matrix $A = V^*(\mathbf{p})$ defines a linear flow $\mathbf{w} \mapsto e^{tA}\mathbf{w}$ on the tangent space $T_{\mathbf{p}}$, which is a vector space isomorphic to E. The intuitive idea is that this linear flow on the fiber $T_{\mathbf{p}}$ should in some sense approximate the actual flow of the vector field near that point.

However this is not quite right as the matrix A changes along the path $\gamma(t)$. That is, rather than just this flow on $T_{\mathbf{p}}$ we should consider a flow on the tangent bundle $T(E)$.

Thus, if we think of the path $\gamma(t)$ as giving the time coordinate, then our stationary vector field V on E gives us a *nonstationary* vector field, on E, along the time as given by the path.

That is, $V^t_{\mathbf{p}} = V^*(\gamma_{\mathbf{p}}(t))$ is a nonstationary differential equation, so by Theorem 37.21 we have a unique solution. In fact this is a nonstationary linear differential equation, see Example 59.

Moreover Theorem 37.21 tells us our solutions are continuous in space. However, this is continuity in the tangent space, not in the base space! We still have no information about that form of continuity. What we need is a more global approach, to consider the whole tangent bundle rather than just along an orbit.

To make this precise, we first note that for each $\mathbf{p}$ fixed, $V^*_{\mathbf{p}}$ is a linear map on the tangent space $T_{\mathbf{p}}$. Equivalently, it is a linear vector field on that space, denoted $V^*_{\mathbf{p}}$; the value of this vector field at the point $\mathbf{w} = \mathbf{w}_{\mathbf{p}} \in T_{\mathbf{p}}$ is given by the matrix multiplication: $V^*_{\mathbf{p}}(\mathbf{w}) = V^*_{\mathbf{p}}\mathbf{w}$.

Thus V^* is a *field of (linear) vector fields*, one on each tangent space.

Now, given a curve $\gamma : \mathbb{R} \to E$, we define an associated matrix-valued curve $\gamma^* : \mathbb{R} \to E^*$ whose value is the spatial derivative of the vector field at that point; that is,

$$\gamma^*(t) = V^*(\gamma(t)) = V^*_{\gamma(t)}.$$

We can write this as $\gamma^* = V^* \circ \gamma = V^*_\gamma$.

Now consider a $\mathcal{C}^2$ nonstationary vector field $(V^t)_{\mathbb{R}}$ on $E = \mathbb{R}^d$, the derivative is $(V^{t*})_{\mathbb{R}}$, where for each t, $V^{t*} = (V^t)^*$. Writing $\widehat{E} = E \times \mathbb{R}$, and $E^t = E \times \{t\} \subseteq \widehat{E}$, then V^t is a vector field on E^t.

We take as an inital condition the pair $(\mathbf{p}, \mathbf{w})$ with $\mathbf{p} \in E^{t_0}$ and $\mathbf{w} \in E_{t_0}$. The point $\mathbf{p}$ will move in time t to $\tau_t(\mathbf{p}) = \widetilde{\mathbf{p}}$, while the tangent space $T_{\mathbf{p}}$ is transformed to $T_{\widetilde{\mathbf{p}}}$ by a nonstationary linear flow.

Given now a vector-valued curve $\eta : \mathbb{R} \to E$, then applying to this the matrix-valued curve γ^* gives an equation:

$$\eta'(t) = V^*_{\gamma(t)}(\eta(t))$$

or equivalently

$$\eta'(t) = \gamma^*(t)\eta(t).$$

We claim that this defines a differential equation, but to make this precise we need to go back to the tangent bundle.

When we apply the matrices $\gamma^*(t)$ to the curve $\eta(t)$ with values in E, we have the formal equation written above which makes sense as for t fixed $\gamma^*(t) \in L(E, E)$. However, as noted it is mathematically more correct to think of η as a curve with values not in E but rather in $TE = T(E)$. That is, for fixed t, $\eta(t) \in T_{\mathbf{p}}$ where $\mathbf{p} = \gamma(t)$. But since E is identified with $T_{\mathbf{p}}$, we can do calculations in E.

The initial value of this curve is $\eta(0) = \mathbf{w} \in E$, but actually, $\eta(0) = (\mathbf{p}, \mathbf{w}) = \mathbf{w}_{\mathbf{p}}$ where $\mathbf{p} = \gamma(0)$. This is important because there can be two points $\mathbf{p}, \widetilde{\mathbf{p}}$ with the same $\mathbf{w} \in E$ but which have different solution curves, apparently violating the uniqueness of solutions. To resolve this, we should not lose track of the base point $\gamma(t)$.

Thus, consider the part of the tangent bundle which projects to the curve $\gamma(t)$. We denote this as $T_{\gamma(t)}$. We identify each $T_{\gamma(t)}$ with E and the path $\gamma(t)$ with a topological factor of $\mathbb{R}$ via the map $\gamma(t) \mapsto t$, and denoted $\mathbb{R}/\sim$ where $\sim$ is the equivalence relation on $\mathbb{R}$: $t \sim s \iff \gamma(t) = \gamma(s)$. Then $\mathbb{R}/\sim$ is topologically either $\mathbb{R}$, a circle or a point, depending on whether γ is injective, periodic or constant. $T_{\gamma(t)}$ is thus identified with $(\mathbb{R}/\sim) \times E = \widehat{\mathbb{E}}$. We have a linear vector field $V^*_{\gamma(t)}$ on each $T_{\gamma(t)}$. This corresponds to a nonstationary vector field V^t on $E_t \subseteq \widehat{E}$.

Remark 37.9. This passage from the tangent space $T_{\mathbf{p}}$ to E is possible because TE is a trival bundle, i.e. just the product $E \times E$. Thus we can identify $T_{\mathbf{p}}$ with E and think of η as a curve with values in E. In a general manifold, this identification depends on a *connection*, which gives a way of connecting the tangent spaces along our curve γ.

Definition 37.13. Given a vector bundle F over a manifold M, by a *vector field on the vector bundle* we mean we have a vector field on each fiber $F_{\mathbf{p}}$ which varies continuously in the base point $\mathbf{p}$.

For example, the spatial derivative V^* our vector field V defines a linear vector field on the tangent bundle TE. For curve denoted $\eta : \mathbb{R} \to E$, the equations

$$\eta'(t) = V^*_{\gamma(t)}(\eta(t))$$

or

$$\eta'(t) = \gamma^*(t)\eta(t)$$

define from our from our stationary vector field V on E, a *nonstationary* linear differential equation on the part of the bundle above each chosen curve γ, that is, on $T_{\gamma(t)}$.

We emphasize that this is *not* a differential equation on E itself, nor on $TE \cong E \times E$. Rather we need the definition just given.

Now $T_{\gamma(t)}$ is identified with $(\mathbb{R}/\sim) \times E$. For example, if $\mathbb{R}/\sim = \mathbb{R}$ this is a linear space; however, the only part of the vector space structure used in considering the vector field, and below when defining the Picard operator, is in the fiber.

We summarize this as follows:

Definition 37.14. Given vector field W on a vector bundle F over a manifold M, and given a curve γ in M, with W_γ denoting the vector field over the curve, then

$$\eta'(t) = W_{\gamma(t)}(\eta(t))$$

is the *W-differential equation over γ*. Here $\eta : \mathbb{R} \to F$, with $\eta(t) \in F_{\gamma(t)}$. Thus the values of η are in the fiber over that point.

In the particular case where V is a vector field on M, and γ is an integral curve for V, then we call the pair of equations

$$\begin{cases} \gamma'(t) = V(\gamma(t)) \\ \eta'(t) = W_{\gamma(t)}(\eta(t)) \end{cases}$$

the (V, W) joint differential equation.

A special case is when $W = V^*$, and then we have

$$\begin{cases} \gamma'(t) = V(\gamma(t)) \\ \eta'(t) = V^*_{\gamma(t)}(\eta(t)) \end{cases}$$

or equivalently:

$$\begin{cases} \gamma'(t) = V(\gamma(t)) \\ \eta'(t) = \gamma^*(t)\eta(t) \end{cases}$$

Choosing a joint initial condition $(\mathbf{v}, \mathbf{w}) \in E \times E$, but more properly in $E \times T_{\gamma(0)}(E)$, we call this the *system of equations of variations* for the vector field V.

We can apply the existence and uniqueness theorem proved above (Theorem 37.21) and conclude that fixing $\gamma(t)$, and choosing an initial condition $\mathbf{w} \in T_{\gamma(0)}$, there exists a unique solution curve $\eta(t)$ satisfying the equation. This is continuous in the spatial variable $\mathbf{w}$ in the fiber.

That is, there is a unique solution (γ, η) with this initial condition to the pair of equations.

There is no reason not to have started with a nonstationary vector field V^t, since we ended up with one in any case!

In our notation, the nonstationary equations are

$$\begin{cases} \gamma'(t) = V^t(\gamma(t)) \\ \eta'(t) = V^*_{\gamma(t)}(\eta(t)) \end{cases}$$

or equivalently:

$$\begin{cases}\gamma'(t) = V^t(\gamma(t)) \\ \eta'(t) = \gamma^*(t)\eta(t)\end{cases}$$

The (joint) inital condition is now given as $(t_0.\mathbf{v}, \mathbf{w})$ with $\mathbf{w} \in T_{\mathbf{v}}(E)$ for $\mathbf{v} = \gamma(t_0)$.
So far we have shown:

Theorem 37.22.
(i) Given a K-Lipschitz vector field W on a vector bundle F over a manifold M, and given a curve γ in M, with W_γ denoting the vector field over the curve, then W-differential equation over γ

$$\eta'(t) = W_{\gamma(t)}(\eta(t))$$

with initial condition $(t_0, \mathbf{v}, \mathbf{w})$ has a unique solution $\eta(t)$. If W is $\mathcal{C}^k$ then η is $\mathcal{C}^{k+1}$. The solution is continuous in $\mathbf{w}$, i.e. in the fiber.
(ii) Given a $\mathcal{C}^2$ nonstationary vector field $(V^t)_{\mathbb{R}}$ on $E = \mathbb{R}^d$, and given a joint initial condition $(\mathbf{v}, \mathbf{w}) \in E \times T_{\gamma(t_0)}(E)$, there exists a unique joint solution $(\gamma, \eta)(t)$ to the syatem of variations

$$\begin{cases}\gamma'(t) = V^t(\gamma(t)) \\ \eta'(t) = \gamma^*(t)\eta(t)\end{cases}$$

which is differentiable in t. If the vector field is $\mathcal{C}^k$ for $k \geq 2$ then the solution is $\mathcal{C}^{k+1}$. The solution is continuous in $(\mathbf{v}, \mathbf{w})$ in the topology of uniform convergence on compact subsets of $\mathbb{R}$.

See Arnold's book [Arn12], (2) on p. 279.
In Arnold's notation the (nonstationary) system is written as:

$$\begin{cases}\dot{\mathbf{x}} = \mathbf{v}(\mathbf{x}, t),\ \mathbf{x} \in E \\ \dot{\mathbf{y}} = \mathbf{v}_*(t, \mathbf{x})\mathbf{y},\ \mathbf{y} \in T_{\mathbf{x}}(E).\end{cases}$$

Whether the joint solution is continuous in the joint initial conditions is glossed over by Arnold; it does not follow directly from Theorem 37.21.

$V^*_{\gamma(t)}(\eta(t))\gamma^*(t)\eta(t)$

We consider the Picard operator for η:

$$(\mathcal{P}_{s,\mathbf{w}}(\eta))(t) = \mathbf{w} + \int_s^t V^*_{\gamma(r)}(\eta(r))\mathrm{d}r = \eta(s) + \int_s^t \gamma^*(r)\eta(r)\mathrm{d}r =$$

To describe Arnold's next step let us first consider replacing the vector-valued path $\eta : \mathbb{R} \to E$ by a matrix-valued path $A : \mathbb{R} \to L(E, \widetilde{E})$ where $E = \mathbb{R}^d$ and $\widetilde{E} = \mathbb{R}^m$.
This gives the equation

$$A'(t) = \gamma^*(t)A(t)$$

Fixing a $(d \times d)$ matrix M- in this case $M = \gamma^*(t)$- multiplication on the left by M defines (via the distributive law for matrix multiplicaion) a linear transformation

of the vector space W of $(d \times m)$ matrices, $\mathcal{M}_{d,m}$. and hence a vector field V on W, by simply $V(\mathbf{w}) = A\mathbf{w}$.

The $(d \times d)$ matrix $\gamma^*(t)$ depends on the point: it equals $V^*(\mathbf{p})$ where $\gamma(t) = \mathbf{p}$. So again we have a field of vector fields; we denote this as V^{*m}. Thus for each $\mathbf{p} \in E$, $V^{*m}_{\mathbf{p}}$ is a vector field on the fiber $E^{*m}_{\mathbf{p}} \cong \mathcal{M}_{d,m}$. This is a fiber in the $\mathcal{M}_{d,m}$ matrix bundle, denoted $E^{*m} \cong E \times \mathcal{M}_{d,m}$, since as before the bundle is trivial.

This setup includes the first case above of the tangent bundle, where $M_{d,1} \cong E^t$ is the space of column vectors, so $M_{d,1} : E^t \to E^t$.

Given a path $\gamma \in \mathcal{C}(\mathbb{R}, \widehat{R})$ with each $\gamma(t) \in E^t$, and the associated square matrix-valued path $\gamma^* = D\gamma$, then we consider the vector field V^{*m} on $L(E^d, E^m)$; its value at the point $\mathbf{p} = \gamma(t)$ in E^d is $\gamma^*(t) = V^{*m}(\mathbf{p})$.

Thus for $A \in L(E^d, E^m)$, $V^{*m}(A(t)) = \gamma^*(t)A(t)$, since the value of the vector field is given by matrix multiplication.

For each fixed initial condition $(t_0, \mathbf{v})$, this curve of matrices $A(t) = \gamma^*(t)$, defines a second nonstationary linear differential equation over the curve γ.

$$\eta'(t) = A(t)\eta(t) = \gamma^*(t)\eta(t).$$

The initial condition at time t_0 is $(t_0, \mathbf{w}) = (t_0, \eta(t_0))$. The equation is linear in $\mathbf{w}$, hence the solution is linear in $\mathbf{w}$, and also is a linear operator on the space of curves $\eta : \mathbb{R} \to E$.

Next we let the initial condition $\mathbf{w}$ be a function $\mathbf{w} = \mathbf{w}(\mathbf{p})$, in other words, a vector field $\mathbf{w} : E \to E$. An interesting choice will be the initial condition of the curve γ, i.e. $\mathbf{w}(\mathbf{v}) = \mathbf{v}$. Then the nonlinear equation $\gamma'(t) = V^t(\gamma(t)$ and the linear equation $\eta'(t) = A(t)\eta(t)$ have the same initial value: starting at time t_0, then $\gamma(t_0) = \mathbf{v} = \eta(t_0) \in E$.

As a check, if for example V^t is nonstationary but constant in space, then $\gamma^* = \gamma$ and indeed the solutions are identical: $\gamma(t) = \eta(t)$; $\gamma(t_0) = \mathbf{v} = \eta(t_0)$.

Given paths $\gamma : \mathbb{R} \to E$ and $\eta : \mathbb{R} \to E$, we extend our curves and vector fields to the product space of values, setting $\widehat{\gamma} = (\gamma, \eta) : \mathbb{R} \to E \times E$; the two vector fields V^t and $A(t)$ define a vector field $\widehat{V} = (V, A)$ on $E \times E = \mathbb{R}^{2d}$.

We can unify these into a single differential equation:

$$\widehat{\gamma}'(t) = \widehat{V}(\widehat{\gamma}(t))$$

with initial condition $(t_0, \mathbf{v}, \mathbf{w})$; thus at time t_0 we have $\gamma(t_0) = \mathbf{v}$; $\eta(\mathbf{v}, t_0) = \mathbf{w}$.

In the particular case $A = \gamma^*$, this is

$$\widehat{\gamma}'(t) = (\gamma'(t), \eta(t))) = (V^t(\widehat{\gamma}(t), \gamma^*(t)\eta(t)).$$

.

Arnold writes this as a system:

$$\begin{cases} \gamma'(t) = V^t(\gamma(t)) \\ \eta'(t) = \gamma^*(t)\eta \end{cases}$$

with initial condition at time t_0: $\gamma(t_0) = \mathbf{v}$; $\eta(\mathbf{v}, t_0) = \mathbf{w}$

in other words

$$\begin{cases}\gamma'(t) = V^t(\gamma(t)) \\ \eta'(t) = A(t, \gamma(t))\mathbf{w}\end{cases}$$

Arnold calls this a *system of equations of variations.*

Now Arnold takes one more crucial step: he generalizes the vector $\mathbf{w}$ to a $(\mathrm{d} \times d)$ matrix Z, that is

$$\begin{cases}\gamma'(t) = V^t(\gamma(t)) \\ Z'(t) = A(t, \gamma(t)) \cdot Z(t)\end{cases}$$

With initial condition $\widehat{\gamma}(t_0) = (t_0, \mathbf{v}, Z)$.

But why is this a DE? What is the space, and the vector field? The observation we need here is that multiplication on the left $Z \mapsto A \cdot Z$ is linear on $\mathcal{M}_d \cong L(E, E)$ and so defines a vector field $\widehat{V}^*$ on $E^2 \cong \mathbb{R}^{2d}$.

This means that defining the vector field along the path, $A^t(Z(\gamma(t))) = v$. we could rewrite the above system as:

$$\begin{cases}\gamma'(t) = V^t(\gamma(t)) \\ Z'(t) = A^t(Z(\gamma(t)))\end{cases}$$

Paths are matrix-valued, and we define the Picard operator for this vector field exactly as before:

Let $\widehat{\mathcal{C}}^* = \mathcal{C}(\mathbb{R}, E^*)$. We can define a on $\widehat{\mathcal{C}}^*$ as follows.

$$(\mathcal{P}^*_{s,\mathbf{v}}(\gamma^*))(t) = M + \int_s^t (\widehat{V}^r)^*(\gamma^*(r))\mathrm{d}r = M + \int_s^t (V^r)^*(\gamma^*(r))\mathrm{d}r,$$

since application of the vector field is given by multiplication by the matrix. This is Arnold's equation.

Examples of linearization.

For a classical example of this we consider the equation for the pendulum, a second-order nonlinear equation in one dimension. Its linearization at position 0 is the harmonic oscillator, with equation

$$x'' = -x,$$

where $x = x(t)$ is position, x' is velocity or momentum, and x'' is acceleration or force.

The derivation of the harmonic oscillator equation comes from Hooke's Law in physics, which states that the restorative force of a spring is proportional to the distance displaced from the rest position $x = 0$. Since by Newton's law $F = ma$, then taking $m = 1$ yields the above equation

$$x'' = -x,$$

with the negative sign due to the force being "restorative" i.e. pushing the mass back torards the rest point.

Introducing a second variable y by setting $y = x'$ then $y' = x''$ and we have the pair of equations in (x, y). Since momentum is mv where $v = x'$ is velocity, then x is position (say horizontal) while y is velocity=momentum. Thus the coordinates are $(x, y) = (x(t), y(t))$ in position-momentum *phase space*.

$$\begin{cases} x' = y \\ y' = -x \end{cases}$$

hence for $\mathbf{w} = (x, y)$, this has matrix form $\mathbf{w}' = A\mathbf{w}$ where $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. (See Exercise 37.4 where we switched x and y, as we are thinking of it as a second-order equation $y''(t) + y(t) = 0$. In that case, we get the transposed matrix, which yields a counterclockwise rotation flow, since for $B = tA$, and B^t the transpose, then $e^{(B^t)} = (e^B)^t$. Thus the vector solutions are circular curves which project onto $\sin, \cos$ in one dimension, solutions for the position and its derivative, the velocity=momentum.

The pendulum has the equation

$$x'' = -\sin x$$

where now x is *angular* position and $y = x'$ is angular momentum. This equation comes from the fact that the force of gravity is constant vertically, but its component in the angular direction is (radius times sin(angle)).

So we have

$$\begin{cases} x' = y \\ y' = -\sin x \end{cases}$$

This is the system for the vector field V on $\mathbb{R}^2$ with

$$V_1(x, y) = y, V_2(x, y) = -\sin x$$

then $DV = V^* = \begin{bmatrix} (V_1)_x & (V_1)_y \\ V_2)_x & (V_2)_y \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\cos x & 0 \end{bmatrix}$ so the linearization of V at $(0, 0)$ is $V^*(0, 0) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, which generates via $\tau_t = e^{tA}$ the clockwise linear rotation flow. See Fig. 110. The solution is again a point $(x, y) = (x(t), y(t))$ in phase space, but now it is for angular position-momentum. To calculate the Euclidean position from the angle we would take $\sin(\theta)$ so for instance if $\theta(t) = x(t) = \sin(t)$ then this would be $\sin(\sin(t))$. No wonder in textbooks this is always just expressed in the angular form!

The time-dependent case as stationary in one more dimension.

??? REDO.

Now in let $\gamma(s, t) = \gamma^s(t)$, Thus $\gamma^s(0) \in E_s$ and $\gamma^s(t) \in E_{s+t}$; this is a path in the time-parameter t, starting at time s. We write and define $\widehat{\mathcal{C}} = \mathcal{C}(\mathbb{R} \times \mathbb{R}, E)$. So given $\gamma \in \widehat{\mathcal{C}}$, we define conversely $\gamma^s(t) = \gamma(s, t)$.

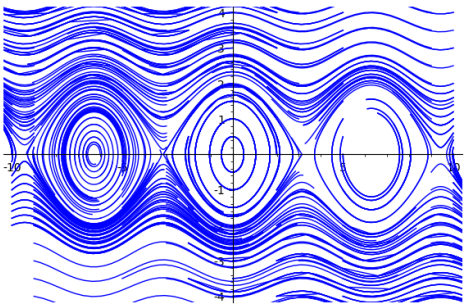


FIGURE 110. A nonlinear rotation: the pendulum, in phase space with angular (position, momentum) coordinates (x, y); the circles flow clockwise, just as for the linearization at $\mathbf{0}$ which is the rotation flow. The upper curves flow to the right , the lower curves to the left; these correspond to the pendulum no longer oscillating but instead going around and around in one direction when it has high enough angular momentum.

Remark 37.10. The *lift* of the nonstationary field $(V^s)_{\mathbb{R}}$ is a vector field V on $\widehat{E}$, i.e. $V : \widehat{E} \to \widehat{E}$, with

$$V(s, \mathbf{v}) = (1, V^s(\mathbf{v})).$$

Thus the t-coordinate of the vector field V is the constant 1. Note that V is a continuous vector field and moreover is K-Lipschitz on $\widehat{E}$. The lift projects to $(V^s)_{\mathbb{R}}$ by $V^s(\mathbf{v}) = V(s, \mathbf{v})$.

This allows one to treat a nonstationary vector field as a vector field in the usual sense (i.e. a stationary vector field) in one higher dimension, of a very special type (constantly 1 in the time coordinate).

Hence, to prove the existence and uniqueness theorem we can simply apply the stationary result to this case. However conceptually the two are quite different. Therefore will be useful to note that the Picard iteration does go through in any case. In what follows we take both approaches.

In the special case where $(V^s)_{\mathbb{R}}$ is constant, the family defines a single vector field on E by the identification of R^n_s with E. In this case we call $(V^s)_{\mathbb{R}}$ a *stationary nonstationary* vector field. It is *not* the same as a single stationary vector field on E, both formally and practically, as one can consider other mathematical objects (functions, measures, other vector fields) together with the family which however *do* vary with time; this can occur naturally.

37.13. **Flow extensions and extensions of vector fields; parametrized vector fields and vector fields on fiber bundles.**

Definition 37.15. (Curves and vector fields on manifolds)

When M is a Banach manifold (Defn???) with tangent space TM, and projection map $\pi : TM \to M$ with $\pi(\mathbf{v}_p) = \mathbf{p}$ for $\mathbf{v_p} \in TM_p$, where each TM_p is a linear spave isomorphic to E, then this tangent bundle is *locally trivial*, i.e. there is some smooth chart Φ for TM such that TM is locally a product $\mathcal{U} \times E$ for some $\mathcal{U}$ an open subset

of E; that is, $\Phi : \mathcal{U} \times E \to TM$ with $\Phi : \mathbf{p} \times E \to TM_{\mathbf{p}}$ linear. Then given a curve $\gamma : \mathbb{R} \to E$, the tangent vector $\gamma'(t)$ is an element of $TM_{\gamma(t)}$.

Thus we have a commutative diagram:

$$\begin{array}{ccc} \mathbb{R} & \xrightarrow{\gamma'} & TM \\ \Big\downarrow \mathrm{id} & & \Big\downarrow \pi \\ \mathbb{R} & \xrightarrow{\gamma} & M \end{array}$$

A vector field on M is a function $V : M \to TM$, with $V(\mathbf{p}) \in TM_{\mathbf{p}}$.

Example 57. We consider $M = E$; then defining as above $\widehat{E} \equiv \mathbb{R} \times E$, denoting the two natural projections by $\widehat{\pi} : (t, \mathbf{v}) \mapsto \mathbf{v}$ with $\widehat{\pi} : \widehat{E} \to E$, and $\pi_{\mathbb{R}} : \widehat{E} \to \mathbb{R}$ with $\pi_{\mathbb{R}} : (t, \mathbf{v}) \mapsto t$. We set $E_t = \pi_{\mathbb{R}}^{-1}(t)$, an indexed copy of E. Thus $E_t = \{t\} \times E \subseteq \mathbb{R} \times E = \widehat{E}$. Now $TM = E \times E = \widehat{E}$. For $\gamma : \mathbb{R} \to E$, $\gamma'(t) = (t, \gamma'(t)) = \widehat{\gamma}'(t) \in \widehat{E}$, as in the second definition above! Thus the definition for manifolds extends this.

We have seen that the infinitesimal version of a flow is a vector field, in both the stationary and nonstationary settings. The time derivative takes you from the flow to the vector field, and solving the ODE defined by the vector field takes us back again. Now we consider extensions to fiber bundles.

Examples:

(1) Time-varyng vector field:

-base space is real line; fiber is E_t; bundle is $\widehat{E} = \mathbb{R} \times E$. Time-varying vector field V^t on $E_t = \{t\} \times E \subseteq \widehat{E}$. Field on $\mathbb{R}$ is constant 1: $W(t) = 1$. Curve in base is $\eta(t) = t$. W lifts to $\widehat{W}$ on $\widehat{E}$, $\widehat{W}(t, \mathbf{v}) = (1, V^t(\mathbf{v}))$. Solution curve: $\widehat{\gamma}(t) = (\eta(t), \gamma(t))$, $\gamma(0) = \mathbf{v}$.

γ is solution to $\gamma'(t) = V^t(\gamma(t))$. $\widehat{\gamma}$ is solution to lifted equation: $\widehat{\gamma}'(t) = \widehat{W}(\widehat{\gamma}(t))$.

(2)"Stack of records":

-base space is real line; fiber is E_t; bundle is $\widehat{E} = \mathbb{R} \times E$. Time-varying vector field V^t on $E_t = \{t\} \times E \subseteq \widehat{E}$. Field on $\mathbb{R}$ is constant 0: $W(t) = 0$. Curve in base is $\eta_t(t) = t$, fixed point, for each t. W lifts to $\widehat{W}$ on $\widehat{E}$, $\widehat{W}(t, \mathbf{v}) = (t, V^t(\mathbf{v}))$. Only movement is is each fiber independently. Solution curve: $\widehat{\gamma}(t) = (t, \gamma_t(t))$, $\gamma(0) = \mathbf{v}$.

γ is solution to $\gamma'(t) = V^t(\gamma(t))$. $\widehat{\gamma}$ is solution to lifted equation: $\widehat{\gamma}'(t) = \widehat{W}(\widehat{\gamma}(t))$.

Fiber $= E$,

???

37.14. **Nonstationary flows.**

Definition 37.16. Let $(\tau_t)_{t \in \mathbb{R}}$ be a family of continuously varying homeomorphisms of a topological space X. We call this a *pre-flow*. It is a *flow* iff it satisfies

(i) τ_0 is the identity map;

(ii) the *flow condition*:

$$\tau_{t+s} = \tau_s \circ \tau_t.$$

Taken together, these imply that each map τ_t is a bijection, as it has inverse τ_{-t}.

Given a pre-flow $(\tau_t)_{t \in \mathbb{R}}$ on a space X we set $X_t = X \times \{t\}$ and write $x_t = (x, t) \in X_t$. We define $\widetilde{\tau}_t : X_s \to X_{s+t}$ by $\widetilde{\tau}_t(x, s) = (\tau_t(x), s + t)$. We call $\widetilde{\tau}_t$ a *nonstationary flow*

or a *flow family*. For example the *time-one map* is $\widetilde{\tau}_1$, and this maps X_s to X_{s+1} to $X_{s+2}\ldots$, iterating the map (but not flow) τ_1.

We set $\widehat{X} = X \times \mathbb{R}$ by $\widehat{\tau}_t(x,s) = (\tau_t(x), t+s)$ $\widehat{\tau}_t : \widehat{X} \to \widehat{X}$ by $\widehat{\tau}_t(x,s) = (\tau_t(x), t+s)$. We call $(\widehat{X}, \widehat{\tau}_t)$ a *total flow*.

Lemma 37.23. *The following are equivalent: a pre-flow, a nonstationary flow. The total flow is a flow on the space $\widehat{X}$ with the special property that it preserves fibers. The following diagram is commutative:*

37.15. **Nonstationary dynamical systems.**

Remark 37.11. This conceptual difference between nonstationary and stationary dynamics becomes more striking in the situation where E is replaced by a compact manifold M, as now the covering space $\widehat{E}$ is non-compact.

From the dynamical systems viewpoint, a nonstationary discrete-time system, i.e. a sequence of maps along a sequence of spaces, can always be lifted to a single map on a larger space, the *coproduct* i.e. the disjoint or indexed union. See [AF05]; we call this the *total map* in that setting. Now this defines an intrinsically wandering dynamical system, where one might expect none of the standard notions of dynamics (stable manifolds, hyperbolicity, Markov partitions, transverse dynamics...) to make sense. Nonetheless, certain ideas go through, while on the other hand some striking new phenomena can occur as well. See [AF01], [Fis09], [dJMA17].[Ace18b], [Ace18a],[MR20], [Sil17].

A good source of examples is the study of *random dynamical systems*; choice of a single ω in the relevant probability space then determines a nonstationary system, either for discrete or continuous time.

Our point of view is that we need also to study nonstationary systems in their own right, not just for those associated to some random system. For differential equations this certainly has been the case historically, where many of the DEs important in applications, also in one dimension, are nonautonomous equations.

Consider for an example the curve $\gamma : \mathbb{R} \to E$, with $\gamma(t) = (\cos t, \sin t)$. Then the *image* of the curve is the unit circle, while the *graph* of the curve is by definition $\{(t, \gamma(t)) : t \in \mathbb{R}\} \subseteq \mathbb{R} \times \mathbb{R}^2$. We can identify this with a subset of $\mathbb{R}^3$, that is, $\{(t, \cos t, \sin t) : t \in \mathbb{R}\}$, which is a helix (a circular spiral). This is in turn a new curve, now in $\mathbb{R}^3$: $\widehat{\gamma}(t) = (t, \gamma(t)) = (t, \cos t, \sin t)$. Thus the graph of γ becomes the image of the related curve $\widehat{\gamma}$. Note that the tangent vector is now $\widehat{\gamma}'(t) = (1, \gamma'(t))$.

Now γ is a solution of the autonomous, linear DE $\mathbf{x}' = A\mathbf{x}$ with initial condition $\mathbf{x}_0 = (1,0)$, since as we saw in Proposition 35.15, for $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, then

$$e^{tA} = R_t = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix},$$

whence $\mathbf{x}(t) = e^{tA}\mathbf{x}_0$. The curve $\widehat{\gamma}$ is a solution of the nonstationary DE with parametrized but constant vector field $V^s = A$ for all $s \in \mathbb{R}$.

Thus we are distinguishing between n.s but constant and truly nonstationary vector fields.

This perspective unifies the stationary and nonstationary cases. A further unification comes by adding a dimension; this will gives us a flow even in the nonstationary case.

Given a time-dependent vector field $V^s = V(s, \cdot)$, i.e. a one- parameter family of K-Lipschitz vector fields, continuous in t, thus $V : \mathbb{R} \times E \to E$ be continuous, and K-Lipschitz in the E-coordinate. We define a stationary vector field $\widehat{V}$ on $\mathbb{R}^{n+1}$ by $\widehat{V}(\mathbf{v}, s) = \widehat{\mathcal{V}^s(\mathbf{v})}$ where $\widehat{(x_1, x_2, \dots, x_n)} = (x_1, x_2, \dots, x_n, 1)$.

$V : \mathbb{R} \times E \to E$ be continuous, and K-Lipschitz in the E-coordinate. We define $E_s = \{s\} \times E \subseteq \mathbb{R} \times E$. We write $V^s = V(s, \cdot)$; this is a one- parameter family of K-Lipschitz vector fields (with fixed K), changing continuously in time. Thus V^s is a vector field on E_s. We call $V^s = V(s, \cdot)$ a *nonautonomous*, *nonstationary (n.s.)* or *time-dependent* vector field.

Now consider the flow on $\mathbb{R} \times E$, defined by $\widehat{\tau}_t : (\mathbf{x}, s) \mapsto (\tau_t(\mathbf{x}, s + t)$. Thus for the above linear example,

$$\widehat{\tau}_t(\mathbf{x}, s) = (e^{tA}\mathbf{x}, s + t)$$

which in matrix form is

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ t \end{bmatrix}$$

This has derivative at $t = 0$: $\widehat{A} + (0, 0, 1)$ where $\widehat{A}$ is the (3×3) matrix

$$\widehat{A} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Let $(X_i, d_i)_{i \in I}$ be metric spaces for $I = \mathbb{N}$ or $\mathbb{Z}$. Let $T_i : X_i \to X_{i+1}$ be continuous maps. This is a *nonstationary dynamical system* or *mapping family*, as introduced in [AF05]. The *total map* of the mapping family is defined as follows: we set $\widehat{X} = \coprod_I X_i = \cup_{i \in I}(X_i, i)$ (this *coproduct* is the same thing as the disjoint union or indexed union) and define $\widehat{T} : \widehat{X} \to \widehat{X}$ by $(x_i, i) \mapsto (T_i(x_i), i + 1)$. This is the *total map* associated to the mapping family.

We think of the mapping family as a sequence of maps along a sequence of spaces, while the total map is an actual map (that is, a map of a space to itself) which contains all the same information.

It is a *two-sided* or *biinfinite* mapping family iff $I = \mathbb{Z}$, otherwise a *one-sided* family. If the family is bilateral and each map is a homeomorphism, it is an *invertible* family, and this is true iff the total map is a homeomorphism.

We note that in the definition of mapping family, the metric is allowed to change with time. This becomes important in examples. See §42.3. Also note that if the individual spaces X_i are compact, the total space nevertheless will be noncompact. The dynamics of the total map is wandering; nevertheless, one can make sense of hyperbolicity, stable manifolds, and transverse dynamics.

The continuous-time version of mapping family is a nonstationary flow, which we next define. For ODEs it is closely linked to the classical idea of a time-varying (or nonstationary, or nonautonomous) vector field, as we shall explain.

....

A student asked me if this might be true for all differential equations. This is a great question!

So far we can verify it for:

–all one-dimensional linear equations (including the nonautonomous case)

–homogeneous linear autonomous vector DEs.

Equations and flows in one higher dimension

Convergence to flows:

–Picard iteration and nonstationary flows

Picard iteration for flows and vector fields: Flow contraction

Defining a nonstationary flow $\tau_t : E \to E$ by $\tau_t(\mathbf{v}) = \gamma_{\mathbf{v}}(t)$

$$\tau_t^{(1)}(\mathbf{v}) = (\gamma_1)_{\mathbf{v}}(t) = (\mathcal{P}_{\mathbf{v}}(\gamma))(t) = \mathbf{v} + \int_0^t V(\gamma(s))\, \mathrm{d}s = \mathbf{v} + \int_0^t V(\tau_s(\mathbf{v}))\, \mathrm{d}s =$$

.....

Example 58. Consider the ....

37.16. **Picard iteration: further examples.**

Example: Picard iteration for the nonstationary one-dimensional case

In (138) above we considered the homogeneous linear (nonautonomous) one-dimensional equation $x'(t) = a(t)x(t)$. We found the solution to be $x(t) = e^{\widehat{a}(t)}x_0$ where $\widehat{a}(t) = \int_0^t a(s)\, d\, s$, thus $\widehat{a}'(t) = a(t)$, $\widehat{a}(0) = 0$ and $x_0 = x(0)$ is the initial condition.

.

We next see if we can derive this by Picard iteration. We have $x_0(t) \equiv x_0$, $x_1(t) = (\mathcal{P}_{x_0}(x_0))(t) = \mathbf{x}_0 + \int_0^t V(x_0(s))\, \mathrm{d}s = \mathbf{x}_0 + \int_0^t a(s)\mathbf{x}_0\, \mathrm{d}s = \mathbf{x}_0 + (\int_0^t a(s)\, \mathrm{d}s)\mathbf{x}_0$....

Example 59. **Solution for nonstationary linear vector fields; Picard iteration.**

Before we consider the general nonstationary vector case, we take the instructive special case of a parameterized family V^t of linear vector fields, that is, given by a matrix family $A(t)$, $t \in \mathbb{R}$.

This is exactly analogous to the homogeneous linear (nonautonomous) case on $\mathbb{R}$ just treated: the one-dimensional homogeneous linear equation $x' = a(x)x$.

In E, this corresponds to the nonstationary equation

$$\mathbf{x}' = A(t)\mathbf{x}(t)$$

with initial condition $\mathbf{x}(0) = \mathbf{x}_0$, where for each t, $A(t)$ is an $(n \times n)$ matrix. We immediately check that the same formula gives solution curves:

$$\mathbf{x}(t) = e^{\widehat{A}(t)}\mathbf{x}_0$$

where $\widehat{A}'(t) = A(t)$. (Since $\widehat{A}(t)$ is a curve in $\mathbb{R}^{n^2}$, the tangent vector makes sense; it is just the derivative of the matrix coordinates).

We note that adding a constant matrix does not change the derivative $\widehat{A}'(t)$. Therefore we require here without loss of generality that $\widehat{A}(0) = 0 \cdot I$; if not we can achieve this by subtracting the constant matrix $\widehat{A}(0)$. Then the initial condition is $\mathbf{x}(0) = \mathbf{x}_0$.

Equivalently, $\widehat{A}(t) = \int_0^t A(s)\ \mathrm{d}s$ and $\widehat{A}(0) = 0 \cdot I$.

For example, fixing $A \in \mathcal{M}_n(\mathbb{R})$, and given a continuous function $f : \mathbb{R} \to \mathbb{R}$, we consider the nonstationary vector field $A(t) = f(t)A$. Then $\widehat{A}(t) = \int_0^t A(s)\ \mathrm{d}s = \int_0^t f(s)A\ \mathrm{d}s = \int_0^t f(s)\ \mathrm{d}s \cdot A = F(t)A$ where $F' = f$, $F(0) = 0$, and the solution of the nonstationary DE $\mathbf{x}' = f(t)A\mathbf{x}(t)$ with initial condition $\mathbf{x}(0) = \mathbf{x}_0$, is

$$\mathbf{x}(t) = e^{\widehat{A}(t)}\mathbf{x}_0$$

where $\widehat{A}(t) = F(t)A$.

We next see if we can derive this by Picard iteration.

Example 60. We consider a specific example: a periodically varying harmonic oscillator:

$$f(t) = \sin(t) \text{ and } y'' = -f(t)y.$$

Then $F(t) = -\cos(t)$ and the vector DE is

$$\mathbf{x}' = A(t)\mathbf{x} = \sin(t) \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \mathbf{x}$$

with nonstationary flow solution therefore: $e^{\widehat{A}(t)} = -\cos(t)R_t$ and solution curves

$$-(\cos t)R_t\mathbf{x}_0 = -(\cos t) \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

leading to the one-dimensional solutions

$$y(t) = (-\cos t)(a \sin t + b \cos t)$$

with initial condition $y(0) = b, y'(0) = a$. Physically, this could model, in music, a periodically varying volume oscillation.

37.17. **Volume change: determinant, divergence and trace.**

Theorem 37.24. *Given a $\mathcal{C}^2$ nonstationary vector field $(V^t)_{\mathbb{R}}$ on $E = \mathbb{R}^d$, the derivative is $(V^{t*})_{\mathbb{R}}$, where for each t, $V^{t*} = (V^t)^*$. Let $(\tau_t\}_{t\in\mathbb{R}}$ be the corresponding nonstationary flow.*

Then $det\tau_t(\mathbf{x}) = \exp(tr(V^{t*}(\mathbf{x}))) = \exp(divV(\mathbf{x}))$.

In particular, $det\tau_t(\mathbf{x}) = 1$ iff $tr(V^{t*}(\mathbf{x})) = 0$, so τ_t preserves volume iff $(V^t)_{\mathbb{R}}$ has divergence 0.

?? volume form and defn of divergence...

37.18. **Nongeodesic curves in group and nonstationary flows.** ex: random walk in group with trace 0 Lie algebra= det 1: SL(n,C)

–nonhomogeneous "group": stationary= geodesic in M, nonst= curve in M. Acting on....

–time-varying M....geodesic or curve....

.....

The flow generated by a Hamiltonian vector field Hf preserves the symplectic form ... p.79 ff "TaylorMichaelGREATnotes"

38. Appendix: A simple proof of the Birkhoff ergodic theorem

We present a simple proof of the almost-sure (Birkhoff) Ergodic Theorem. Our treatment is original only in some aspects of the presentation; we borrow from Keane [Kea91] Keller [Kel98] and Kaznelson-Weiss [KW82]. There have been a number of "simple" proofs recently, the starting points being a proof of Ornstein-Weiss for amenable groups [OW83], which inspired Shields' treatments [Shi87] [Shi96] as well as Kamae's nonstandard analysis proof [Kam82]. The insights there led in turn to the (standard analysis) proofs of Keane and of Kalznelson-Weiss. Keane's approach most recently sparked an article with Petersen [KP06]; see also [KK97].

We find most readable Keane's proof [Kea91]; this was written for an introductory survey course for graduate students, and is incomplete as it only deals with the case of bounded functions. The remaining (nontrivial) details have been nicely completed by Keller, but for the more general case of $\mathbb{Z}^d$ actions; this adds further complications, so we have felt it worthwhile to present here the proof for the simplest case, that of a single map.

Theorem 38.1. *Let T be a measure-preserving transformation of a probability measure space $(X, \mathcal{A}, \mu)$. Then for any $f \in L^1$, the limit*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) = f^*(x)$$

exists almost surely; the function f^* is invariant and integrable, and for any measurable $E \subseteq X$ which is invariant i.e. $T^{-1}(E) = E$, $\int_E f d\mu = \int_E f^* d\mu$.

Proof. Writing $f = f^+ - f^-$, it will be sufficient to prove the theorem for $f \geq 0$. The hard part is proving the limit exists. We write the partial sums of f along an orbit as $S_n = S_n(f)$ with $S_0 = 0$ and $S_n(x) = \sum_{i=0}^{n-1} f(T^i(x))$ for $n \geq 1$. We define the averaged sums $A_n(x) = S_n(x)/n$ and write

$$\overline{A} = \lim \sup_{n\to+\infty} A_n$$

and

$$\underline{A} = \lim \inf_{n\to+\infty} A_n.$$

These are invariant extended-real valued functions; we claim it will be enough to show:

$$\int \underline{A}\,\mathrm{d}\mu \geq \int f\,\mathrm{d}\mu \geq \int \overline{A}\,\mathrm{d}\mu. \tag{150}$$

For then since $\overline{A} \geq \underline{A}$, (150) implies that

$$0 \geq \int \overline{A} - \underline{A}\,\mathrm{d}\mu \geq 0.$$

Hence $\underline{A} = \overline{A}$ $\mu-$a.s, and so the limit f^* exisits. Note that since $\underline{A}$ and $\overline{A}$ are invariant, so is f^*.

We further claim that in fact it will be enough to show half of this: for any $f \geq 0$ in L^1,

$$\int f\,\mathrm{d}\mu \geq \int \overline{A}\,\mathrm{d}\mu. \tag{151}$$

For this argument, borrowed from [Kel98], we shall indicate the dependence of $\overline{A}$ on a function g as $\overline{A}(g)$.

First assume that $0 \leq f(x) \leq M$ for all x. Considering the function -f, we have $\overline{A}(-f) = -\underline{A}(f)$. We now apply (151) with a different function, $(M - f)$, which is by the assumption ≥ 0 and in L^1; this gives

$$\int (M - f)\,\mathrm{d}\mu \geq \int \overline{A}(M - f)\,\mathrm{d}\mu. \tag{152}$$

On the left-hand side we have $\int (M - f)\,\mathrm{d}\mu = M - \int f\,\mathrm{d}\mu$ and on the right-hand side, $\int \overline{A}(M - f)\,\mathrm{d}\mu = M + \int \overline{A}(-f)\,\mathrm{d}\mu = M + \int -\underline{A}(f)\,\mathrm{d}\mu = M + - \int \underline{A}(f)\,\mathrm{d}\mu$ so therefore

$$-\int f\,\mathrm{d}\mu \geq -\int \underline{A}(f)\,\mathrm{d}\mu$$

and hence

$$\int \underline{A}(f)\,\mathrm{d}\mu \geq \int f\,\mathrm{d}\mu.$$

This takes care of the bounded case. Next, not assuming $f \leq M$, and writing

$$f \wedge M(x) = \min\{f(x), M\}$$

we have for $f \geq 0$,

$$\int \underline{A}(f)\,\mathrm{d}\mu \geq \int \underline{A}(f \wedge M)\,\mathrm{d}\mu \geq \int (f \wedge M)\,\mathrm{d}\mu$$

the last of which, by the Monotone Convergence Theorem, converges to $\int f\,\mathrm{d}\mu$. Thus we have shown that knowing the second inequality of (150), for all functions ≥ 0 and integrable, leads to the first inequality as well.

Thus it remains

To show: for $f \geq 0$ in L^1, then

$$\int f\,\mathrm{d}\mu \geq \int \overline{A}\,\mathrm{d}\mu.$$

As in [Kea91], we precede by cases of increasing difficulty, for pedagogical purposes.

CASE 1: $f \in L^\infty$. Fix $\varepsilon > 0$ and define:

$$\tau(x) = \inf\{n \geq 1 : \ A_n(x) \geq (1-\varepsilon)\overline{A}\}.$$

This is a **stopping time** in the language of probability theory. **Note:** τ is finite a.s. since $\overline{A}(x) \leq ||f||^\infty < \infty$.
CASE 1a: $\tau(x) \leq \tau_0$ for some constant τ_0.
CLAIM: for all $n \geq 1$, for a.e. x, we have:

$$S_n(x) \geq (1-\varepsilon)(n-\tau_0)\overline{A}(x). \tag{153}$$

The idea will be to divide the orbit of x into segments of length given by the stopping time, and where we therefore have a good estimate. Thus we define $x_0 = x$, and inductively

$$x_{k+1} = T^{\tau(x_k)}(x_k).$$

First we note that for $l = \tau(x)$,

$$S_l(x) \geq (1-\varepsilon)(l)\overline{A}(x) \tag{154}$$

by the definition of the stopping time τ, and that the same is true for any time l of the form $l = \tau(x_0) + \tau(x_1) + \cdots + \tau(x_k)$. For a general time $n \geq 0$, let x_k be the last point in the orbit such that l as just defined is $\leq n$, and hence $l + \tau(x_{k+1}) > n$. We have the good estimate (154) up to time l; after this we only know the values are ≥ 0, but fortunately by hypothesis, $\tau(x_{k+1}) < \tau_0$. Hence simply neglecting the last $n - l$ terms gives us the estimate (153), since $l > n - \tau_0$.

Next we use this to prove (151) for this case. From (153), we have

$$\frac{S_n(x)}{n} \geq (1-\varepsilon)\frac{(n-\tau_0)}{n}\overline{A}(x) \tag{155}$$

and so, using for the first equality the invariance of the measure,

$$\int f \,\mathrm{d}\mu = \int \frac{S_n}{n} \,\mathrm{d}\mu \geq (1-\varepsilon)\frac{(n-\tau_0)}{n} \int \overline{A} \,\mathrm{d}\mu$$

and since this is true for all n,

$$\int f \,\mathrm{d}\mu \geq (1-\varepsilon) \int \overline{A} \,\mathrm{d}\mu,$$

finishing the proof for Case (1a).
CASE 1b: $\tau(x)$ is now unbounded (but is a.s. finite as observed above).

We choose $\tau_0 > 0$ and define G (the good set) to be the set of all x where $\tau(x) < \tau_0$ and the bad set B to be its complement. Note that $\mu(B) \to 0$ as $\tau_0 \to +\infty$. Then we set

$$\tilde{f}(x) = \begin{cases} f(x) & \text{for } x \in G \\ \sup f & \text{for } x \in B \end{cases}$$

We define $\widetilde{S}_n(x) = S_n\tilde{f}(x)$ and
CLAIM: for all $n \geq 1$, for a.e. x:

$$\widetilde{S}_n(x) \geq (1-\varepsilon)(n-\tau_0)\overline{A}(x). \tag{156}$$

Proof of Claim: We have always

$$\widetilde{S}_{\tau(x)}(x) \geq S_{\tau(x)}(x) \geq (1-\varepsilon)(\tau(x))\overline{A}(x).$$

So if x_k (defined as before) is in G, the same estimate works as for part (1a), and we have:

$$\widetilde{S}_n \geq S_n \geq (1-\varepsilon)(n-\tau_0)\overline{A}.$$

If $x_k \notin G$, the value of $\tilde{f}$ on the point x_k has been boosted to $\sup f \equiv c$. If $T(x_k) \notin G$, the value of $\tilde{f}$ here has been boosted as well. However it is possible that some first point after x_k in the orbit, say w, is again in G (and so the value has not been boosted). But then we simply use the previous estimate along this next piece of orbit, since w is in G. Now along the orbit segment immediately after x_k until w, of length $\tilde{l}$, the partial sum is $c \cdot (\tilde{l}) \geq \overline{A} \cdot (\tilde{l}) \geq (1-\varepsilon)\overline{A} \cdot (\tilde{l})$, and the part following w is estimated as before. After $\tau(w)$ we may again enter B and boost the values, continuing in this way until we reach n. In summary, we have now the estimate $(1-\varepsilon)\overline{A} \times$ (length) along the whole orbit segment until time n, and so

$$\widetilde{S}_n \geq (1-\varepsilon)\overline{A} \cdot n.$$

Integrating as before,

$$\int_X \tilde{f}\,\mathrm{d}\mu \geq (1-\varepsilon)\int_X \overline{A}\,\mathrm{d}\mu.$$

Now,

$$\int_X \tilde{f}\,\mathrm{d}\mu = \int_G \tilde{f}\,\mathrm{d}\mu + \int_B \tilde{f}\,\mathrm{d}\mu = \int_G \tilde{f}\,\mathrm{d}\mu + c\mu(B)$$

and

$$\begin{aligned}\int_X f\,\mathrm{d}\mu &= \int_G f\,\mathrm{d}\mu + \int_B f\,\mathrm{d}\mu \geq \int_G f\,\mathrm{d}\mu = \int_G \tilde{f}\,\mathrm{d}\mu \\ &= \int_X \tilde{f}\,\mathrm{d}\mu - c\mu(B) \geq (1-\varepsilon)\int_X \overline{A}\,\mathrm{d}\mu - c\mu(B).\end{aligned}$$

Letting $\tau_0 \to \infty$, $\mu(B) \to 0$ and we have

$$\int_X f\,\mathrm{d}\mu \geq (1-\varepsilon)\int \overline{A}\,\mathrm{d}\mu.$$

This holds for any $\varepsilon > 0$, finishing the proof of the Claim.

CASE 2: This is the general case: $f \geq 0$, $f \in L^1$.

Choosing $M > 0$, we define

$$\overline{A}_M(x) = \min\{\overline{A}(x), M\}$$

and define now:

$$\tau(x) = \inf\{n \geq 1 :\ A_n(x) \geq (1-\varepsilon)\overline{A}_M(x)\}.$$

Note: Again we see that τ is finite a.s., since even though $\overline{A}(x)$ may be infinite, by definition $\overline{A}_M(x) < \infty$. Choosing $\tau_0 > 0$, we define the sets B and G to be where τ is $>$ or $< \tau_0$, as before.

Next we set

$$\tilde{f}_M(x) = \begin{cases} f(x) & \text{for } x \in G \\ f(x) + (1-\varepsilon)\overline{A}_M(x) & \text{for } x \in B. \end{cases}$$

We now
CLAIM: for all $n \geq 1$, for a.e. x, we have:

$$S_n \tilde{f}_M(x) \geq (1-\varepsilon)(n-\tau_0)\overline{A}_M(x). \tag{157}$$

The proof is like that of part (1b); with x_k the initial point of the last segment as before, if $x_k \in G$ the argument is exactly the same, as $S_n \tilde{f}_M(x) \geq S_n(x)$ and we neglect the last piece just as before, with the values of the function boosted in defining $\tilde{f}_M$, and so we again get a good bound, completing the estimate for the whole length of orbit to time n.

Next, integrating, we have for each $n \geq 1$,

$$\int_X \tilde{f}_M \, \mathrm{d}\mu \geq (1-\varepsilon)\frac{(n-\tau_0)}{n} \int_X \overline{A}_M \, \mathrm{d}\mu.$$

Also,

$$\int_X f \, \mathrm{d}\mu = \int_X \tilde{f}_M \, \mathrm{d}\mu - (1-\varepsilon)\int_B \overline{A}_M \, \mathrm{d}\mu \geq (1-\varepsilon)\frac{(n-\tau_0)}{n} \int_X \overline{A}_M \, \mathrm{d}\mu - M(1-\varepsilon)\mu(B)$$

for each n. Therefore, for each τ_0 fixed, and $\varepsilon > 0$ fixed,

$$\int_X f \, \mathrm{d}\mu \geq (1-\varepsilon) \int_X \overline{A}_M \, \mathrm{d}\mu - M(1-\varepsilon)\mu(B).$$

As τ_0, $\mu(B) \to 0$ so $\int_X f \, \mathrm{d}\mu \geq (1-\varepsilon)\int_X \overline{A}_M \, \mathrm{d}\mu$. This is true for each $\varepsilon > 0$, so we have shown that for any choice of $M > 0$, $\int_X f \, \mathrm{d}\mu \geq \int_X \overline{A}_M \, \mathrm{d}\mu$. Finally, by the Monotone Convergence Theorem, the right hand side increases to $\int_X \overline{A} \, \mathrm{d}\mu$ as M increases to $+\infty$, completing the proof of the Claim and hence of the Theorem. □

Remark 38.1. We have borrowed from Keane in the pedagogical presentation (breaking the proof into palatable cases of increasing complexity) and from Keller in two important technical points: the proof that (2) implies (1), and in the definitions of $\tilde{f}_M$ and $\overline{A}_M$. Note how the using the Monotone Convergence Theorem comes in nicely in both these arguments. (Our definition of τ for Case (2) is slightly simpler than Keller's.)

39. Appendix: The Hopf argument for ergodicity

Theorem 39.1. *(Birkhoff Ergodic Theorem, flow version) Let τ_t be a measure-preserving flow of a probability measure space $(X, \mathcal{A}, \mu)$. Then for any $f \in L^1$, for almost any $x \in X$, this limit exists:*

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(\tau_t(x))\, dt = f^+(x).$$

This function is measurable and invariant, and satisfies for any invariant measurable set E, $\int_E f d\mu = \int_E f^+ d\mu$. In particular, $\int_X f^+ d\mu = \int_X f^+ d\mu$.

Of course, the same holds for the limits at -∞, since $\widetilde{\tau}_t = \tau_{-t}$ is a measure-preserving flow; we call this limit f^-.

The proof is as a corollary of the theorem for transformations.

Lemma 39.2. *In the above situation, there is a set G of full measure such that for all $x \in G$, $f^+(x) = f^-(x)$.*

Proof. From the Birkhoff theorem, f^+ and f^- are defined on an *invariant* set of full measure. Now if the statement fails, then there is an invariant set E with $\mu(E) > 0$ such that $f^+ < f^-$ on E. But then $\int_E f d\mu = \int_E f^+ d\mu < \int_E f^- d\mu = \int_E f d\mu$, a contradiction. □

Remark 39.1. For transformations, the fact that f^+ is defined on an *invariant* set of full measure follows once we know the limit exists a.s., because if the limit exists for some x it exists for any $y \in \mathcal{O}(x)$. For flows this is not so obvious. It is true if f is integrable along any compact subset of the orbit of x. This can be proved with some work, e.g. by using the Ambrose-Kakutani flow-under-a-function representation any applying Fubini's theorem, but it is easier here to simply make use of the Birkhoff theorem, which tells us that f^+ is invariant.

We shall need:

Theorem 39.3. *(Lusin's Theorem)Let X be a locally compact Hausdorff space with μ a Borel probability measure on X and let $f : X \to \mathbb{R}$ measurable. Let $\varepsilon > 0$. Then there exists a continuous function g with compact support such that $g = f$ on a set of measure at least $1 - \varepsilon$, and moreover $||g||_\infty \leq ||f||_\infty$.*

Note that since g has compact support, it is in fact uniformly continuous. In words, any measurable function is nearly continuous. The proof uses Urysohn's Lemma. See e.g. [Rud70] p. 53.

Proposition 39.4. *Assume the situation of Theorem 39.1, with X a locally compact Hausdorff space and μ a Borel measure. Then for any $x \in G$, the function f^+ is constant on $W^s(x)$ and the function f^- is constant on $W^u(x)$.*

Proof. We show that for $x \in G$, for all $y \in W^s(x)$, the limit defining $f^+(y)$ exists and $f^+(x) = f^+(y)$.

Assume first that f is uniformly continuous. Then ...

□

The rough idea behind Hopf's argument for proving ergodicity is the following. We wish to show that an invariant integrable function f is constant. From the Birkhoff theorem we know that there is a set G of full measure such that $f^+ = f^- \equiv \overline{f}$ exists. Given two points $x, y \in G$, we would like to show $\overline{f}$ agrees on x, y. Assuming hyperbolicity, we can connect (one hopes) x and y by finitely many paths along stable and unstable manifolds. Since $\overline{f}$ is constant there, the values are equal.

How hyperbolicity allows one to connect two points in this way is illustrated for two quite different examples: a hyperbolic toral automorphism and a geodesic flow on the upper half space $\mathbb{H}$, see Figs. ??

The actual details can be tricky, involving the idea of the **absolute continuity** of the natural maps between nearby stable/unstable leaves, the so-called **holonomy maps**.

This is true even in the most basic case of product measure. But it is worth being precise here, because the somewhat hand-waving arguments sometimes encountered only add to the possible confusion.

Theorem 39.5. *(Fubini's Theorem)*

Proposition 39.6. *(The concluding Hopf argument for product measure) Given probability spaces (X, μ_X) and (Y, μ_Y), let $G \subseteq X \times Y$ be of full measure, and let*

..... removed Feb 2020 for students

Remark 39.2. Interesting but not so useful:

$$I1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \; J1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix} \; K1 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$I1{*}J1 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \; J1{*}I1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{bmatrix} \; J1{*}K1 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \; K1{*}I1 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\text{inverse of } K1 * I1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

39.1. **Cayley graphs: choosing a point randomly from an infinite group.** We have seen above that a key idea of dynamics- the Birkhoff ergodic theorem- depends for its intuitive statement, "time average equals space average", on having a good notion of time average, that is to say, of average value for a function defined on the semigroup $\mathbb{N}$ (for a noninvertible transformation), or the groups $\mathbb{Z}$ or $\mathbb{R}$ (for an invertible transformation or a flow). And the Birkhoff theorem says exactly that, for measure-preserving dynamics on a probability space, the Cesáro average provides an appropriate notion of time average.

In these notes- even though our main focus will be on maps and flows- we will also find it both interesting and natural to treat more general actions. But then, with what should we replace the notion of time average?

One pictures the orbit of a point as a copy of this semigroup or group, wrapped in some perhaps complicated manner through the dynamical space. So the idea will be to first define an appropriate averaging method over the group, and then transport this to the orbits.

To carry this out, we need first a way to visualize the group itself.

Before describing this, we mention the variety of ways group and semigroup actions will be of interest in these notes:

– Group actions will often be used to define our dynamical spaces, on which our transformations or flows will act. Thus for example the factor space $E/\mathbb{Z}^n$ defines the torus, see below, while other manifolds can be realized e.g. by hyperbolic space $\mathbb{H}^n$ acted on by a discrete group of isometries. Examples of the dynamical systems we study on these spaces include toral rotations and automorphisms, geodesic and horocycle flows.

– The boundary at infinity of a nonamenable group is an interesting space geometrically; examples are the beautiful fractal objects, such as Kleinian limit sets; one studies several related types of dynamics: the group action on the boundary, the geodesic and horocycle flows. In addition there may be an interesting map on the boundary space with the same space of orbits as the group;

–Many other fractal sets can be thought of as analogous to this, as they are the limit sets of a semigroup action. Indeed, beginning with a noninvertible map, say two-to-one, then the collection of inverse images of a point forms an orbit for the action of a free semigroup on two generators. An example is given by the Julia set for the map $z \mapsto z^2 + c$; choosing any point $z_0 \in \mathbb{C}$, the "boundary at infinity" of its tree of preimages will be the Julia set, which is itself acted on by the free semigroup. Further examples include hyperbolic Cantor sets and Iterated Function Systems (IFSs).

To describe the geometry of a discrete group, we begin with a finitely generated group G or semigroup S, and a list of generators, $\mathcal{G} = (g_1, \dots g_n)$. The **Cayley graph** of G or S consists of one vertex for each element, connected by edges labelled by the generators. For a semigroup draw an edge labelled by $g_i \in \mathcal{G}$ from vertex g to h iff $g_i g = h$. For the case of a group, we do the same for the augmented list of generators together with their inverses, $\widetilde{\mathcal{G}} = (g_1, g_1^{-1}, \dots g_n, g_n^{-1})$.

A **word** is a finite string of generators. We consider a finite collection $\mathcal{R}$ or words, with $\widehat{\mathcal{R}}$ denote the subgroup generated by $\mathcal{R}$. A **relation** is an element of $\widehat{\mathcal{R}}$.

We denote by F_n the free group on n generators, and form the factor group $F_n/\widehat{\mathcal{R}}$.

For semigroups we proceed similarly: we write FS_m for the free semigroup on m generators (also called **letters**); we can get from this construction the free group as follows: beginnign with $m = 2n$ generators, labelled $(g_1, g_1^{-1}, \dots g_n, g_n^{-1})$, we factor by the collection of relations $\widehat{\mathcal{R}}$ generated by $\mathcal{R} = \{g_i g_i^{-1} :, 1 \leq i \leq n\}e$. That is, we mod out by the relations $gg^{-1} = e$.

Conversely, any finitely generated group G can be represented in this way, as a factor group of F_n, and so as a facto semigroup of FS_{2n}. The relations $\widehat{\mathcal{R}}$ are, geometrically, the words which form closed loops starting at e in the Cayley graph.

For the case of an abelain group, the law $ab = ba$ is achieved by including the relation $f^{-1}g^{-1}fg$.

The Cayley graph in the case of a group is **homogeneous** in that its geometry everywhere is the same, and is just like that n the identity e.

A homomorphism from a group G to a group H can be visualized by a continuous map of the Cayley graphs; a good example to keep in mind is the homomorphism from the free group F_2 on two generators (a, b) to the free abelian group on two generators, $\mathbb{Z}^2$, and from $\mathbb{Z}^2$ to $\mathbb{Z}_6 = \mathbb{Z}_2 \oplus \mathbb{Z}_3$. See Fig. ??.

In what follows, a key notion will be that of a *random walk* on G.

factor groups, free semigroup, free group, free abelian group, finite abelian group

fundamental domain; lattice subgroup

random walk

boundary at infinity

hitting measure

example: Parry measure

normal subgroup

left/right actions; free semigroup boundary and IFS/ Cantor set
Free semigroup and group automorphisms
Kleinian limit set, Patterson measure

In this section we briefly touch on a number of matters which call out for a much deeper look, and so are returned to later in these notes. However, we must at least mention these now so the reader should keep this broader picture in mind as she or he progresses through the rest.

Let G be a discrete group, by which we mean a finitely generated group; examples to keep in mind are the finite groups ...$\mathbb{Z}_d$, semidirect prod, perm, and infinite groups .... free, infinite abelian; fund group of surface, matrix groups

We can turn these into geometric objects by way of the picture of the group known as a Cayley graph. ...

role of independence: prob measure on G; random walk; convolution we have already encountered convol on reals (of functions/measures)

random walk on $\mathbb{Z}$: limit laws;
random walk on free: limit laws Proof!
random walk on graph: Markov
random walk on $\mathbb{Z}$: convergence to Brownian; asip
harmonic function; average of function via random walk; harmonic projection
local CLT; Cesaro average
Benford
not every process is iid (next section)

Example 61. *Choosing a point randomly from an amenable group.* Infinite groups split into two types, depending on exactly this question we have been considering. An **amenable** group is by definition one for which there exists an invariant mean for the action of the group on itself by left translation. That is, it is exactly for amenable groups that the idea of an average value for a bounded function makes sense.

To get ahold of this idea, we need to consider some basic examples of groups which are amenable, and some which are not.

...
free group boundary at infinity
The importance of this notion is indicated by these several equivalent properties:

Theorem 39.7.
(i) G is amenable;
(ii) there exists a Følner sequence for G
(iii) given an ergodic measure-preserving action of G on a probability space, ???
convex affine action fixed point
In the case where G is finitely generated, we have additionally:
(iv) there exists a probability measure μ on the generators so that there are no bounded $\mu-$harmonic functions;
(v) the $\mu-$boundary at infinity of the group is trivial (a single point);

random walks

Example 62. *Choosing a point randomly from a nonamenable group or semigroup.* By what we have said above, nonamenable groups are exactly those for which the idea of an average value does *not* makes sense. Nevertheless, let us propose two possible approaches and see what happens !

For simplicity we again restrict to the case where G is finitely generated, and we are given a probability measure μ on the generators.

The most basic example of a nonamenable group is the free group on two generators $F_2 = \langle a, b\rangle$. This is pictured in Fig. ??, which shows its **Cayley graph**; for this graph, the vertices are the group elements, while the edges are taken from the set of generators and their inverses, $\mathbb{E} = \{a, b, a^{-1}, b^{-1}\}$, translated to each vertex; that is, there is an edge e from g to h iff $ge = h$ for $e \in \mathbb{E}$.

Actually it is easier to begin one step back, with the **free semigroup** on two generators, FS_2, since this consists of all finite **words** in the alphabet $\mathcal{A} = \{a, b\}$, $a_0 \dots a_n$ with $a_i \in \mathcal{A}$, together with the **empty word** denoted e which is the identity. Multiplication in FS_2 is given by concatenation of two words; that is, for $g = a_0 \dots a_n$ and $h = b_0 \dots b_m$ then $gh = a_0 \dots a_n b_0 \dots b_m$. Now the Cayley graph is a binary tree, Fig. ??.

We define F_2 by starting with the free semigroup on four generators $\mathcal{A} = \{a, b, \widetilde{a}, \widetilde{b}\}$ and then factoring out by the group generated by the relation set $\mathcal{R} = \{a\widetilde{a}, \widetilde{a}a, b\widetilde{b}, \widetilde{b}b\}$ (this is the best formal way to give the definition, although all this means is that we are defining $a\widetilde{a} = e$, in other words taking $\widetilde{a}$ to be equal to a^{-1}.) The elements in F_2 are then the reduced words $a_0 \dots a_n$ together such that $a_i a_{i+1}$ cannot belong to the set $\mathcal{R}$. Multiplication in F_2 is now given by concatenation *followed by* cancellation of all $a_i a_{i+1} \in \mathcal{R}$.

Note now that the number of words of length n in this group grows exponentially, compared to $\mathbb{Z}^d$ where the growth is polynomial (like d^n).

removed removed Aug 2016 for students .....

39.2. **Choosing a point randomly from an amenable group, or from Euclidean space.** An examination of this for other noncompact groups brings us to the idea of **amenability**, and to the beginning of a long and fascinating story.

A group G is termed **amenable** (a pun; it should be "a*mean*able!") iff there exists an invariant mean on it (on $l^\infty(G)$ if the group is discrete; on $L^\infty(G, m)$ where m is Haar measure if G is a continuous group which is locally compact, so Haar measure exists). [Gre69]

There are several equivalent notions:

Furst defn of amenable.

Harmonic projection.

Save rest for after boundary- in examples section!

39.3. **Choosing a point randomly from a nonamenable group, or from hyperbolic space.** Equivalent notions of non-amenable.

Basic example: F_2.

Appendix: Aside: Cayley graph of a semigroup or group. F_2 to $\mathbb{Z}^2$. Generators and relations.

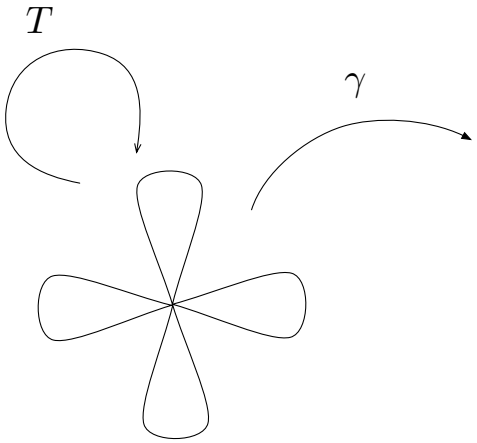
Simplest case: finitely generated discrete.

μ on generators. Top invariant mean: Harmonic function. Harm projection. Equivalent defs. Boundary at ∞. Boundary values. Mokobodski mean.

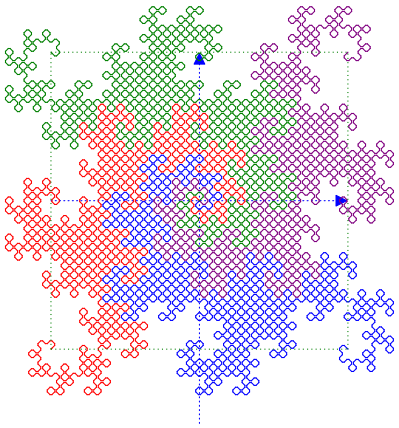
Test functions: group actions. Random/ Markov ergodic theorem.

Fractal sets, again (top invt measures; IFS)

end of removed for students .....



(a) The space-filling curve γ gives a measure-preserving semiconjugacy from the Markov map on the cloverleaf to the doubling map on the torus.



(b) A fundamental domain for the the Ito-Misutani Tetradragon, showing the four elements of a Markov partition for the doubling map of the torus.

40. Transitive points and Baire category Take 2

The usual understanding of a property holding for "almost all" points of a space is measure theoretical, that the complement of the subset where this is valid have measure zero. A complementary, purely topological notion is provided by sets of second Baire category. The fascinating relationship between these two very different but in many ways parallel ideas is explored in Oxtoby's wonderful little book [Oxt80] (which cannot be recommended too highly, e.g. for graduate students refining their knowledge for analysis qualfying exams, for professors preparing for a lecture course, or for anyone with the time to pursue beautifully presented ideas for their own sake). Here we bring in some basic definitions and one result, specifically related to dynamics. The methods involved may remind one of Poincaré recurrence, specifically of Walters' proof of the previous section.

Definition 40.1. Let $(X, \mathcal{T})$ be a topological space. A subset A is **meagre** or **of first (Baire) category** iff it is a subset of a countable union of nowhere dense sets. It is **residual** or **of second category** iff it contains a countable intersection of dense open sets.

We recall that a countable union of closed sets is called an $F_\sigma-$set, while its complement, a countable intersection of open sets, is a G_δ. Thus a meagre set is contained in an F_σ of a special type, while a residual set contains a G_δ. Also, the complement of a meagre set is residual and vice-versa. Much of the utility of this notion comes from the **Baire Category Theorem**, which guarantees a **dense** G_δ subset:

Theorem 40.1. *Let X be a Polish space. Then a residual set is dense.*

Proof. Let G_i be open and dense, for $i = 1, 2, \ldots$. We shall show that

$$E = \cap_{i=1}^{\infty} G_i$$

is dense. Let $\mathcal{U}$ be an open subset of X, and assume that $d(\cdot, \cdot)$ is a metric compatible with the topology of X, for which X is complete. Since G_1 is dense, there exists $x_1 \in \mathcal{U} \cap G_1$, and there exists $\delta_1 > 0$ such that for the ball of that radius, $B_{\delta_1}(x_1) \subseteq \mathcal{U} \cap G_1$. Now there exists $x_2 \in B_{\delta_1}(x_1) \cap G_2$ and $\delta_2 > 0$ such that $B_{\delta_2}(x_2) \subseteq B_{\delta_1}(x_1) \cap G_2$. Continuing in this manner, the sequence $(x_i)_{i \geq 1}$ is Cauchy; by completeness this sequence has a limit point x, and by construction $x \in \mathcal{U} \cap G_k$ for all k. □

Definition 40.2. Let T be a homeomorphism of X. A point $x \in X$ is **transitive** iff it has a dense orbit. The map T is transitive iff there exists a transitive point.

If T is continuous but not necessarily invertible, we say a point is **forward transitive** iff it has a dense forward orbit, and the map is forward transitive iff there exists a forward transitive point.

Proposition 40.2. *Let $(X, \mathcal{T})$ be a Polish space with no isolated points.*
(i)Let T be a homeomorphism. Then if T is transitive, the set E of forward transitive points is residual.
(ii)Let T be a continuous map. Then if T is forward transitive, the set E of forward transitive points is residual.

We note that in (i), by having biinfinite orbits in the hypothesis and forward orbits in the conclusion, the statement is stronger in both respects. That is, the existence of a single biinfinitely transitive point implies existence of (many) forward transitive points: a residual set hence (by the Baire Category Theorem) a dense G_δ of them. Without the assumption of no isolated points this can fail, as shown by a simple example on p. 129 of [Wal82], of a homeomorphism with a dense biinfinite orbit but no dense forward orbit (imagine the left shift map on the two-point compactification $\mathbb{Z} \cup \{-\infty, +\infty\}$of $\mathbb{Z}$).

Proof. With metric d as above, since X is a separable metric space there exists a countable base $\{\mathcal{U}_i\}_{i \geq 1}$ for the topology. Then E is the set of points x such that for each $j \geq 1$, the forward orbit of x meets $\mathcal{U}_j$. That is, for each j, $E \subseteq G_j \equiv \cup_{n \geq 0} T^{-n}(\mathcal{U}_j)$, so we can write:

$$E = \inf G_j = \lim_{j \geq 1} \sup_{n \geq 0} T^{-n}(\mathcal{U}_j) = \cap_{j \geq 1} \cup_{n \geq 0} T^{-n}(\mathcal{U}_j).$$

We claim that each of the open sets G_j is dense. We wish to show that for each $i \geq 1$, $\mathcal{U}_i$ meets G_j. Now there exists a transitive point w; that is, for (i), the biinfinite orbit $(T^n(w))_{n \in \mathbb{Z}}$ is dense; for (ii) we know this for the forward orbit. Furthermore,

since X has no isolated points this collection of points must be infinite. Now any dense infinite sequence of distinct points must meet an open set $\mathcal{U}$ infinitely often: singletons are closed sets in a metric space, so $\mathcal{U} \setminus \{x\}$ is again nonempty open and we can find the next such element. Given $i, j \geq 0$, therefore, in either case, the orbit of w enters both $\mathcal{U}_i$ and $\mathcal{U}_j$ infinitely often, one of them first. If we know w is forward transitive, then from this we know there is a pair of times such that $\mathcal{U}_i$ occurs first. That is, there exists a point x and an $k > 0$ such that $x \in \mathcal{U}_i$ and $T^k(x) \in \mathcal{U}_j$, equivalently, $x \in \mathcal{U}_i \cap T^{-k}(\mathcal{U}_j) \subseteq \mathcal{U}_i \cap G_j$. Thus $\mathcal{U}_i \cap G_j$ is nonempty and hence G_j is dense as claimed, so E is a countable intersection of open dense sets and hence is residual.

If we only know w is biinfinitely transitive, we have to be slightly more careful. Now if $\mathcal{U}_i$ occurs first, the rest of the argument is the same. But for a general argument, we know there exists $n \geq 0$ such that either $\mathcal{U}_i \cap T^n(\mathcal{U}_j)$ or $\mathcal{U}_j \cap T^n(\mathcal{U}_i)$ is nonempty. Call this set $\mathcal{U}$ in either case. We claim that there exists $k > 0$ such that $\mathcal{U}_i \cap T^k(\mathcal{U}_j)$ is nonempty, and then will proceed as before. But the transitive point enters $\mathcal{U}$ infinitely many times, so there exists $m > 0$ and x such that $x \in \mathcal{U}$ and $T^m(x) \in \mathcal{U}$, and we take simply $k = m$ in the first case, or $k = m - n$ in the second and then proceed as before.

□

Something similar can be proved for semiflows and flows, with the "no isolated points" property replaced by an appropriate condition (that compact pieces of flow orbits cannot fill up an open subset). Also, there is a general result which holds for continuous group actions.

Now we move on to flows and semiflows. First we have these definitions:

Definition 40.3. Given a topological space $(X, \mathcal{T})$, a **continuous flow** τ_t on X is a jointly continuous map $\tau : X \times \mathbb{R} \to X$ which satisfies the **flow property**: writing $\tau_t(x) = \tau(x, t)$, this is $\tau_{t+s} = \tau_t \circ \tau_s$. To define a continuous **semiflow** we replace $\mathbb{R}$ above by $\mathbb{R}^+ = [0, +\infty)$.

Given a flow on X, a point $x \in X$ is **transitive** iff it has a dense biinfinite orbit, $\{\tau_t(x) : t \in \mathbb{R}\}$; the flow is transitive iff there exists a transitive point. A point x is **forward transitive** for a flow or semiflow if the forward orbit $\{\tau_t(x) : t \geq 0\}$ is dense, and a (semi)flow is forward transitive iff there exists such a point.

For our statement we need a replacement for the notion of no isolated points.

Definition 40.4. Given a topological (semi)flow τ_t on X, we say the flow has **no isolated orbit segments** iff given any nonempty open set $\mathcal{U}$, $x \in X$ and $J \subseteq \mathbb{R}$ a compact interval, then $\mathcal{U} \setminus \{\tau_t(x) : x \in J\}$ is a nonempty open set (it is open since the continuous image of a compact set is compact).

Here is a topological property of X which guarantees this. A **curve** in X is a continuous bijective map $\gamma : J \to X$ where J is some subinterval of $\mathbb{R}$. We say the space X has **no isolated compact curves** iff given any nonempty open set $\mathcal{U}$, and any curve γ defined on J a compact interval, then $\mathcal{U} \setminus \gamma(J)$ is a nonempty open set. Since orbits of τ_t are curves, this implies the no isolated orbit segment property.

Note that if X is a metric space, it is equivalent to require that this be nonempty for all balls $\mathcal{U} = B_\delta(x)$.

Proposition 40.3. *Let $(X, \mathcal{T})$ be a Polish space, and let τ_t be a (semi)flow on X, with no isolated orbit segments.*
(i) Then if τ_t is a transitive flow, the set E of forward transitive points is residual.
(ii)If τ_t is a forward transitive semiflow, the set E of forward transitive points is residual.

The same holds if instead of asumming no isolated orbit segments, we assume that there exists t_0 such that the time-t_0 map $T \equiv \tau_{t_0}$ has a (forward) transitive point.

Proof. As before, let $\{\mathcal{U}_i\}_{i\geq 1}$ be a countable base for the topology $\mathcal{T}$ of the separable metric space (X, d). Now we have

$$E = \inf G_j = \cap_{j\geq 1} \cup_{t\geq 0} \tau_{-t}(\mathcal{U}_j)).$$

We wish to show each of the open sets G_j is dense. Thus we claim that for each $i \geq 1$, $\mathcal{U}_i$ meets G_j.

For part (i), we are given that there exists a transitive point w with $\{\tau_t(x) : t \in \mathbb{R}\}$ dense; for (ii) we know this for $\mathbb{R}^+$. Since the flow space X has no isolated orbit segments, this orbit cannot be periodic.

Since there are no isolated orbit segments, there are no isolated points. Thus X contains at least two disjoint open balls.

Let $\mathcal{U}$ be an open set. Since τ_t is jointly continuous, given $w \in X$, the curve $\tau_t(w)$ is a continuous map from $\mathbb{R}$ to X, so the inverse image in $\mathbb{R}$ of U is open hence a countable union of disjoint intervals. Since there are no isolated orbit segments, this inverse image is unbounded at $+\infty$ for a semiflow, and also at -∞ for a flow.

Now we argue as for the case of a map. Given $i, j \geq 0$, the orbit of w enters both $\mathcal{U}_i$ and $\mathcal{U}_j$ in an unbounded, infinite number of time intervals. So it enters one of them first. If it is $\mathcal{U}_i$, we are done, exactly as before, as $\mathcal{U}_i$ meets G_j. Now if the flow or semiflow is forward transitive, this is always the case for some point, as the intervals corresponding to $\mathcal{U}_i$ and $\mathcal{U}_j$ each occur infinitely often towards $+\infty$. Lastly, in the flow case, given a biinfinitely transitive point and that $\mathcal{U}_j$ occurs first, then there is an interval of times $J = [a, b]$ such that $\mathcal{U} \equiv \mathcal{U}_j \cap \{\tau_{-t}(\mathcal{U}_i) : \ t \in J\}$ is nonempty. By the property of no isolated orbit segments, there is some $s > b$ such that $\mathcal{U} \cap \tau_{-s}(\mathcal{U})$ is nonempty. Therefore, reasoning as for discrete time, $\mathcal{U}_i$ meets G_j and we are done. □

Remark 40.1. One can extend the notion of transitive point to an action of a group G action on a topological space X by continuous maps (hence by homeomorphisms, since we have inverses). Note that the group itself is not required to have a topology. We shall say the action is **dynamically** transitive iff there exists a transitive point. The reason we have added the modifier "dynamically" is because of this much stronger property: the action is called **transitive** iff for each $x, y \in X$ there exists $g \in G$ with $g(x) = y$. Thus a transitive transformation or flow is generally not transitive as a $\mathbb{Z}-$ or $\mathbb{R}-$action!

41. Nonstationary transitions

Defining coordinate functions on Σ by $X_i(x) = x_i$ for $x = (\dots x_{-1}x_0x_1\dots) \in \Sigma$, then the sequence of measurable functions X_i on the measure space $(\Sigma, \mathcal{B}, \mu)$ give in

the language of probability theory a **stochastic process**, also called a **sequence of random variables**. This is a **stationary process** if and only if the measure μ is shift-invariant. The Shannon-Parry measure μ is invariant and so defines a stationary Markov process. However, as we mentioned above, a general Markov measure on Σ^+ or Σ need not be invariant; an example is given by the Shannon-Parry eigenmeasure ν, which is defined from the same transition matrix P but beginning with the non-invariant initial distribution $\mathbf{w^t}$.

Nevertheless, for this example the transition probabilities are stationary, while for a general Markov process this also need not be the case, with the single matrix P replaced by a sequence of row-stochastic matrices. In the case we shall study, these may be rectangular, as the alphabets may change as well.

Given a sequence $\underline{M} = (M_i)_{i\in\mathbb{Z}}$ of $(l_i \times l_{i+1})$ nonnegative matrices, each matrix M_i acts on the left on column vectors and on the right on row vectors. Therefore, writing C_i for the space of column vectors, and R_i for the space of row vectors, both isomorphic to $\mathbb{R}^{l_i}$, we have these diagrams:

$$\cdots R_{-1} \xrightarrow{M_{-1}} R_0 \xrightarrow{M_0} R_1 \xrightarrow{M_1} R_2 \xrightarrow{M_2} R_3 \cdots$$

and

$$\cdots C_{-1} \xleftarrow{M_{-1}} C_0 \xleftarrow{M_0} C_1 \xleftarrow{M_1} C_2 \xleftarrow{M_2} C_3 \cdots$$

We write C_i^+ for the cone of nonnegative vectors in C_i, i.e. column nonzero vectors with each entry ≥ 0, and similarly for row vectors, $R_i^+ \subseteq R_i$.

Let $\mathrm{Proj} : (C_i^+ - \underline{0}) \to \Delta_i$ be the projection $\mathbf{v} \mapsto \mathbf{v}/||\mathbf{v}||$ where $||\mathbf{v}|| \equiv \sum |v_k|$ as before, and also write $\mathrm{Proj} : (R_i - \underline{0}) \to \Delta_i^t$ for the projection $\mathbf{v}^t \mapsto \mathbf{v}^t/||\mathbf{v}^t||$. We consider the induced projective action of the matrices M_i on the positive simplices $\Delta_i \subseteq C_i^+$, $\Delta_i^t \subseteq R_i^+$, and define for $k \in \mathbb{Z}$ and $n \geq 0$ the sets $\Delta_{k(+n)} \subseteq \Delta_k$ by

$$\Delta_{k(+n)} = \mathrm{Proj}(M_k M_{k+1} \cdots M_{k+n} \Delta_{k+n+1}),$$

and we define the set $\Delta_{k(-n)}^t \subseteq \Delta_k^t$ by

$$\Delta_{k(-n)}^t = \mathrm{Proj}(\Delta_{k-n}^t M_{k-n} \cdots M_{k-2} M_{k-1}).$$

Note that these are nested:

$$\Delta_k = \Delta_{k(+0)} \supseteq \Delta_{k(+1)} \supseteq \ldots$$

and

$$\Delta_k^t = \Delta_{k(-0)}^t \supseteq \Delta_{k(-1)}^t \supseteq \cdots .$$

We write $\Delta_{k(-\infty)}^t, \Delta_{k(+\infty)}$ for the intersections.

41.1. **Nonstationary Shannon-Parry measures.** We shall next describe such a generalization of the Shannon-Parry measure. In the nicest case, when we have the future and past Perron-Frobenius properties (see below), we will construct Shannon-Parry measures much as before, and these will be unique. However for what follows it will be important to consider the more general situation.

For this, we specialize to transition matrices. We begin with a sequence of finite alphabets, $\mathcal{A}_i$ for $i \in \mathbb{Z}$, with $\#\mathcal{A}_i = l_i$, and with $\mathcal{A}_i = \{0, 1, \ldots, l_i - 1\}$. Given a sequence $\underline{L} = (L_i)_{i\in\mathbb{Z}}$ of $(l_i \times l_{i+1})$ $0-1$ matrices, we let $\Sigma_{\underline{L}} \subseteq \Pi_i \mathcal{A}_i$ denote the set of

allowed strings, $x = (\dots x_0 x_1 \dots)$ such that the $(x_i x_{i+1})$ entry of L_i equals 1 for all $i \in \mathbb{Z}$.

We next define

$$\Omega_{\underline{L}} = \{\underline{\widehat{\mathbf{w}}} = (\dots \widehat{\mathbf{w}}_{-1}\widehat{\mathbf{w}}_0\widehat{\mathbf{w}}_1 \dots) \text{ such that } \widehat{\mathbf{w}}_i = L_i\widehat{\mathbf{w}}_{i+1} \text{ and } \widehat{\mathbf{w}}_i \in (C_i^+ - \underline{0})\}$$

and

$$\widehat{\Omega}^t_{\underline{L}} = \{\underline{\widehat{\mathbf{v}}}^t = (\dots \widehat{\mathbf{v}}^t_{-1}\widehat{\mathbf{v}}^t_0\widehat{\mathbf{v}}_1 \dots) \text{ such that } \widehat{\mathbf{v}}_i L_i = \widehat{\mathbf{v}}_{i+1} \text{ and } \widehat{\mathbf{v}}_i \in (R_i^+ - \underline{0})\}.$$

Lemma 41.1. *$\widehat{\Omega}_{\underline{L}}$ is a nonempty convex compact set, and the number of its extreme points is at most $\liminf_{i\geq 0}(l_i)$. The same is true for $\widehat{\Omega}^t_{\underline{L}}$, with extreme points $\leq \liminf_{i\leq 0}(l_i)$.*

Proof. We can build a sequence $\underline{\widehat{\mathbf{w}}}$ constructively: first choose $\widehat{\mathbf{w}}_k \in \Delta_{k(+\infty)}$ for some $k \in \mathbb{Z}$, and then define $\widehat{\mathbf{w}}_i$ for $i < k$ by $\widehat{\mathbf{w}}_{-1} = L_{-1}\widehat{\mathbf{w}}_0, \widehat{\mathbf{w}}_{-2} = L_{-2}\widehat{\mathbf{w}}_{-1}, \dots$. For times $> k$, we note that the L_i may not be invertible, so we have to make choices in the inverse images by matrix multiplication, possibly at each level: we choose first $\widehat{\mathbf{w}}_{k+1}$ such that $L_0\widehat{\mathbf{w}}_{k+1} = \widehat{\mathbf{w}}_k$. The preimage exists since $\widehat{\mathbf{w}}_k \in \Delta_{k(+\infty)}$. Next we choose $\widehat{\mathbf{w}}_{k+2}$ such that $L_1\widehat{\mathbf{w}}_{k+2} = \widehat{\mathbf{w}}_{k+1}$, and so on. Note that this constructive procedure yields the entire collection $\widehat{\Omega}_{\underline{L}}$.

For $\widehat{\Omega}^t_{\underline{L}}$ a sequence $\underline{\widehat{\mathbf{v}}}^t \in \widehat{\Omega}^t_{\underline{L}}$, the procedure is similar: choose $\widehat{\mathbf{v}}^t_k \in \Delta^t_{k,-\infty}$ for some k and then define $\widehat{\mathbf{v}}^t_i$ for $i > k$ by $\widehat{\mathbf{v}}^t_{k+1} = \widehat{\mathbf{v}}^t_k L_k$ and so on, and for $\widehat{\mathbf{v}}^t_{k-m}$ making choices in the successive preimages.

The extreme point count is like the proof of Lemma 16.3, taking into account the inverse images. □

Given $\underline{\widehat{\mathbf{w}}} \in \widehat{\Omega}_{\underline{L}}$, we define a second sequence $\underline{\mathbf{w}} = (\mathbf{w}_i)$ by $\mathbf{w}_i = \widehat{\mathbf{w}}_i/||\widehat{\mathbf{w}}_i|| = \text{Proj}(\widehat{\mathbf{w}}_i)$; these have been normalized so as to be in the simplex Δ_i. We write $\Omega_{\underline{L}}$ for the collection of these normalized sequences.

Next, we define a sequence of real numbers λ_i by $\lambda_i = ||\widehat{\mathbf{w}}_i||/||\widehat{\mathbf{w}}_{i+1}||$. Therefore, for each $i \in \mathbb{Z}$,

$$L_i\mathbf{w}_{i+1} = \lambda_i\mathbf{w}_i.$$

We say the normalized positive column vectors $\mathbf{w}_{i+1}$ form an **eigenvector sequence with eigenvalues** λ_i for the matrix sequence $\underline{L}$; the $\widehat{\mathbf{w}}_i$ form an eigenvector sequence with constant eigenvalue sequence 1.

Given a sequence $\underline{\widehat{\mathbf{v}}}^t \in \widehat{\Omega}^t_{\underline{L}}$ define the sequence $\underline{\mathbf{v}}^t = (\mathbf{v}^t_i)$ for $i \in \mathbb{Z}$ by $\mathbf{v}^t_i = \widehat{\mathbf{v}}^t_i/(\widehat{\mathbf{v}}^t_i\mathbf{w}_i)$. Thus, the vectors $\mathbf{v}_i$ are normalized so their inner product with $\mathbf{w}_i$ is 1.

Lemma 41.2. *$\mathbf{v}^t_i$ is an eigenvector sequence with the same eigenvalues as $\mathbf{w}_i$.*

Proof. If we define $\tilde{\lambda}_i$ by $\mathbf{v}^t_i L_i = \tilde{\lambda}_i\mathbf{v}^t_{i+1}$, then we have $\tilde{\lambda}_i = \lambda_i$, because:

$$\tilde{\lambda}_i = \tilde{\lambda}_i(\mathbf{v}^t_{i+1}\mathbf{w}_{i+1}) = (\mathbf{v}^t_i L_i)\mathbf{w}_{i+1} = \mathbf{v}^t_i(L_i\mathbf{w}_{i+1}) = \mathbf{v}^t_i(\lambda_i\mathbf{w}_i) = \lambda_i.$$

□

We define a sequence of positive row vectors $\underline{\boldsymbol{\pi}^t} = (\boldsymbol{\pi}^t_i)_{i\in\mathbb{Z}}$ by $(\boldsymbol{\pi}_i)_k = (\mathbf{v}_i)_k(\mathbf{w}_i)_k$ for $i \in \mathbb{Z}$ and where k is the index for the i^{th} alphabet. By the chosen normalization $\mathbf{v}^t_i\mathbf{w}_i = 1$, $\boldsymbol{\pi}^t_i$ is an element of Δ^t_i.

Next we define a matrix sequence $\underline{P} = (P_i)_{i\in\mathbb{Z}}$ by $P_i = \frac{1}{\lambda_i} W_i^{-1} L_i W_{i+1}$, where W_i is the $(l_i \times l_i)$ diagonal matrix with the entries of the vector $\mathbf{w}_i$ on the diagonal. Writing $\mathbf{1}_i$ for the column vector with l_i entries all equal to 1, we have as before for the case of a single matrix,

$$P_i \mathbf{1}_{i+1} = \mathbf{1}_i, \tag{158}$$

and

$$\boldsymbol{\pi}_i^t P_i = \boldsymbol{\pi}_{i+1}^t. \tag{159}$$

These are right and left eigenvector sequences with constant eigenvalue 1.

Given our $0-1$ matrix sequence $(L_i)_{i\in\mathbb{Z}}$ and a choice of $\underline{\mathbf{w}} \in \Omega_{\underline{L}}$, $\underline{\mathbf{v}}^t \in \Omega_{\underline{L}}^t$ we define the measures μ and ν exactly as before, but now using the nonstationary row-stochastic transition matrices P_i.

Thus, for $k, m \in \mathbb{Z}$ with $k < m$, we define $\lambda^{(k,m)} = \Pi_{i=k}^{m-1} \lambda_i$. The measure of a cylinder set is now:

$$\mu([x_k \dots x_m]) = (\boldsymbol{\pi}_k^t)_{x_k} (P_k)_{x_k x_{k+1}} \cdots (P_{m-1})_{x_{m-1} x_m} = (1/\lambda^{(k,m)})(\mathbf{v}_k)_{x_k} (\mathbf{w}_m)_{x_m}. \tag{160}$$

The eigenmeasure ν is defined, on $\Sigma_{\underline{L}}+$, as before:

$$\nu([x_0 \dots x_m]) = \mu([x_0 \dots x_m])/(\mathbf{v}_0)_{x_0} = (1/\lambda^{(0,m)})(\mathbf{w}_m)_{x_m}. \tag{161}$$

We now have an even nicer formula for ν, in terms of the non-normalized eigenvector sequence $\widehat{\mathbf{w}}_i$:

$$\nu([x_0 \dots x_m]) = 1/(\widehat{\mathbf{w}}_m)_{x_m}. \tag{162}$$

We indicate the dependence on our choices $\underline{\widehat{\mathbf{w}}} \in \widehat{\Omega}_{\underline{L}}$, $\underline{\widehat{\mathbf{v}}}^t \in \widehat{\Omega}_{\underline{L}}^t$ or, equivalently, of the normalized sequences $\underline{\mathbf{w}} \in \Omega_{\underline{L}}$ and $\underline{\mathbf{v}}^t \in \Omega_{\underline{L}}^t$ by: $\underline{\boldsymbol{\pi}}^t_{\underline{\mathbf{w}},\underline{\mathbf{v}}}$, $\underline{P}_{\underline{\mathbf{w}},\underline{\mathbf{v}}}$, and similarly for the measures defined from these: $\mu_{\underline{\mathbf{w}},\underline{\mathbf{v}}}$ and $\nu_{\underline{\mathbf{w}},\underline{\mathbf{v}}}$. We call these collections respectively the **Shannon-Parry invariant measures** and **Shannon-Parry eigenmeasures** for $\underline{L}$.

41.2. **The one-sided case.** Next we consider the situation where the sequence (L_i) is only defined for $i \geq 0$.

We choose a strictly positive vector $\mathbf{v}$ in Δ^t, and normalize it to the vector $\mathbf{v_0}$ which satisfies $\mathbf{v_0^t w}_0 = 1$. Beginning with this vector, we then define the rest of the sequence $\mathbf{v}_i^t$ as before, for $i \geq 0$. This again gives a positive eigenvector sequence with eigenvalues λ_i.

From this we define the row probability vectors $\boldsymbol{\pi}_i^t$ and transition matrices P_i, and the measures μ and ν on the one-sided space Σ_L^+, exactly as before.

To indicate the dependence on these choices, we write $\mu = \mu_{\mathbf{w},\mathbf{v}_0}$. Noting that by (162) the definition of ν does not depend on the choice of $\mathbf{v} = \mathbf{v}_0$ (while μ does), we write it as $\nu = \nu_{\mathbf{w}}$.

Here is an equivalent procedure: extend (L_i) to a two-sided sequence, choosing L_i for $i < 0$ to be the identity $l_0 \times l_0$ matrix and apply the two-sided construction. Now $\Delta^t_{(0,-\infty)}$ is equal to Δ_0^t, so this procedure will give the same freedom for the choice of $\mathbf{v}$.

41.3. **Mixing for nonstationary one-sided sequences.** The key to the proof of the Bowen-Marcus lemma in the nonstationary context will be to extend the mixing condition in the appropriate way.

We first give a definition for the case of a general Markov measure on a one- or two-sided nonstationary space.

Definition 41.1. Let P_i for $i \geq 0$ be a sequence of $(l_i \times l_{i+1})$ row-stochastic matrices. Let Σ^+ denote $\Pi_{i=0}^{\infty} \mathcal{A}_i$, where $\mathcal{A}_i$ are alphabets with cardinality l_i. For $0 \leq k \leq n$ integers, write $P^{(k,n)} = P_k P_{k+1} \cdots P_n$. Let $\boldsymbol{\pi}_0^t$ be an element of Δ_0^t, and define a measure μ on Σ^+ by the first half of formula (160). Thus $P^{(k,n)}$ is $(l_k \times l_{n+1})$. We call (Σ^+, μ) a **nonstationary Markov chain**.

Define the sequence of row vectors $\boldsymbol{\pi}_0^t, \boldsymbol{\pi}_1^t = \boldsymbol{\pi}_0^t P_0, \ldots, \boldsymbol{\pi}_m^t = \boldsymbol{\pi}_0^t P^{(0,m)}$. Since P_i is row-stochastic, $\boldsymbol{\pi}_i^t \in \Delta_i^t$, and this is a eigenvector sequence with eigenvalue 1.

We write $Q_{\boldsymbol{\pi}_m^t}^{(k,m)}$ for the $(l_k \times l_{m+1})$ matrix all of whose rows are $\boldsymbol{\pi}_m^t$.

Recall that, for $k \leq m$, $\mathcal{B}_k^m$ is the $\sigma-$algebra generated by the collection of thin cylinder sets $\mathbb{C}_k^m$ (the sets such that the symbols $x_k, x_{k+1}, \ldots, x_m$ are fixed).

We say: the nonstationary Markov chain (Σ^+, μ) is **(future) mixing** iff for any fixed $k \geq 0$, given $\varepsilon > 0$, for $m > k$ sufficiently large we have that, for every $A \in \mathcal{B}_0^k$ and $B \in \mathcal{B}_m^{\infty}$, then $\mu(A \cap B) = (1 \pm \varepsilon)\mu(A)\mu(B)$.

We define a metric on the nonnegative $(m \times n)$ matrices:

$$d(A, B) = \sup_i d(A_{i*}, B_{i*}) \tag{163}$$

where A_{i*} indicates the i^{th} row of A, and d is the projective metric on the standard cone $\mathbb{R}^{n+}$, and have, exactly as in Proposition **??**:

Lemma 41.3. *The Markov nonstationary chain is future mixing iff: for k fixed, given $\varepsilon > 0$, then for $m > k$ sufficiently large we have $d(P^{(k,m)}, Q_{\boldsymbol{\pi}_m^t}^{(k,m)}) \leq \varepsilon$.*

Definition 41.2. The sequence $\underline{M}_i$ of $(l_i \times l_{i+1})$ nonnegative matrices is **(future) Perron-Frobenius** iff for all $k \in \mathbb{Z}$, $\Delta_{k(+\infty)}$ is a singleton. We say $\underline{M}_i$ is **(future) focussing** iff $k \geq 0$, given $\varepsilon > 0$ for $m > k$ sufficiently large then the projective diameter of $\Delta_k^t M_k \cdots M_m$ is $< \varepsilon$.

If the matrix sequence is two-sided (indexed by $i \in \mathbb{Z}$ rather then $i \geq 0$), we define the properties **(past) mixing, Perron-Frobenius, focussing** in the obvious the symmetric way; all results hold here for the past conditions, so we only state the future versions.

Now we see the power of Birkhoff's contraction formula (92) and Corollary 24.9:

Lemma 41.4. *The sequence $\underline{M}_i$ is (future) Perron-Frobenius iff it is focussing.*

Proof. By Corollary 24.9, the opening of $M_k \cdots M_m$ is equal to that of $(M_k \cdots M_m)^t$ and these are the projective diameters of $M_k \cdots M_m \Delta_m$ and $\Delta_k^t M_k \cdots M_m$ respectively. The images $M_k \cdots M_m \Delta_m$ are nested decreasing sets, and the limiting projective diameter is zero iff the intersection is a singleton. □

Remark 41.1. The two conditions, Perron-Frobenius and focussing, are apparently quite different as the *order of applying* the matrices is reversed. And since for

$\Delta_k^t M_k \cdots M_m$ more matrices are added in post-composition of the maps, rather than pre-composition, this changes everything: the sets are certainly not nested, indeed, they do not even belong to the same space (for instance the dimension= l_{m+1} may change!)

We shall also need:

Lemma 41.5. *Let X be an affine space and let $C \subseteq X$ be a convex set which satisfies the no-line property, and write d_C for the Hilbert metric on C. Let $D \subseteq C$ be a compact convex set with a finite number of extreme points, $\{e_1 \ldots, e_k\}$.*

Assume that $d_C(x, y) \leq \varepsilon$ for all pairs of extreme points. Then the diameter of D is bounded above by $2(k-1)\varepsilon$.

Proof. We will be done if we show that for $p \in D$, then $d_C(p, e_i) \leq (k-1)\varepsilon$ for each extreme point. The proof is by induction on k. Consider $k = 2$; then since p is on the segment bewteen e_1 and e_2, $d_C(p, e_i) \leq \varepsilon$. Now for three points e_1, e_2, e_3, consider $d_C(e_1, p)$; let $\tilde{p}$ be the point on the $(e_2, e_3)-$segment where the line through e_1 and p meets it; then $d_C(e_1, p) \leq d_C(e_1, \tilde{p})$ and by the previous step, $d_C(\tilde{p}, e_2) \leq \varepsilon$, and since also $d_C(e_1, e_2) \leq \varepsilon$ we have that $d_C(e_1, p) \leq 2\varepsilon$.

For $k = 4$, we project p to $\tilde{p}$ in the simplex generated by e_2, e_3, e_4 and apply the previous induction step. Continuing in this manner we get the claimed bound. □

Now we consider mixing for Shannon-Parry type measures. Let $\underline{L}$ be a sequence of $(l_i \times l_{i+1})$ $0-1$ matrices, with either index $i \geq 0$ or $i \in \mathbb{Z}$.

Lemma 41.6. *The following properties are equivalent:*
(a) For some choice of $\underline{\mathbf{w}} \in \Omega_{\underline{L}}$ and $\underline{\mathbf{v}}^t \in \Omega^t_{\underline{L}}$ the nonstationary Markov chain $(\Sigma^+_L, \mu_{\underline{\mathbf{w}},\underline{\mathbf{v}}})$ is mixing. For $\underline{P} = \underline{P}_{\underline{\mathbf{w}},\underline{\mathbf{v}}}$:
(b) The sequence (P_i) is future Perron-Frobenius.
(c) The sequence (P_i) is future focussing.
(d) The sequence (L_i) is future Perron-Frobenius.
(e) The sequence (L_i) is future focussing.
(f) The chain $(\Sigma^+_L, \mu_{\underline{\mathbf{w}},\underline{\mathbf{v}}})$ is future mixing for any choice $\underline{\mathbf{w}}$, $\underline{\mathbf{v}}$.

Proof. We have $(b \Longleftrightarrow c)$ and $(d \Longleftrightarrow e)$ from Lemma 41.4.
$(b \Longleftrightarrow d)$:We fix a choice of $\underline{\mathbf{w}}$, $\underline{\mathbf{v}}$ and write $\underline{P} = \underline{P}_{\underline{\mathbf{w}},\underline{\mathbf{v}}}$. We have

$$P^{(k,m)} = (1/\lambda^{k,m}) W_k^{-1} L^{(k,m)} W_{m+1}.$$

Now we compare $P^{(k,m)}\Delta_{m+1}$ and $L^{(k,m)}\Delta_{m+1}$; since W_{m+1} is diagonal, it is a bijection on the simplex, so $W_{m+1}\Delta_{m+1} = \Delta_{m+1}$, and this does not affect the image; on the other hand, W_k^{-1} is an isometry of Δ_k by Corollary 23.23. Therefore the projective diameters of $P^{(k,m)}\Delta_{m+1}$ and $L^{(k,m)}\Delta_{m+1}$ are equal. (A similar reasoning gives a direct proof of $(c \Longleftrightarrow e)$.)
$(c \implies f)$: Here we use for (f) the equivalent condition of Lemma 41.3. Assuming (c), we know the projective diameter of $\Delta_k^t P^{(k,m)}$ is $\leq \varepsilon$ for m large. Now let $\boldsymbol{e}_i^t$ denote the row vector with 1 in the i^{th} coordinate, 0 elsewhere; then $\boldsymbol{e}_i^t P^{(k,m)}$ is the i^{th} row of $P^{(k,m)}$. Hence all the rows of this matrix are within distance ε. And $\Delta_0^t P^{(0,k-1)} \subseteq \Delta_k^t$.

Hence the vector $\boldsymbol{\pi}_m^t$ (the unique row of $Q_{\boldsymbol{\pi}_m^t}^{(k,m)}$) is in this set also, and so

$$d(P^{(k,m)}, Q_{\boldsymbol{\pi}_m^t}^{(k,m)}) < \varepsilon,$$

as claimed.
$(a \implies c)$: Assume each row of $P^{(k,m)}$ is $\varepsilon-$close to the unique row of $Q_{\boldsymbol{\pi}_m^t}^{(k,m)}$; then by Lemma 41.5 the diameter of the image $\Delta_k^t P^{(k,m)}$ is less than $2(l_k+1)\varepsilon$, since by Lemma 16.3 the number of extreme points in the image simplex is bounded by the number in Δ_k^t; this proves (c). Finally, note that $(f \implies a)$ a fortiori. $(a \iff f)$: Since statements (d) and (e) do not refer to the choice of $\underline{\mathbf{w}}$ and $\underline{\mathbf{v}}$,..... □

Remark 41.2. For two-sided sequences, the corresponding past statements are all equivalent.

41.4. **A nonstationary Bowen-Marcus lemma.** We now extend Lemma 18.1 to the nonstationary situation.

Lemma 41.7. *Given a sequence $l_i \geq 1$, if L_i for $i \geq 0$ is a sequence of $(l_i \times l_{i+1})$ $0-1$ matrices which are future Perron-Frobenius, then for a measure m on $\Sigma_{\underline{L}}^+$ which has the Bowen-Marcus property, m is a constant multiple of ν.*

Proof. We follow nearly line-for-line the proof for the stationary case, Lemma 18.1. The first main change is this: we choose and fix $\underline{\mathbf{w}} \in \Omega_{\underline{L}}$ and $\underline{\mathbf{v}}^t \in \Omega_{\underline{L}}^t$; this defines the measures $\mu_{\underline{\mathbf{w}},\underline{\mathbf{v}}}$ and $\nu_{\underline{\mathbf{w}},\underline{\mathbf{v}}}$. Then we define $\gamma_{t,s}$ by

$$m([* * \cdots * s]) = \gamma_{t,s}\nu([* * \cdots * s])$$

as before.

From Lemma 41.6, the future Perron-Frobenius condition implies that the sequence is future focussing which implies the measure $\mu_{\underline{\mathbf{w}},\underline{\mathbf{v}}}$ is future mixing, so from equation (161), we have for t sufficiently large, for every $a \in \mathcal{A}_0$, $s \in \mathcal{A}_t$

$$\nu([a * \cdots * s]) = \mu([a * \cdots * s])/\mathbf{v}_a = (1 \pm \varepsilon)\nu[a] \cdot \mu[* * \cdots * s];$$

and so

$$m[a] = \sum_{s \in \mathcal{A}_t} m([a * \cdots * s]) = \sum_{s \in \mathcal{A}_t} \gamma_{t,s} \cdot \nu([a * \cdots * s]) = (1 \pm \varepsilon) \sum_{s \in \mathcal{A}_t} \gamma_{t,s} \cdot \nu([a])\mu([** \cdots * s])$$

hence for a different ε',

$$(1 \pm \varepsilon')m[a]/\nu[a] = \sum_{s \in \mathcal{A}_t} \gamma_{t,s} \cdot \mu([* * \cdots * s]).$$

This holds for each $t \geq 0$ sufficiently large.

Therefore the limit

$$\lim_{t \to \infty} \sum_{s \in \mathcal{A}_t} \gamma_{t,s} \cdot \mu([* * \cdots * s])$$

exists, and this is our constant γ. □

41.5. **Nonstationary minimality and unique ergodicity.**

Definition 41.3. We recall the a continuous map on a topological space is **minimal** iff every orbit is dense. See e.g. [Wal75], [Fur81].

We shall say a sequence $\underline{M}$ of nonnegative $(l_i \times l_{i+1})$ matrices is **past primitive** if for any $k \in \mathbb{Z}$, there exists $m > 0$ (depending on k) such that $M_k^{k+m} \equiv M_k M_{k+1} \dots M_{k+m}$ has entries all nonzero. It is **future primitive** if M for any $k \in \mathbb{Z}$, there exists $m > 0$ such that $M_{k-m}^k > 0$.

In the nonstationary situation, an interesting new phenomenon arises, that *primitive does not necessarily imply Perron-Frobenius*. See §§42.2, 42.4 below.

Here is the main result towards which we have been heading:

Theorem 41.8. (Minimality and unique ergodicity for nonstationary adic transformations) Let L_i for $i \geq 0$ be a sequence of $(l_i \times l_{i+1})$ $0-1$ matrices, and let $\mathcal{O}$ be a one-dependent symbol order. Then:
(i) If the sequence (L_i) is future-primitive, then the adic transformation $(\Sigma_{\underline{L}}^+, T_{\mathcal{O}})$ is minimal.
(ii) The adic transformation is uniquely ergodic if and only if this sequence is future Perron-Frobenius, and then and $m = \nu/\nu(\Sigma_{\underline{L}}^+)$ is the unique $T_{\mathcal{O}}-$invariant non-atomic measure on Σ_A^+.

Proof.
(ii): Assume future Perron-Frobenius; then unique ergodicity follows from Lemma 41.7 just as for the stationary case, see the proof of Theorem 18.3.

For the converse, if the future Perron-Frobenius property does not hold, then there are at least two right eigenvector sequences $\underline{\mathbf{w}} \in \Omega_{\underline{L}}$, and we have at least two measures $\nu = \nu_{\mathbf{w}}$; see §41.2.
(i): We are to show that every orbit is dense, so, given a point $x = (x_0 x_1 \dots) \in \Sigma_{\underline{L}}$, and a thin cylinder set $B = [b_0 b_1 \dots b_k] \in \mathbb{C}_0^k$ for some k, we are to show that there exists $n \in \mathbb{Z}$ with $T_{\mathcal{O}}^n(x) \in B$. Since $\underline{L}$ is future primitive, there exists m such that all the entries of $L_k \cdots L_m$ are > 0. Thus there exists an allowed string $(b_0 b_1 \dots b_k w_{k+1} \dots w_m)$ such that $w_m = x_m$.

Therefore the infinite string $w = (b_0 b_1 \dots b_k w_{k+1} \dots w_{m-1} x_m x_{m+1} \dots)$ is allowed, and hence the two points x and w are comparable with respect to the order, so either $x \leq w$ or $w \leq x$. In the first case, there exists $n \geq 0$ with $T_{\mathcal{O}}^n(x) = w$, in the second case n is ≤ 0. □

More generally we have:

Proposition 41.9. *Given the sequence $\underline{L}$ and a one-dependent symbol order, the set of measures $\{\nu(\mathbf{w}_0)\}$ is a convex compact set and its extreme points are the ergodic measures for the adic transformation, with cardinality at most* $\inf_{i \geq 0}(l_i)$.

Proof. From Lemma 41.1, the set of invariant measures is indexed by the simplex Δ_∞. That the extreme points of the collection of invariant measures for a transformation are the ergodic measures is a well-known fact; see e.g. p. 38 of [Bil65]. □

As for the stationary case we have:

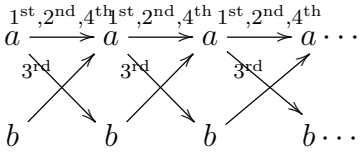


FIGURE 111. The edge-labelled Bratteli diagram for the Chacon adic transformation; the 4 edges entering into the symbol a are ordered a, a, b, a.

Proposition 41.10. *Let L_i for $i \geq 0$ be a sequence of $(l_i \times l_{i+1})$ $0-1$ matrices satisfying the future Perron-Frobenius property, then the action of the group of finite coordinate changes on $\Sigma^+_{\underline{L}}$ is uniquely ergodic; future primitive implies the action is minimal.* □

42. PLANNED: Nonstationary substitutions and adic transformations

42.1. **The Chacon example (with Thierry Monteill and Julien Cassaigne) (warning: this section is a rough version!)** First we consider a stationary example, coming from the well-known Chacon substitution dynamical system. This is defined from the substitution

$$a \mapsto aaba,$$

$$b \mapsto b.$$

To this we associate in the natural way a one-sided stationary ordered Bratteli diagram, Figure 111. The order is given by the substitution. Note that the diagram contains exactly the same information as the substitution. The **matrix of the substitution** or **abelianization** is in this case $L = \begin{bmatrix} 3 & 1 \\ 0 & 1 \end{bmatrix}$; the top row indicates that there are 3 arrows coming from a to a, 1 from b to a, and the bottom row that there is one arrow going from b to b. One can show (as is well known):

Proposition 42.1. *The substitution dynamical system is infinitely decodable.*

As a consequence:

Proposition 42.2. *The natural homomorphism from the adic transformation to the substitution dynamical system is one-to-one, if we remove the countable set of points in the adic which do not have complete orbits, and also a countable set in the substitution dynamical system.*

As a consequence, minimality and unique ergodicity are true for one if and only if they are true for the other. Now the matrix L is certainly not primitive. Yet the Chacon transformation is minimal, as is well known. There is however a unique positive eigendirection; this is the Perron-Frobenius condition, which by Theorem 41.8 implies unique ergodicity, for the adic and hence for the substitution (this latter statement was previously known).

42.2. **A determinant one** (3×3) **counterexample (with S. Ferenczi).** Here we give an example of a class of adic transformations which are primitive hence minimal but not uniquely ergodic, and whose matrices have determinant one. By definition and by Theorem 41.8 respectively, primitivity and unique ergodicity only depend on the matrix sequence and not on the order on the Bratteli diagram.

We define

$$L_j = \begin{pmatrix} m_j - 1 & m_j & 0 \\ n_j & n_j & n_j - 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Note that $\det L_j = 1$.

Let $\widetilde{L_j}x$ be $\frac{L_j(x)}{|L_j(x)|}$ where $|y| = \sum_{i=1}^{3} y_i$. We define $\widetilde{B_k}$ to be the product $\widetilde{L_1}\dots\widetilde{L_k}$.

Lemma 42.3. *For any $x \in \Delta$, the unit simplex in $\mathbb{R}^3$, $(\widetilde{L_j}x)_3 \le \frac{1}{n_j}$.*

Proof. We have $(L_j x)_3 = 1$ while $|L_j(x)| = n_j + (m_j+1)x_1 + m_j x_3 \ge n_j$. □

Lemma 42.4. *Suppose $n_{k+1} \ge 2m_k$ for all k, and let $e_1 = (1,0,0)$; then for $k \ge 1$, $(\widetilde{B_k}e_1)_2 \ge 1 - \frac{2}{n_1}$.*

Proof. Let $x^k = \widetilde{L_k}e_1$, $x^{k-1} = \widetilde{L}_{k-1}x^k$, …, $x^1 = \widetilde{L_1}x^2$; then, in view of Lemma 42.3, we just have to prove that $x_1^1 \le \frac{1}{n_1}$, and we shall prove by induction that for each j we have $x_1^j \le \frac{1}{n_j}$; this is true for $x^{k+1} = e_1$, and

$$x_1^j = \frac{m_j x_1^{j+1} + (m_j - 1)x_3^{j+1}}{n_j + (m_j+1)x_1^{j+1} + m_j x_3^{j+1}};$$

this last quantity will certainly be smaller than $\frac{1}{n_j}$ as soon as $m_j x_1^{j+1} + (m_j - 1)x_3^{j+1} \le 1$, and this is true because of Lemma 42.3, the condition $n_{j+1} \ge 2m_j$ and the induction hypothesis $x_1^{j+1} \le \frac{1}{n_{j+1}}$. □

Lemma 42.5. *Suppose $m_k \ge \frac{3n_k+1}{2}$ for all k, and let $e_2 = (0,1,0)$; then for $k \ge 1$, $(\widetilde{B_k}e_2)_1 \ge \frac{1}{3}$.*

Proof. We define a sequence x_i in the same way as in the previous lemma; under the induction hypothesis $x_1^{j+1} \ge \frac{1}{3}$, the expression of x_1^j being as in the previous lemma, we have just to prove that $(2m_j - 1)x_1^{j+1} \ge n_j + (3 - 2m_j)x_3^{j+1}$, which is a consequence of $m_j \ge \frac{3n_j+1}{2}$. □

Consequences: Since the product $L_i L_{i+1}$ has entries all > 0, the sequence L_i is primitive hence any adic transformation defined from it by choosing an order on the Bratteli diagram is minimal, by Theorem 41.8. Writing now Δ_k for Δ, the unit simplex in $\mathbb{R}^3$, where k serves only to indicate the time, then the image $\widetilde{B_k}(\Delta_k) \subseteq \Delta_0$ contains the points $\widetilde{B_k}(e_2)$, with first coordinate $\ge 1/3$, and $\widetilde{B_k}(e_1)$ whose second coordinate is $> 1 - 1/n_1 \ge 2/3$. so these points are in disjoint regions of the simplex for all k, and hence the image simplices cannot possibly nest down to a singleton. See Fig. ??. We have shown:

Corollary 42.6. *Any adic transformation defined by taking the sequence L_i as edge matrices, $i \geq 0$, and then fixing a one-dependent symbol order, is minimal but is not uniquely ergodic.* □

42.3. **Anosov families and the** (2×2) **case.** The next theorem shows that a (2×2) determinant one counterexample cannot exist.

Theorem 42.7. *Let $(L_i)_{i=0}^{\infty}$ for $i \geq 0$ be a sequence of (2×2) determinant one integer matrices. Then any adic transformation defined by taking the L_i as edge matrices, $i \geq 0$, and then fixing a one-dependent symbol order, and which is minimal, is then necessarily also uniquely ergodic.*

Proof. The collection of such matrices is $SL(2,\mathbb{N})$, and as is well known (see Lemma 3.11 of [AF05]) this semigroup is generated freely by the matrices $M = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ and $N = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Given a sequence $(L_i)_{i=0}^{\infty}$ in $SL(2,\mathbb{N})$, let us factor each successively to produce the new sequence $(A_j)_{j=0}^{\infty}$ with each $A_j = M$ or N; this is what is called an **additive sequence** in [AF05], [**?**]. Now this is clearly future-minimal iff A_j is not eventually always equal to M or N. And in this case, we can then take partial products to produce a **multiplicative** sequence of the form $\widehat{A}_i = \begin{bmatrix} 1 & 0 \\ n_i & 1 \end{bmatrix}$ or $\widehat{A}_i = \begin{bmatrix} 1 & n_i \\ 0 & 1 \end{bmatrix}$, for n_i positive integers. with the choice of upper or lower triangular alternating for i even or odd. Proposition 4.1 of [AF05] then shows that the sequence $(\widehat{A}_i)$ is future Perron-Frobenius; indeed, the unique positive right eigenvector is $\mathbf{w} = (c,d)$ where c/d or d/c is equal to the continued fraction $[n_0 n_1 n_2 \dots]$ depending on the parity of $\widehat{A}_0$. □

42.4. **Keane's counterexample.** Keane's construction uses a sequence of (4×4) matrices which have a combinatorics similar to that just discussed; indeed our matrices are simply sumatrices of Keane's. However, since they are larger, the combinatorics is somewhat more complicated. Our idea in the last section was twofold: to imitate his combinatorial argument in this simpler situation, thus perhaps facilitating an understanding of Keane's example; secondly, to answer the natural question (once things have been put into an adic framework) as to whether such a (3) "Keane-type counterexample" exists – one knows already that no exchange on 3 intervals will work!!!

Here are Keane's matrices:

$$L_j = \begin{pmatrix} 0 & 0 & 1 & 1 \\ m_j - 1 & m_j & 0 & 0 \\ n_j & n_j & n_j - 1 & n_j \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Again, these satisfy $\det L_j = 1$.

(to be continued.....)

42.5. **Veech and Masur's theorem.**

43. Appendix: invariant means; the harmonic projection

44. Appendix:Measure theory, functional analysis background

$\delta-$continuity property: Countable additivity is equivalent to: for $(A_n)_{n\geq 0}$ nested decreasing sets of finite measure and $A_n = \lim A_n = \inf A_n = \cap A_n$, then $\lim \mu(A_n) = \mu(A)$.

Lebesgue Dominated Convergence Theorem:

On a finite measure space, If $f_n \to f$ a.s and $|f_n| \leq g$ a.e. for $g \in L^1$ then

$$\int f_n \to \int f.$$

45. Tensor products

In this section we build on [Spi65], [Spi79], [GP74], [War71].

Definition 45.1. Given a vector space V of dimension k, with dual space denoted V^*, a $p-$**tensor** T is a function on the product of p copies of V, $T: V \times \cdots \times V \to \mathbb{R}$, which is $p-$**multilinear**, i.e. is separately linear in each coordinate. Thus a one-tensor is just a linear functional. The space of all $p-$tensors is denoted $\mathcal{T}^p(V^*)$, so $\mathcal{T}^1(V^*) = V^*$.

Given $\varphi \in \mathcal{T}^p(V^*)$ and $\psi \in \mathcal{T}^q(V^*)$, then the **tensor product** is $\varphi \otimes \psi \in \mathcal{T}^{p+q}(V^*)$ defined by

$$\varphi \otimes \psi(\mathbf{v}, \mathbf{w}) = \varphi(\mathbf{v})\psi(\mathbf{w})$$

(which is indeed multilinear). This map is not onto, but the image, denoted $\mathcal{T}^p(V^*) \otimes \mathcal{T}^q(V^*)$, spans the vector space $\mathcal{T}^{p+q}(V^*)$, a point we return to below. The tensor product clearly satisfies the distributive laws $(T_1 + T_2) \otimes S = T_1 \otimes S + T_2 \otimes S$, and $T \otimes (S_1 + S_2) = T \otimes S_1 + T \otimes S_2$, as well as the associativity of scalar multiplication, $(aT) \otimes S = aT \otimes S$, $T \otimes (aS) = aT \otimes S$. Note that the tensor product is not commutative (even if $p = q$) but it is associative, so $T_1 \otimes T_2 \otimes \cdots \otimes T_n$ is well-defined.

From these laws, $\Phi(T, S) = T \otimes S$ itself defines a bilinear function $\Phi : \mathcal{T}^p(V^*) \times \mathcal{T}^q(V^*) \to \mathcal{T}^p(V^*) \otimes \mathcal{T}^q(V^*) \subseteq \mathcal{T}^{p+q}(V^*)$, and more generally, $\Phi(T_1, T_2, \ldots, T_n) = T_1 \otimes T_2 \otimes \cdots \otimes T_n$ defines a multilinear function, though these do not fit our definition of tensor products, as the spaces are not the same.

A $p-$**multi-index** or $p-$**index sequence** I is a function $I : \{1, 2, \ldots, p\} \to \{1, 2, \ldots, k\}$. We write $I_1 I_2 \ldots I_p$ for I, and define $\mathcal{I}^p$ to be the collection of all $p-$index sequences. Note that $\#\mathcal{I}^p = k^p$.

Definition 45.2. We recall that the **symmetric group** S_p is the the group of permutations of p symbols, and is generated by the **transpositions**, those permutations which interchange two elements. The symmetric group has $p!$ elements. The function sgn : $S_p \to (\{\pm 1\}, \cdot) \cong (\mathbb{Z}_2, +)$ is a homomorphism which takes the value +1 on the **even** and value -1 on the **odd** permutations; these are, respectively, the product of an even or odd number of transpositions. The **alternating group** A_p is the subgroup of all even permutations; it is a normal subgroup of index 2, with the odd permutations being its other coset.

A $p-$tensor is **symmetric** iff the value is unchanged whenever any permutation is applied to the vectors, equivalently iff this holds for any transposition. It is **antisymmetric**, also called *skew-symmetric*, iff the sign changes whenever an odd permutation is applied, again equivalently iff true for any transposition; thus, iff for every such pair $\mathbf{v}, \mathbf{z}$, we have $T(\mathbf{w}_1, \ldots, \mathbf{v}, \ldots, \mathbf{z}, \ldots, \mathbf{w}_p) = -T(\mathbf{w}_1, \ldots, \mathbf{z}, \ldots, \mathbf{v}, \ldots, \mathbf{w}_p)$. A $p-$tensor is **alternating** iff whenever two vectors of $\mathbf{v}_I$ for $I \in \mathcal{I}^p$ happen to be the same, then the value is zero. It is clear that this is equivalent to being antisymmetric, in our setting of vector spaces over the field $\mathbb{R}$ (see below regarding the one exceptional case, the field $\mathbb{Z}_2$!)

The alternating $p-$tensors will be of particular interest; the collection of these is a vector subspace, denoted $\Lambda^p(V^*) \subseteq \mathcal{T}^p(V^*)$.

Choose now a basis $\mathbf{v}_1 \ldots, \mathbf{v}_k$ for V, and define $\varphi_1 \ldots, \varphi_k$ to be the dual basis: the linear functionals satisfying $\varphi_i(\mathbf{v}_j) = \delta_{ij} = 0$ or 1 iff $i \neq j$, $i = j$ respectively. Given a $p-$index sequence I, we write $\mathbf{v}_I \equiv (\mathbf{v}_{I_1}, \ldots, \mathbf{v}_{I_p})$ and $\varphi_I = \varphi_{I_1 I_2 \ldots I_p} \equiv \varphi_{I_1} \otimes \cdots \otimes \varphi_{I_p}$.

Proposition 45.1. *dim*$(\mathcal{T}^p(V^*)) = k^p$*, with basis* $\{\varphi_I\}_{I \in \mathcal{I}^p}$.

Before giving the proof, we consider the most basic examples: an inner product, and the determinant.

Using the standard basis for $\mathbb{R}^n$, with $\langle \mathbf{v}, \mathbf{w} \rangle$ the standard inner product, then for $T(\mathbf{v}, \mathbf{w}) \equiv \langle \mathbf{v}, \mathbf{w} \rangle$, T is a $2-$tensor. We note that $T = \varphi_{11} + \varphi_{22} + \cdots + \varphi_{nn}$; this is a symmetric $d-$tensor.

Next, let $\mathbf{v}_1 \ldots, \mathbf{v}_n$ be the columns of an $(n \times n)$ matrix. Then the determinant is an alternating $d-$tensor, since it is multilinear and switching any two vectors changes the sign. For $k = 2$, with $\mathbf{v} = (a, b)$ and $\mathbf{w} = (c, d)$, then for

$$S(\mathbf{v}, \mathbf{w}) = \det \begin{bmatrix} a & c \\ b & d \end{bmatrix} = ad - bc$$

we have $S = \varphi_1 \otimes \varphi_2 - \varphi_2 \otimes \varphi_1 = \varphi_{12} - \varphi_{21}$.

For $k = 3$, expanding along the top row, the determinant is $S = \varphi_1 \otimes (\varphi_{23} - \varphi_{32}) - \varphi_2 \otimes (\varphi_{13} - \varphi_{31}) + \varphi_3 \otimes (\varphi_{12} - \varphi_{21}) = \varphi_{123} - \varphi_{132} - \varphi_{213} + \varphi_{231} + \varphi_{312} - \varphi_{321}$.

Noting that these basis elements are indexed by the six permutations of three letters, this hints at the formula for the determinant for arbitrary k. Indeed, starting with the single basis element φ_{123}, we have applied all the elements of the permutation group S_3, with coefficient equal to the sign of the permutation. And this furthermore suggests a way to bulid an alternating tensor, perhaps with $p \neq n$, beginning with a single tensor. We encounter exactly this general procedure below: the **Alt** operator.

Unifying these two examples we get the following.

Proposition 45.2. *Each* $2-$*tensor* T *can be represented as a* $(k \times k)$ *matrix* A *via*

$$T(\mathbf{v}, \mathbf{w}) = \mathbf{v}^t A \mathbf{w}$$

where

$$T = \sum A_{ij} \varphi_{ij}.$$

T *is respectively symmetric or antisymmetric iff that holds for its representing matrix* A*, i.e. iff* $A = A^t$*, respectively* $A = -A^t$*.* T *is an inner product iff* A *is symmetric and invertible.*

Proof. The first statement uses Proposition 45.1 and the second is immediate. The third comes from Lemma 35.51.

□

Note that the dimension of the collection of $(k \times k)$ matrices agrees with Proposition 45.1: it is k^2.

Remark 45.1. All these definitions make sense with the field $\mathbb{R}$ replaced by any field, and more generally in the setting of modules over a ring. There is one small difference: an alternating tensor is always antisymmetric, but if we have a vector space over a field of characteristic 2, then the converse may not hold. Such fields are unusual: every element has the property that $a = -a$. Thus for example, with the standard inner product $(1,1) \cdot (1,1) = 1 + 1 = 0$. The inner product is a tensor product T with matrix representation given by I. Taking $\mathbf{v} = (1,0)$, then $T(\mathbf{v}, \mathbf{v}) = 1 \neq 0$ (so it cannot be alternating) yet $T(\mathbf{v}, \mathbf{v}) = -T(\mathbf{v}, \mathbf{v})$ and indeed that hols for any vector $\mathbf{w} = (a, b)$, $\mathbf{w} \cdot \mathbf{w} = a^2 + b^2 = -(a^2 + b^2)$ so T is antisymmetric.

Another way to see this is that the tensor is alternating iff $A = -A^t$ and the diagonal entries are 0. But for this field

gives the inner product tensor T, and for

Now that we have this matrix representation for 2−tensors, we can see by example why the tensor product is not onto. Let $V = \mathbb{R}^2$. Since $\mathcal{T}^1(V^*) = V^* = \mathbb{R}^2$, consider $T = (a, b) = a\varphi_1 + b\varphi_2$ and $S = (c, d) = c\varphi_1 + d\varphi_2$, so $T \otimes S = ac\varphi_{11} + ad\varphi_{12} + bc\varphi_{21} + bd\varphi_{22}$ which corresponds to the matrix

$$\begin{bmatrix} ac & ad \\ bc & bd \end{bmatrix}$$

But not all (2×2) matrices are of this form: if say $bc = 0$ then one of b, c is zero, say b; but then the whole second row is zero. So in fact neither the standard inner product nor the determinant is of this form; they correspond respectively to the matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

If, however we take linear combinations of products we get everything, as the Proposition shows.

We remark that this same example shows that not all measures on a product space are product measures; here the space is a two-point space with point masses of weights a, b and c, d.

But that is not surprising, as we have this infinite dimensional example of tensor product:

Example 63. Let $V = \mathcal{C}(I, \mathbb{R})$ (the space of continuous functions on the unit interval); so the dual space V^* is the space $\mathcal{M}(I)$ of finite signed countably additive measures on I. Then for $\mu, \nu \in \mathcal{M}$, $\mu \otimes \nu$ is the product measure $\mu \times \nu$.

Exercise: show that the while the product measures are not all of $\mathcal{M}(I \times I)$, their linear combinations form a dense subset.

Now we return to prove Proposition 45.1.

Proof. First we show $\{\varphi_I\}_{I\in\mathcal{I}^p}$ is a basis for $\mathcal{T}^p(V^*)$. To show it spans, we claim that for

$$S = \sum_{I\in\mathcal{I}} T(\mathbf{v}_I)\varphi_I$$

then $S = T$. Note first that $\varphi_I(\mathbf{v}_J) = \delta_{IJ}$. Hence $S(\mathbf{v}_J) = T(\mathbf{v}_J)$ for all $\mathbf{v}_J \in \mathcal{I}$. But by multilinearity, knowing T on all the $\{\mathbf{v}_J\}_{J\in\mathcal{I}}$ determines it on $V \times \cdots \times V$ (p times), since we have a basis for each factor individually, so $S = T$.

Next we show $\{\varphi_I\}_{I\in\mathcal{I}^p}$ is linearly independent. Suppose

$$S = \sum_{I\in\mathcal{I}} a_I\varphi_I = 0.$$

Applying this to $\mathbf{v}_J$ gives $S(\mathbf{v}_J) = a_J = 0$. Thus it is a basis.

The dimension statement is now clear since as noted above, $\#\mathcal{I}^p = k^p$.

□

Exercise 45.1. Show that any $2-$tensor T is the sum of a symmetric and an antisymmetric tensor.

Hint: given an $(n \times n)$ matrix A, consider $A + A^t$ and $A - A^t$.

Question: is this true for $p-$tensors?

As a hint for the last exercise, consider the following question: starting from $T \in \mathcal{T}^p(V^*)$, how can we produce a symmetric or an antisymmetric tensor?

Note that S_p acts on $\mathcal{T}^p(V^*)$ as follows: $(\sigma T)(\mathbf{w}_1, \ldots, \mathbf{w}_p) = T(\mathbf{w}_{\sigma(1)}, \ldots, \mathbf{w}_{\sigma(p)})$. So simply averaging over the group action, $\overline{T} \equiv \frac{1}{p!}\sum_{\sigma\in S_p} \sigma(T) \in \mathcal{T}^p(V^*)$ will be symmetric. Dividing by $\#S_p = p!$ assures that for T already symmetric then $\overline{T} = T$.

For an antisymmetric version of this construction, we define $\mathrm{Alt}(T) \in \mathcal{T}^p(V^*)$ by

$$\mathrm{Alt}(T) \equiv \frac{1}{p!}\sum_{\sigma\in S_p} \mathrm{sgn}(\sigma)(\sigma T).$$

It is clear that this is an alternating tensor. In fact we encountered this formula above for the determinant, except now we are normalizing by $\#S_p$; for T already alternating then $\mathrm{Alt}(T) = T$. We note that for T symmetric, then $\mathrm{Alt}(T) = 0$.

The main reason for considering this operator is to then define the **wedge product**. Given tensors T, S this is: $T \wedge S \equiv \mathrm{Alt}(T \otimes S)$.

In particular, for T, S alternating, $T \wedge S$ defines a map from $\Lambda^p(V^*) \times \Lambda^q(V^*)$ to $\Lambda^{p+q}(V^*)$.

More generally, we define $T_1 \wedge \cdots \wedge T_n \equiv \mathrm{Alt}(T_1 \otimes \cdots \otimes T_n)$. The most basic properties of the wedge product, the distributive law and linearity, are easy to verify. But the problem with this definition is that we don't know if it agrees with what has been done before. That is, we have $(T \wedge S) \wedge R = \mathrm{Alt}(\mathrm{Alt}(T \wedge S) \otimes R)$ but is this really equal to $T \wedge S \wedge R \equiv \mathrm{Alt}(T \otimes S \otimes R)$?

What is needed is:

Proposition 45.3. *The wedge product is associative:* $(T \wedge S) \wedge R = T \wedge (S \wedge R)$.

Our proof follows [Spi65] [Spi79] or [GP74]. First we need:

Lemma 45.4. *If T is symmetric, then $T \wedge S = 0$ and similarly for $S \wedge T$.*

Proof. ... □

Proof. (of Proposition).....

We know the tensor product is associative, but the difficulty when the Alt operator is applied is that several subgroups of $G = S_{p+q+r}$ are involved: those permutation which only affect the first p, middle q and final r coordinates. These are subgroups G_p, G_q, and G_r of G which are isomorphic to S_p, S_q, and S_r respectively. Also involved are G_{p+q} and G_{q+r}, defined similarly.

We know $(T \otimes S) \otimes R = T \otimes (S \otimes R)$ so if we can show $(T \wedge S) \wedge R = \text{Alt}((T \otimes S) \otimes R)$ and $T \wedge (S \wedge R) = \text{Alt}(T \otimes (S \otimes R))$ and we'll be done.

Now $(T \wedge S) \wedge R \equiv \text{Alt}((T \wedge S) \otimes R)$. .....

□

Note that this is exactly how we produced the determinant, since e.g. for $n = 3$, we started with $\varphi_{123} = \varphi_1 \otimes \varphi_2 \otimes \varphi_3$ (and one-tensors are trivially alternating).

In particular, if $\varphi, \psi \in V^*$, then $\varphi \wedge \psi = \frac{1}{2}(\varphi \otimes \psi - \psi \otimes \varphi)$. Therfore, $\varphi \wedge \psi = -\psi \wedge \varphi$ and $\varphi \wedge \varphi = 0$. This property of the wedge product is called **skew-symmetry** (or **anticommutativity** or **antisymmetry**). More generally, if $\varphi \in \Lambda^p(V^*), \psi \in \Lambda^q(V^*)$, then $\varphi \wedge \psi = (-1)^{p+q} \psi \wedge \varphi$. But on what space is this product defined? On the **exterior algebra** $\Lambda(V^*) = \Lambda(V^*) \equiv \Lambda^0(V^*) \oplus \ldots \Lambda^p(V^*) \oplus \ldots$; here one defines $\Lambda^0(V^*) = \mathbb{R}$. Now the next statement shows this is finite dimensional, as indeed:

Proposition 45.5.
(i)For V with dimension d, $dim(\Lambda^p(V^*)) = \binom{k}{p}$, with basis $\{\varphi_I\}_{I=i_1 \ldots i_p : i_1 \leq \cdots \leq i_p}$.
(ii) $\Lambda(V^*) = \Lambda^0(V^*) \oplus \cdots \oplus \Lambda^n(V^*)$; $dim(\Lambda(V^*)) = 2^n$.

Proof. Recalling that

$$\binom{n}{p} \equiv \frac{n!}{p!(n-p)!} = \frac{n(n-1)\cdots(n-p+1)}{p!},$$

"d choose p", is the number of ways of choosing p objects from n without order, the dimension statements follows once we show this is a basis. Now $\{\varphi_I\}_{I \in \mathcal{I}^p}$ span since this is a subspace of the tensor product. And by anticommutativity, any change of order is redundant, and any repetition gives the zero vector.

□

It is important to note how linear maps act: let $A : V \to W$ be linear; then the transpose map $A^* : W^* \to V^*$ is defined on $\varphi \in W^*$ by $(A^*\varphi)(\mathbf{v}) = \varphi(A\mathbf{v})$. This extends coordinatewise to $A^* : \Lambda^p(W^*) \to \Lambda^p(V^*)$:

$$(A^*T)(\mathbf{v}_1 \ldots \mathbf{v}_p) = T(A(\mathbf{v}_1) \ldots A(\mathbf{v}_p)).$$

It therefore extends to Λ, where we have:

$$A^*(T \wedge S) = A^*T \wedge A^*S.$$

We have the important special case where $p = n$ and since $\dim\Lambda^n(V^*) = \binom{n}{n} = 1$, the linear map A^* must be multiplication by a constant. Expressing A as an $(n \times n)$ matrix, then indeed, A^* is the number $\det A$.

45.1. **Tensor product of vector spaces: linearity and the Universal Mapping Property.** Two points are raised by the above treatment of multilinear algebra. First, although it is a natural path to take didactically, given our familiarity with inner products, with the determinant, and also with differential forms, it may seem curious in retrospect to have defined tensor products for the dual space rather than for the vector space itself. Secondly, now that we understand multilinear maps on V, isn't it possible that they can somehow be viewed as linear maps, on a different space? In particular, this might clarify our understanding of the action of a linear map, just given.

The answer to both queries turns out to be the same. Indeed, one can begin by defining the tensor product of vectors, and can then see that a multilinear map is exactly a linear map on this new (and much larger) vector space.

To define the tensor product on V, the easiest approach given what we have already done (the definition for the dual space V^*) is to note that since V is naturally isomorphic to $(V^*)^*$, we can just apply the previous definition.

But the cleanest method from an abstract point of view is to build the multilinearity, and later (for wedge products) the anticommutativity, directly into the vector space structure, in just the same way one defines a group via generators and relations.

Thus, let $F(V,W)$ denote the vector space with basis all of the nonzero elements of $V \times W$, and define $R(V,W)$ to be the subspace generated by elements of the form:

$$(\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}) - (\mathbf{v}_1, \mathbf{w}) - (\mathbf{v}_2, \mathbf{w})$$
$$(a\mathbf{v}, \mathbf{w}) - a(\mathbf{v}, \mathbf{w})$$

for all $a \in \mathbb{R}$, $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v} \in V, \mathbf{w} \in W$ and similarly for the second coordinate. Then we define $V \otimes W$ to be the quotient vector space $F(V,W)/R(V,W)$, with an element $\mathbf{v} \otimes \mathbf{w}$ being the corresponding coset [War71] p. 54. This yields:

$$(\mathbf{v}_1 + \mathbf{v}_2) \otimes \mathbf{w} = \mathbf{v}_1 \otimes \mathbf{w} + \mathbf{v}_2 \otimes \mathbf{w}$$
$$(a\mathbf{v}) \otimes \mathbf{w} = a(\mathbf{v} \otimes \mathbf{w})$$

and similarly in the second coordinate.

We have:

Proposition 45.6. *The dual space of $V \otimes W$ is isomorphic to $\mathcal{T}^1(V^*) \otimes \mathcal{T}^1(W^*)$. The two definitions of $V \otimes W$ agree. The similar statement holds for $p-$tensors.*

Proof. With $\Phi : V \times W \to V \otimes W$ denoting the map $\mathbf{v} \times \mathbf{w} \to \mathbf{v} \otimes \mathbf{w}$, which is bilinear, then given a vector space Z, and a bilinear map $\varphi : V \times W \to Z$, there is a unique **linear** map $\widehat{\varphi}$ such that the map φ lifts to it: $\widehat{\varphi} \circ \Phi = \varphi$. See 2.2(a) of [War71] p. 55. This is the **Universal Mapping Property** of tensor products. □

Tensor algebra of V.....

The operator **Alt** and hence the wedge product also makes sense for vectors: we write $\otimes \mathbf{v}_I = \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_p$ and then note that S_p acts on $V \otimes \cdots \otimes V$ (p times) by: $\sigma(\otimes \mathbf{v}_I) = \sigma(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_p) = \mathbf{v}_{\sigma(1)} \otimes \cdots \otimes \mathbf{v}_{\sigma(p)}$ and then we define

$$\mathrm{Alt}(T) \equiv \frac{1}{p!} \sum_{\sigma \in S_p} \mathrm{sgn}(\sigma)(\sigma \mathbf{v}_I).$$

So then $\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_p \equiv \mathrm{Alt}(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_n)$. We define $\Lambda_p(V) = V \wedge \cdots \wedge V$ to be the vector space generated by all of these elements. ???first via ideal; also V times W

The **exterior algebra** of V is now $\Lambda(V) = \Lambda^0(V) \oplus \ldots \Lambda^p(V) \oplus \ldots$.

As before for V^* we have:

Proposition 45.7.
(i)For V with dimension n, $dim(\Lambda^p(V)) = \binom{k}{p}$, with basis $\{\otimes \mathbf{v}_I\}_{I=i_1 \ldots i_p : i_1 \leq \cdots \leq i_p}$.
(ii) $\Lambda(V) = \Lambda^0(V) \oplus \cdots \oplus \Lambda^n(V)$; $dim(\Lambda(V)) = 2^n$.

Next we check how linear maps act on $V \otimes V$. Let $A : V \to W$ be linear, then we define $A(\mathbf{v}_1 \otimes \mathbf{v}_2) = (A\mathbf{v}_1) \otimes (A\mathbf{v}_2)$. This restricts to $\Lambda(V)$, where we have: $A(\mathbf{v}_1 \wedge \mathbf{v}_2) = (A\mathbf{v}_1) \wedge (A\mathbf{v}_2)$.

Proposition 45.8. *Let $A : V \to V$ be a linear map, and suppose that we are given vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ with $A\mathbf{v}_1 = \lambda_1 \mathbf{v}_1, \ldots, A\mathbf{v}_k = \lambda_k \mathbf{v}_k$. Then $\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_k$ is an eigenvector for $A : V \otimes \cdots \otimes V$, with eigenvalue $\lambda = \lambda_1 \cdots \lambda_k$.*
, , with generalized eigenvalues $\lambda_1 \ldots \lambda_n$. Then

Proof. For every pair $\mathbf{v}_1, \mathbf{v}_2$,
???
?? generalized eigenvectors?? □

$$(A^*(\varphi_1 \otimes \varphi_2))(\mathbf{v}_1, \mathbf{v}_2) = (\varphi_1 \otimes \varphi_2)(A\mathbf{v}_1, A\mathbf{v}_2) = \varphi_1(A\mathbf{v}_1)\varphi_2(A\mathbf{v}_2) = \lambda_1\lambda_2\varphi_1 \otimes \varphi_2(\mathbf{v}_1, \mathbf{v}_2),$$

$$(A^*(\varphi_1 \otimes \varphi_2))(\mathbf{v}_1, \mathbf{v}_2) = (\varphi_1 \otimes \varphi_2)(A\mathbf{v}_1, A\mathbf{v}_2) = \varphi_1(A\mathbf{v}_1)\varphi_2(A\mathbf{v}_2) = \lambda_1\lambda_2\varphi_1 \otimes \varphi_2(\mathbf{v}_1, \mathbf{v}_2),$$

so

$$(A^*(\varphi_1 \otimes \varphi_2)) = \lambda_1\lambda_2\varphi_1 \otimes \varphi_2.$$

Now this passes to the action of A on $V \otimes \cdots \otimes V$, by duality.

45.2. **Homology and cohomology.** Our references here are [War71], [Mas91], [GP74], [Spi65], [Spi79], [Hat02], [BT82]. For V a vector space of dimension n, as above $\Lambda^k(V^*)$ denotes the alternating $k-$ tensors and $\Lambda(V^*)$ the exterior algebra, $\Lambda(V^*) = \Lambda^0(V^*) \oplus \cdots \oplus \Lambda^n(V^*)$. Now we consider M a compact differentiable manifold of dimension n. We write $\Lambda^k(M)$ for the bundle of alternating tensors, so $\Lambda^k(M) = \cup_{p \in M} \Lambda^k((TM_p)^*)$ and similarly for $\Lambda(M)$.

We consider the smooth sections of these bundles, denoting these by $\Omega^k(M)$ and $\Omega(M)$.

Definition 45.3. A (differential) $k-$ *form* on M is an element of $\Lambda^k(M)$.

Consider a $\mathbb{C}^\infty$ function $f : M \to \mathbb{R}$; taking $M = \mathbb{R}^n$, then e.g. $dx_1 \in \Lambda^1(M) = (\mathbb{R}^n)^*$, $f(x)dx_1$ gives an example of a differential form. A general differential form can be written in coordinates as $\Sigma_I f_I dx_I$ where the sum is over multi-indices.

The derivative map in this context is given by the operator

$$d : \Lambda^k(M) \to \Lambda^{k+1}(M)$$

defined by:

Bott Tu p 13

The p^{th} de Rahm cohomology group $H^p_{deR}(M)$ is the closed $p-$forms modulo the exact ones. Writing $H^*_{deR}(M) = H^0(M) \oplus \dots H^n(M)$, the wedge product passses to this **graded algebra**. (A graded algebra simply means an algebra which is a direct sum of levels like this, with a product as here, such that an element of level p times one of level q belongs to level $p+q$).

We need:

–de Rahm's theorem: $H^p_{deR}(M)$ is isomorphic to $H^p(M, \mathbb{R})$, the singular cubical cohomology group with real coefficients; moreover, this respects the d operators, so it is an isomorphism of the "cohomology theories", i.e. there is an isomorphism of the commutative diagrams. This correspondence is via integration of forms against singular chains;

–this isomorphism is an isomorphism of the $\mathbb{R}-$modules, and moreover it takes the wedge product to the cup product, giving an algebra isomorphism.

–Künneth Formula: [BT82] p. 47: $H^*(M \times N) = H^*(M) \otimes H^*(N)$, that is, for each $0 \leq m \leq n$, $H^m(M \times N) = \oplus_{p+q=m} H^p(M) \otimes H^q(N)$;

–Poincaré Duality: see Massey [Mas91], p. 365: for any abelian group G, $H^p(M, G) \cong H_{n-p}(M, G)$.

Now let M be the $n-$torus, $\mathbb{R}^n/\mathbb{Z}^n$.

Theorem 45.9. *For $1 \leq p \leq n$, $H^p(M, \mathbb{R})$ is isomorphic to $\Lambda^p(\mathbb{R}^*)$, a real vector space of dimension $dim(\Lambda^p(V^*)) = \binom{k}{p}$, with basis $\{\varphi_I\}_{I=i_1 \dots i_p : i_1 \leq \dots \leq i_p}$.*
(ii) $\Lambda(V^*) = \Lambda^0(V^*) \oplus \dots \oplus \Lambda^n(V^*)$; $dim(\Lambda(V^*)) = 2^n$.

Proof. By Proposition 45.7, □

M^r is conjugate to the transpose of $(M^k)^{-1}$ for $k = n - r$. Functoriality of intersection form; wedge product and duality.

46. Semidirect products and skew products

Definition 46.1. Let G be a group. The identity element of a group is denoted by $\mathbf{e}$, and the identity subgroup $\{\mathbf{e}\}$ is denoted I. That H is a subgroup of G is written $H < G$. Given $H, K < G$ then $HK = \{hk : h \in H, K \in K\}$. Similarly, one defines $gH = \{gh : h \in H\}$; this is the collection of **left cosets** of H, denoted G/H, while $H\backslash G$ are the **right cosets** $\{Hg\}$.

We say G is a **product** of H and K iff
(i) $G = HK$ and
(ii) $H \cap K = \{1\}$.

This is equivalent to:
(i') for every $g \in G$ there exist $h \in H$, $k \in K$ such that $g = hk$;
(ii') this expression is unique.

By taking inverses, the order is not important: G is a product of H and K iff it is a product of K and H.

Recall that an **inner automorphism** of G is a map $\varphi_g : G \to G$ for some $g \in G$ defined by $h \mapsto ghg^{-1}$.

A subgroup $K < G$ is **normal**, written $K \lhd G$, iff

(a) for all $g \in G$, $gKg^{-1} = K$, equivalently $gK = Kg$;
iff:
(b) for all $g \in G$, for all $k \in K$, there exists $\widetilde{k} \in K$ such that $gk = \widetilde{k}g$, equivalently $gkg^{-1} = \widetilde{k}$.

That is, K is normal iff the inner automorphism φ_g of G defines an automorphism of K (which will be an inner automorphism of K itself iff $g \in K$).

(This gives lots of examples of automorphisms which are not inner!)

A third definition is:
(c) The left cosets G/H form a group, with the multiplication $(g_1H)(g_2H)$ just the multiplication of these sets.

Exercise 46.1. Check the equivalence of $(a), (b)$ and (c).

Definition 46.2. We say G is a **(internal) direct product** of H, K iff G is a product and H and K commute, i.e. iff:
(i) for every $1h \in H$, $k \in K$, we have $hk = kh$. It is easily checked that this holds iff
(i') both H and K are normal in G.

The **(external) direct product** of groups H, K is the Cartesian product $H \times K = \{(h, k) : h \in H, k \in K\}$ with the group operation given by coordinatewise multiplication: $(h_1, k_1) \cdot (h_2, k_2) = (h_1h_2, k_1k_2)$.

It is easily checked that if G is an internal direct product of subgroups H and K then G is isomorphic to $H \times K$, and that conversely, $H \times K$ is the internal direct product of its subgroups $\widetilde{H} = H \times \{1\}$ and $\widetilde{K} = \{1\} \times K$.

If only one of the subgroups, say K, is normal, that is, if G is a product of H and K, and $K \lhd G$, then G is said to be an **(internal) semidirect product of K and H with normal subgroup K**. The semidirect product has an external version as well: given groups H, K and for all $h \in H$ an automorphism $\varphi_h : K \to K$, then we define a group $K \rtimes_\varphi H$, the **outer** or **external** semidirect product, by taking the following operation on the product set $H \times K$:

$$(k_1, h_1) \cdot (k_2, h_2) = (k_1\varphi_{h_1}(k_2), h_1h_2).$$

One checks that this is a group. Moreover, the external and internal versions are again equivalent: indeed, given G an internal semidirect product of N and H with N the normal subgroup, then $(n_1h_1)(n_2h_2) = (n_1\widetilde{n}_2)(h_1h_2)$ where $\widetilde{n}_2h_1 = h_1n_2$ whence $\widetilde{n}_2 = h_1n_2h_1^{-1}$, which is in N since that is normal, and which moreover defines an automorphism of N, as we have seen above.

Conversely, given an external semidirect product $K \rtimes_\varphi H$, then one checks that $K \times \{1\}$ is indeed a normal subgroup.

A semidirect product is also called a **split extension**. The reason for this is as follows.

Consider the exact sequence depicted here, recalling what (**exact** means: the image of a map is the kernel of the following map). Thus since the image of the first map is $1 \equiv \{\mathbf{e}\}$, that means the map α is injective, while the last map is surjective, so its kernel is all of K, implying that the image of β is K, whence β is surjective. So this diagram being exact simply says that α is $1-1$ while β is onto. In this case one says

that G is an **extension** of the subgroup H by K. See p. 413 of [**?**]. The fact that ??? implies that $\alpha(H)$ is normal in G.

$$1 \longrightarrow H \xrightarrow{\alpha} G \xrightarrow{\beta} K \longrightarrow 1$$

Now one says that this exact sequence **splits back** iff there exists a map $\varphi : K \to G$ such that $\beta \circ \varphi =$ id.

$$1 \longrightarrow H \longrightarrow G \underset{\varphi}{\overset{\beta}{\rightleftarrows}} K \longrightarrow 1$$

We summarize this by:

Proposition 46.1. *These are equivalent:*
(i) G is an internal semidirect product of H by K;
(ii) G is isomorphic to an external semidirect product of H by K;
(iii) The short exact sequence splits back.

Given a group G suppose we have subgroups $\{\mathbf{e}\} = 1 = G_0 \lhd G_1 \lhd \cdots \lhd G_{n-1} \lhd G_n = G$. This is called a **subnormal series**. If each factor group G_k/G_{k-1} is **simple** (has no nontrivial normal subgroups) then one says this is a **composition series** for G. Equivalently, it is a maximal subnormal series. If each $G_i \lhd G$, it is a **normal** series. If it is abelian this is called a **solvable series**. If there exists a solvable series, G is termed a **solvable** group.

The easiest for us to understand are solvable groups, for (Given $K \lhd G$, there may not exist such an $H-$ otherwise all groups would be solvable).

From semidirect products to skew products. There is an intriguing intuitive parallel between the algebraic notion of external semidirect product and the dynamical construction of skew product transformation. But can this analogy be made real? Here is one way.

Proposition 46.2. *Consider an external semidirect product $G = K \rtimes_\varphi H$, with*

$$(k, h) \cdot (\widetilde{k}, \widetilde{h}) = (k\varphi_h(\widetilde{k}), h\widetilde{h}).$$

Fix $\widetilde{g} = (\widetilde{h}, \widetilde{n}) \in G$. Define $\widehat{T} : G \to G$ by right multiplication by $\widetilde{g}$: $\widehat{T}(g) = g\widetilde{g}$. Let T denote the restriction of $\widehat{T}$ to H, i.e. right multiplication by $\widetilde{h}$. Then $\widehat{T}$ is a skew product transformation over the base map $T : H \to H$, that is, on the fiber bundle G with base H and fiber K, acting on the fibers by translation.

Proof. For any pair $(k, h) \in K \times H$ we have $g\widetilde{g} = (k, h) \cdot (\widetilde{k}, \widetilde{h}) = (k\varphi_h(\widetilde{k}), h\widetilde{h})$, so defining $\Phi_g(k) = k\varphi_h(\widetilde{k})$ then for $\Psi_h : K \to K$ defined by $\Psi_h = \Phi_g$, the map $\widehat{T} : G \to G$ acts on $K \times H$ by

$$\widehat{T}(k, h) = (\Psi_h(k), T(h))$$

which is the claimed skew product transformation.

□

Thus we can think of the skew product as a generalization to dynamics of a semidirect product of groups. In the same way, one can imagine generalizing the notion of a solvable group, by having a sort of central series of extensions. In fact, Furstenberg carried this out, both in the setting of topological dynamics and measure dynamics, the latter together with Zimmer. One can imagine replacing abelian groups by the dynamics of rotations of compact groups, which are isometries. The first step, a single isometric extension, we have examned above in Proposition **??**. See ???

Remark 46.1. (Nov 6 2015)Questions:

(1) Does the flow in the paper with Tom give a related example? The scenery paper with Pierre?

(2) Maybe the defn of skew product generalizes immediately to a group action on the base, and then the semidirect product is such an example??? Let's check it tomorrow!

Remark 46.2. Aug 10 2020 Questions:

(1)groups of matrices , upper trianglular

(2) central series of groups

)3= pos cone, Brat diags, extensions of actions. ex Heisenberg diagram!!!

47. Appendix: What is Functional Analysis?

Fuctional analysis is, in essence, the study of linear algebra in infinite dimensions, when a topology has been added. For finite dimensions, as we have seen in Lemma 35.41, all norms are equivalent, and indeed all reasonable topologies are, but that is far from the case in infinite dimensions, which is what makes the subject so rich and fascinating.

The word "analysis" in the title can be explained in this way: the best examples of these vector spaces are spaces of sequences or functions, and there all the tools of analysis will come into play: convergent series, derivatives, integrals and so on.

A key part of the subject is indicated by its name: linear functionals. Intuitively, the space V^* of continuous linear functionals on a topological vector space V serves to define *coordinates* on the space. We then work with V through analysis of these coordinates, so in summary Functional Analysis can be considered to be "analysis on function spaces making use of linear functionals".

First, beginning with finite dimensions:

48. Invariant means and time averages on the reals

In Example 2 we asked the question of how to choose a point randomly from a compact group. But it is when we move on to the noncompact setting that things get really interesting. And, as we shall see, the answer has everything to do with the notion of time average so familiar from the Birkhoff Ergodic Theorem.

Example 64. *Invariant means: choosing a point randomly from a noncompact group.*

An **invariant mean** on a discrete group G is a finitely additive invariant measure μ which is invariant for the action of the group on itself by left multiplication. Such a μ will provide an answer to our question, but several new issues are raised:

–Can we strengthen finite to countable additivity? (No!)
–Do invariant means always exist? (No: G is termed **amenable** if so).
–If G is amenable, is the invariant mean unique? (No).
–Can we reduce this "nonuniqueness" in some meaningful way? (Yes).
–Can we still do something meaningful in the nonamenable case? (Yes, sort of!)

We shall return to each of these points below.

But what should the definition be for non-discrete groups? First, let us note that amenability is equivalent to the existence of a left-invariant mean λ on the group, i.e. λ is a normalized positive translation-invariant functional on the space of bounded functions, $L^\infty(G)$ where this is with respect to the Haar measure on G (for a discrete group, the Haar measure is counting measure and so this is just $l^\infty(G)$).

This second definition generalizes to a locally compact Hausdorff group G; by a left-invariant mean λ on the group, we mean a normalized positive translation-invariant functional on $L^\infty(G, m)$ where m is Haar measure on G. Of course m itself is infinite when G is noncompact. This does agree with the definition for discrete groups, as then m is just counting measure.

At this point we need some examples. Abelian groups, such as $\mathbb{Z}^n$, $\mathbb{R}^n$, are always amenable; more generally any solvable group is (for example the $ax + b$ or **affine** group). On the other hand, the free group on two generators, F_2, provides a basic example of nonamenability. Another is $PSL(2, \mathbb{R})$, the group of isometries of hyperbolic space; see §??.

But first let us examine more closely the simplest example, $(\mathbb{R}, +)$, the additive group of the real numbers.

.....

Example 65. For this, it is convenient to begin instead with the motivating idea of integration.

By definition a **mean** λ on a locally compact group G is a normalized linear, positive, bounded functional on $L^\infty(G, \mu)$ where L^∞ is defined with respect to Haar measure μ; given a bounded function f, then $\lambda(f)$ is thought of as the integral of f with respect to our (now only finitely additive) measure μ. We recall that Haar measure is the unique (up to normalization) translation-invariant measure on the group; since the group may not be commutative, we have to fix here either the right- or left- action of the group on itself, and choose right- or -left Haar measure. When G is noncompact, Haar measure is infinite; on R^n it is Lebesgue measure of dimension n. One says λ is a (left) **invariant mean** if it is translation-invariant for the action on the left.

To produce an invariant mean λ on $\mathbb{R}$, note that for $f \in L^\infty$, $f(x) - f(x - t)$ has mean zero. Letting E denote the closed vector subspace generated by functions of this form, λ is defined uniquely on E by linearity and continuity. Now we extend λ to L^∞ by the Hahn-Banach extension theorem, and we have produced an invariant mean, and moreover all invariant means result from some such extension.

This is hardly particularly satisfying or useful, as it gives an ambiguous notion of average value even on some obvious examples. Indeed, take $f(x) = \sin(x)$, then

translation-invariance alone will tell us that $\lambda(f) = 0$, but this is no longer true for the function $f(x) = \sin(\sqrt{x})$. And yet, applying a simple time average encountered above, we arrive at 0 as the mean value here as well. Now ideally, in our search for "the" appropriate notion of random choice of a point, we might hope to arrive at something like Lebesgue measure on the circle, with its uniqueness. So perhaps what we need to look for is invariance with respect to other natural operations.

Consider now a nonnegative function $\varphi : \mathbb{R} \to \mathbb{R}$ of integral one. The **convolution** of φ with f is defined as

$$\varphi * f(t) = \int_{\mathbb{R}} \varphi(t-x) f(x) \mathrm{d}x.$$

This is commutative: taking u= t-x so x= t-u; du= -dx then

$$\varphi * f(t) = \int_{-\infty}^{\infty} \varphi(t-x) f(x) \mathrm{d}x = -\int_{x=-\infty}^{x=\infty} \varphi(u) f(t-u) \mathrm{d}u = \int_{u=-\infty}^{u=\infty} \varphi(u) f(t-u) \mathrm{d}u$$
$$= f * \varphi(t)$$

Hence writing the translate of the function $f(t)$ by u as $f_u(t) \equiv f(t-u)$, then this is $\int_{u=-\infty}^{u=\infty} f_u(t)\varphi(u)\mathrm{d}u$ leading to two quite different interpretations for the convolution: from the first equation, f remains still while φ is flipped and then translated, so $\varphi * f$ represents a local smoothing of f (and indeed, if φ has a certain degree of smoothness, that will be passed over to the convolution); from the second, $f * \varphi$ is a weighted average, by the (signed) normalized measure $\varphi(u)\mathrm{d}u$, of the translates of f.

Now via either interpretation, it is reasonable to require of our invariant mean that it is invariant for this operation: that $\lambda(f) = \lambda(\varphi * f)$. But note that the average of translates is given by an integral rather than a sum, and so this **convolution invariance** (also known as **topological invariance** [Gre69]) does not follow from translation invariance, rather it will need to be an additional assumption.

To prove there exists a convolution-invariant mean, we can we can enlarge the subspace E to include functions of the form $f - \varphi * f$ for all such φ, and proceed as before by Hahn-Banach extension.

Here is a further reasonable requirement. For m Lebesgue measure on $\mathbb{R}$, let us define the mean value of f to be:

$$\lambda(f) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f \mathrm{d}m,$$

whenever this limit exists. We define the measurable sets to be the algebra of sets A where this exists for $f = \chi_A$.

Thinking of f as a function of time, then half of this is

$$\lambda^+(f) = \lim_{T\to\infty} \frac{1}{T} \int_{0}^{T} f \mathrm{d}m,$$

the **Cesàro average** of f, giving our usual notion of the "time average" of the function.

A set $A \subseteq \mathbb{R}^+$ has **(Cesàro) density** $c \in [0,1]$ iff the Cesàro average of χ_A exists and equals c, i.e.

$$\lim_{T\to\infty} \frac{1}{T} m(A \cap [0,T]) \to c \text{ as } T \to \infty.$$

The integer version of this, for $A \subseteq \mathbb{N}$, is

$$\lim_{N\to\infty} \frac{1}{N} \#(A \cap [0,N-1]) \to c \text{ as } N \to \infty.$$

Although the above limit only exists for certain functions, we can again one can extend λ^+ to all of $L^\infty(\mathbb{R}, m)$ abstractly (by the Hahn-Banach Theorem).

But again this Hahn-Banach extension is nonunique. The Hardy-Riesz **log average** of f,

$$\lim_{T\to\infty} \frac{1}{\log T} \int_1^T f(x) \frac{1}{x} \mathrm{d}m,$$

might help; and next we can move on to their the **log log** average, defined by

$$\lim_{T\to\infty} \frac{1}{\log\log T} \int_e^T f(x) \frac{1}{x \log x} \mathrm{d}m,$$

and so on.

....

example: $\sin(\log x)$

....

But are these all compatible- that is, if the Hardy-Riesz $\log^{(n)}$ −average exists does the $\log^{(n+1)}$ −average does as well, and with the same value?

To see the answer, it helps to change somewhat our point of view, returning again to convolutions. Now considering the specific choice $\varphi(x) = \chi_{[0,\infty)} e^{-x}$, it follows that the Cesáro averaging operator is the conjugate via an exponential change of variables of the operator $f \mapsto \varphi * f$ on L^∞, and that moreover the $\log^{(n+1)}$ −averaging operator is the exponential conjugate of the $\log^{(n)}$ −operator.

So we can proceed as follows:
1) the Cesàro operator is consistent with any convolution operator; and hence the order –$(n+1)$ operator is consistent with the order – n operator, so all are consistent; moreover, taking this one step further,
2) there exists an invariant mean which is invariant both for convolution and for an exponential change of variables. Since convolution-invariance implies translation-invariance, this mean is invariant with respect to dilation, composition with x^r for $r > 0$, and so on.

Even after this, the Hahn-Banach extension is nonunique. For an attempt to see how far one can proceed, see [Fis87]- there are still some basic open questions here!

Theorem 48.1. [Fis87] *There exists an invariant mean λ on $\mathbb{R}$ which is invariant with respect to compostion with* exp *and is invariant for all the log averages of each order, and which is measure-linear in the sense of Mokobodski.*

The point of this discussion is that two different notions of mean value (the expected value of a function defined on a probability space, and the time average of a function defined on the reals) have quite different natures: for the first we have an actual

(countably additive) measure, while for the second it is only finitely additive. Now one of the nicest ideas in ergodic theory is when these two notions of average value are related: if we are given a flow on a measure space, and a function defined on the space, and a flow-invariant probability measure, then this measure gives an idea of randomness (from the invariance with respect to the action of the group $(\mathbb{R}, +)$.) So on the one hand we have the expected value of this function. On the other hand, we could choose a point randomly with respect to this measure, then observe the values of the function over th orbit of this point- which is a copy of the real line. Then we can ask if the time average of this exists. The positive answer is given by Birkhoff's Ergodic Theorem.

Now for the Birkhoff result, we don't need such a fancy invariant mean; in fact all we need is the Cesàro average.

The proof of this theorem is subtle, but it can be viewed as a sort of Fubini theorem on the interchange of the order of integration, for two measures, one countably and the other finitely additive (the time average). From this point of view, the need for a special proof is due to Fubini's theorem failing when we don't have countable additivity, and the content of the theorem is that the Cesáro averaging method is strong enough to get our result.

Test functions: Almost periodic functions. Ergodic theorem.

From this point of view, for our test functions, we have not made use of the stronger averaging methods (log, log log and so on), it is fair to ask whether these might not also prove useful in some dynamical context. In fact this is so, but only if we move beyond the standard setting, to the dynamics of infinite measures. See §???.

Here is one example: first digit problem Another: PCLT.

49. Infinite measure ergodic theory

50. More on towers; noninvertible towers

50.1. **Return times and induced maps: the noninvertible case.** Here we consider what can still be done when the original map is not invertible.

Beginning with a map $T : X \to X$, we first extend the return-time function to all of X, setting

$$r_A(x) = \inf\{n > 0 : T^n(x) \in A\}$$

and $r_A(x) = \infty$ iff x never returns to A.

Again defining $B = A^c$, we extend the return-time partition to all of X, setting

$$A_k = \{x \in A : r_A(x) = k\} \text{ and } B_k = \{x \in B : r_A(x) = k\}. \tag{164}$$

Note that now:

$$\begin{aligned} A_k &= A \cap T^{-1}(B) \cap \cdots \cap T^{-(k-1)}(B) \cap T^{-k}(A) \text{ and} \\ B_k &= B \cap T^{-1}(B) \cap \cdots \cap T^{-(k-1)}(B) \cap T^{-k}(A). \end{aligned} \tag{165}$$

The dynamics of T is as indicated in the illustration on the right side of Fig. 112 (borrowed from [AW73], where it is used to study a famous infinite measure-preserving, noninvertible map called Boole's transformation): A_1 and B_1 map to A; $B_{k+1} \cup A_{k+1}$ maps onto B_k; A_∞ maps to B_∞ which maps to itself. This Adler-Weiss picture

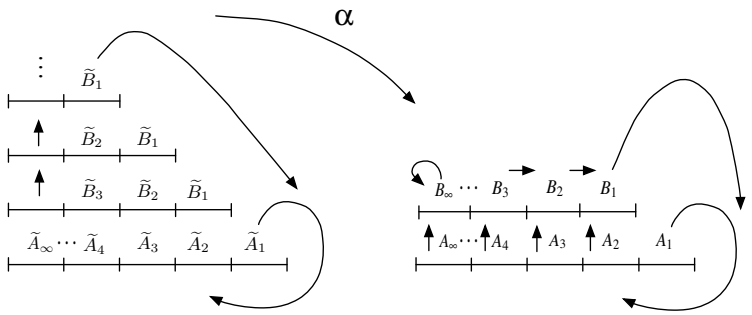


FIGURE 112. The semiconjugacy from the noninvertible tower $(\widetilde{X}, \widetilde{T})$ to the Adler-Weiss model of the original map (X, T); here $\widetilde{B}_i = \alpha^{-1}(B_i)$.

of (X, T) - (perhaps it should be called a split-level house!) replaces the Kakutani skyscraper (the internal tower) of Fig. 18 when the map is not invertible.

The induced map $T_A : A \setminus A_\infty \to A$ is defined as before, and we construct the tower over (A, T_A) with height function r_A, now writing $(\widetilde{X}, \widetilde{T})$ for the tower space and map $(\widehat{A}, \widehat{T}_A, r_A)$. The base of the tower will be denoted $\widetilde{A} \equiv A \times \{0\}$, with $\widetilde{B}$ its complement. We define subsets $\widetilde{A}_i, \widetilde{B}_i$ of $\widetilde{X}$ by replacing A, B with $\widetilde{A}, \widetilde{B}$ in equations (164), equivalently (165). We write $\widetilde{T}_A$ for the map $(x, 0) \mapsto (T_A(x), 0)$ on $\widetilde{A} \setminus \widetilde{A}_\infty$. Note that the map $(\widetilde{X}, \widetilde{T})$ is invertible inside the tower but not on the top.

This external tower maps to the split-level model of (X, T) by $\alpha : \widetilde{X} \to X$ where $\alpha(x, k) = T^k(x)$. Here $(x, k) \in \widetilde{X}$; thus, $x \in A$ and $k < r_A(x)$. We have:

Lemma 50.1.
(i) The map α is a homomorphism from $(\widetilde{X}, \widetilde{T})$ to its image inside of (X, T).
(ii)For each $k \geq 1$, $\alpha^{-1}(A_k) = \widetilde{A}_k$ and $\alpha^{-1}(B_k) = \widetilde{B}_k$.
(iii) If T is invertible, then α is a bijection to its image.

Proof. By definition of α, we have this commutative diagram, where $x \in A$ and $n = r_A(x) - 1$:

$$\begin{array}{ccccc} (x,0) & \xmapsto{\widetilde{T}^n} & (x,n) & \xmapsto{\widetilde{T}} & (T_A(x),0) \\ \downarrow \alpha & & \downarrow \alpha & & \downarrow \alpha \\ x & \xmapsto{T^n} & T^n(x) & \xmapsto{T} & T_A(x) \end{array}$$

Therefore the following diagram commutes, proving (i):

$$\begin{array}{ccc} \widetilde{X} & \xrightarrow{\widetilde{T}} & \widetilde{X} \\ \downarrow \alpha & & \downarrow \alpha \\ X & \xrightarrow{T} & X \end{array}$$

For (ii), since $\alpha \circ \widetilde{T} = T \circ \alpha$, for the action on sets via the inverse image maps we have $\alpha^{-1} \circ T^{-j} = \widetilde{T}^{-j} \circ \alpha^{-1}$. Hence the definition of A_k, B_k in (165) converts, after

the application of α^{-1}, to the corresponding definition of $\widetilde{A}_k, \widetilde{B}_k$. Thus $\alpha^{-1}(A_k) = \widetilde{A}_k$ and $\alpha^{-1}(B_k) = \widetilde{B}_k$.

To prove (iii), suppose $\alpha(x) = \alpha(y) = z$. Suppose $z \in A_k$; then by (ii) both x and y are in $\widetilde{A}_k$, but α is a bijection from $\widetilde{A}$ to A, so $x = y$. If $z \in B_k$, then $x, y \in \widetilde{B}_k$ by (ii). We follow x, y down the tower to the base $\widetilde{A}$, by applying $\widetilde{T}^{-1}$ the appropriate number of times. Since $\widetilde{T}$ is invertible inside the tower, this gives two points x', y'. Now we follow z down to $z' \in A$; assuming T is invertible, this gives a single point z'; we have that $\alpha(x') = \alpha(y') = z'$. But since α is a bijection from $\widetilde{A}$ to A, $x' = y'$ and hence $x = y$.

□

We shall say that a set $A \subseteq X$ is **recurrent** iff every $x \in X$ eventually enters A, so r_A is finite for each $x \in X$ exactly when A is a recurrent set. Equivalently, $\cup_{n=1}^{\infty} T^{-n} A = X$. In particular, every point of A itself returns to A.

Theorem 50.2.
(i) Given a measurable map T of a measurable space $(X, \mathcal{A})$ and $A \in \mathcal{A}$, then A_k, B_k are measurable sets, and r_A, T_A, $\widetilde{T}$, and α are measurable functions.

If μ is an invariant measure and $\mu(A)$ is positive finite, then if A is recurrent:
(ii) $\mu(B_k) \to 0$ as $k \to \infty$;
(iii)(Adler-Weiss [AW73]) the map T_A preserves $\mu|_A$ (the measure μ restricted to A);
(iv) the map $\alpha : \widetilde{X} \to X$ sending (x, n) to $T^n(x)$ is a homomorphism from the tower $(\widetilde{X}, \widetilde{T}, \widetilde{\mu})$ to (X, T, μ). This is onto a.s., hence a factor map;
(v) the map α is an isomorphism iff (X, T, μ) is invertible.

Proof. In the first statement, we are given a $\sigma-$algebra $\mathcal{A}$ with respect to which the map T and the subset A are measurable, and it is understood that $\mathbb{N}^* \equiv \{1, 2, \dots\}$ is equipped with the discrete topology (every set is open) and the associated Borel $\sigma-$algebra (every set is measurable). By (165) A_i and B_i are measurable sets, hence r_A is measurable.

By definition $T_A^{-1}(E)$ consists of all points in A that take some least time k to return to A and then enter it in the subset E. By (165), we can write this set as the disjoint union

$$T_A^{-1}(E) = \cup_{k=1}^{\infty} A_k \cap T^{-k}(E). \tag{166}$$

Therefore T_A is measurable, since T hence each T^k is.

The tower map $\widetilde{T}$ inside the tower simply moves the level of a column upwards, so is measurable. That is, for $E \subseteq \widetilde{X} \setminus \widetilde{A}$, $\widetilde{T}^{-1}(E)$ is measurable. It suffices to show the same for $E \subseteq \widetilde{A}$. Write $\check{T}$ for the map from the base $\widetilde{A}$ to the top $\widetilde{B}_1$, defined on A_k to be $\widetilde{T}^{k-1}$; this map is a bimeasurable bijection. Then for $E \subseteq \widetilde{A}$, $\widetilde{T}^{-1}(E) = \check{T}(T_A^{-1}(E))$ which is measurable.

Next we show that α is measurable. Consider first $E \subseteq A$. But α is essentially the identity map on the bases, that is $\alpha^{-1}(E) = E \times \{0\}$ which is measurable. Next suppose that $E \subseteq B_1$, and let $\widetilde{E} = T^{-1}(E)$. Consider $\widetilde{E}' \equiv \check{T}^{-1}(\widetilde{E})$; this takes $\widetilde{E}$ down the tower to the base, and since $\check{T}$ is a bimeasurable bijection, this set is measurable iff $\widetilde{E}$ is. Now pull back E to a set $E' \subseteq A$ via applications of T.

Now $T^{-1}(B_1) = A_1 \cup B_2$, $T^{-1}(B_2) = A_2 \cup B_3$, $T^{-1}(B_3) = A_3 \cup B_4$, and so on, so $E' = T^{-1}(E \cap B_1) = (E' \cap A_1) \cup (E' \cap A_2) \cup (E' \cap A_3) \cup \dots$ so $E' = \cup_{k=1\infty} E' \cap A_k)$

But $\check{T}^{-1}(\widetilde{E}) = ???$

$\alpha^{-1}(E)$ has the same measure as $T_A^{-1}(E)$ which by part (iii) is equal to $\mu(E)$.

For (ii), to show that $\mu(B_k) \to 0$, note that $T^{-1}(A) = A_1 \cup B_1$, so $\mu(A) = \mu(T^{-1}A) = \mu(A_1) + \mu(B_1)$. Similarly, $T^{-1}(B_1) = A_2 \cup B_2$, so $\mu(B_1) = \mu(T^{-1}B_1) = \mu(A_2) + \mu(B_2)$. Continuing inductively, at stage n we have $\mu(A) = \sum_{k=1}^{n} \mu(A_k) + \mu(B_n)$. So $\mu(B_n) = \mu(A) - \sum_{k=1}^{n} \mu(A_k) = \sum_{k=n+1}^{\infty} \mu(A_k) + \mu(A_\infty)$ for each n. Since A has finite measure the sum $\sum_{k=1}^{\infty} \mu(A_k)$ converges, and by recurrence A_∞ has measure zero. Thus $\mu(B_k) \to 0$ as claimed.

Now we prove (iii). Let $E \subseteq A$. The set $T_A^{-1}(E)$ consists of all points in A that take some time k to return to A and then enter it in the subset E. We can write this set as a disjoint union

$$T_A^{-1}(E) = \cup_{k=1}^{\infty} A_k \cap T^{-k}(E), \tag{167}$$

which is measurable. Using the same idea as before, for each n,

$$\mu(E) = \sum_{k=1}^{n} \mu(A_k \cap T^{-k}E) + \mu(B_n \cap T^{-n}E).$$

Since $\mu(B_n) \to 0$ as $n \to \infty$, we have

$$\mu(E) = \sum_{k=1}^{\infty} \mu(A_k \cap T^{-k}E) \tag{168}$$

and by (167) T_A is measure-preserving as claimed.

Next we prove (iv). By Lemma 50.1, we know that α is a semiconjugacy in the set category; it is clearly measurable.

.....????

If A is transitive then α is onto, since $y \in X$ is $T^n(x)$ for some $x \in A$, and taking the least such n, then $y = \alpha(x, n)$.

By part (i), the map T is transitive, hence α is onto.

.....

We verify that it is measure-preserving.

We note that $\alpha(x, 0) = x \in A$, while for $1 \leq k \leq r(x) - 1$, $\alpha(x, k) = T^k(x) \in A^c = B$.

Consider a set $E \subseteq A$. We have $\alpha^{-1}(E) = E \times \{0\}$ which has the same measure as E.

Next suppose that $E \subseteq B_1$. Let $\widetilde{E} = T^{-1}(E)$. We shall show the set $\alpha^{-1}(E)$ has the same measure as $T_A^{-1}(E)$ which by part (iii) is equal to $\mu(E)$.

First we claim that $\alpha^{-1}(E)$ is contained a subset of the top of the tower. Indeed, for $x \in E$, $Tx \in A$ but $x \notin A$. Thus the inverse image $\alpha^{-1}(x)$ is in the top of the tower, since all points in this set have the form $(w, r(w) - 1)$. We project each such point to $(w, 0)$ in the base. This is in $T_A^{-1}(E) \times \{0\}$; we note that conversely every point in this set has that form.

Pictorially, we take each part of $\alpha^{-1}(E)$ in a given column down its elevator until it hits the base. This preserves measure within the tower. But the resulting set is $T_A^{-1}(E) \times \{0\}$ which has the same measure as $T_A^{-1}(E)$, as stated.

Finally, if T is $1-1$, then α is: the map $\widetilde{T}$ is a bijection up the tower, and T as well; now α is a bijection from the base to A, so that property transports up the tower.

....

False proof that

(ii)The set A is transitive.

We shall say A is **transitive** if $\cup_{n=0}^{\infty} T^n A = X$.

If A is transitive, then α is onto hence a factor map.

Now we prove (ii). Consider $W = \{x : x \in (T^n(A))^c \text{ for each } n \geq 0\}$. We are to show that $\mu(W) = 0$. A priori forward images are not always measurable, but we are assuming the measure space is complete, and will bound this set by small measurable sets.

Since $\{B_i\}$ partition B, $W = \cup_{i=1}^{\infty} W \cap B_i$. Now if $x \in W \cap B_1$ and $T(y) = x$, and $y \notin A$, then $y \in B_2$. Hence $T^{-1}(W \cap B_1) \subseteq W \cap B_2$, and similarly for higher powers. So $\mu(W \cap B_1) = \mu(T^{-1}(W \cap B_1)) \leq \mu(W \cap B_2) \leq \mu(W \cap B_n) \to 0$ as $n \to \infty$. So recurrence does imply transitivity.

??

□

Note that part (iii) above gives a fourth proof of Theorem 5.1, Poincaré recurrence.

50.2. **Noninvertible towers and the natural extension.**

51. Appendix: Transitive points for group actions

Here we extend the ideas of §5.2 to flows and semiflows. The "no isolated points" property will be replaced by an appropriate condition (that compact pieces of flow orbits cannot fill up an open subset). First we have these definitions:

Definition 51.1. Given a topological space $(X, \mathcal{T})$, a **continuous flow** τ_t on X is a jointly continuous map $\tau : X \times \mathbb{R} \to X$ which satisfies the **flow property**: writing $\tau_t(x) = \tau(x, t)$, this is $\tau_{t+s} = \tau_t \circ \tau_s$. To define a continuous **semiflow** we replace $\mathbb{R}$ above by $\mathbb{R}^+ = [0, +\infty)$.

Given a flow on X, a point $x \in X$ is **transitive** iff it has a dense biinfinite orbit, $\{\tau_t(x) : t \in \mathbb{R}\}$; the flow is transitive iff there exists a transitive point. A point x is **forward transitive** for a flow or semiflow if the forward orbit $\{\tau_t(x) : t \geq 0\}$ is dense, and a (semi)flow is forward transitive iff there exists such a point.

For our statement we need a replacement for the notion of no isolated points.

Definition 51.2. Given a topological (semi)flow τ_t on X, we say the flow has **no isolated orbit segments** iff given any nonempty open set $\mathcal{U}$, $x \in X$ and $J \subseteq \mathbb{R}$ a compact interval, then $\mathcal{U} \setminus \{\tau_t(x) : x \in J\}$ is a nonempty open set (it is open since the continuous image of a compact set is compact).

Here is a topological property of X which guarantees this. A **curve** in X is a continuous bijective map $\gamma : J \to X$ where J is some subinterval of $\mathbb{R}$. We say the

space X has **no isolated compact curves** iff given any nonempty open set $\mathcal{U}$, and any curve γ defined on J a compact interval, then $\mathcal{U} \setminus \gamma(J)$ is a nonempty open set. Since orbits of τ_t are curves, this implies the no isolated orbit segment property.

Note that if X is a metric space, it is equivalent to require that this be nonempty for all balls $\mathcal{U} = B_\delta(x)$.

Note further that a special semiflow, that is a suspension over a continuous map with continuous return-time function, such that the base map has no isolated points, certainly satisfies the no isolated orbit segment property.

Proposition 51.1. *Let $(X, \mathcal{T})$ be a Polish space, and let τ_t be a (semi)flow on X, with no isolated orbit segments.*
(i) Then if τ_t is a transitive flow, the set E of forward transitive points is residual.
(ii)If τ_t is a forward transitive semiflow, the set E of forward transitive points is residual.

The same holds if instead of asumming no isolated orbit segments, we assume that there exists t_0 such that the time-t_0 map $T \equiv \tau_{t_0}$ has a (forward) transitive point.

Proof. As before, let $\{\mathcal{U}_i\}_{i\geq 1}$ be a countable base for the topology $\mathcal{T}$ of the separable metric space (X, d). Now we have

$$E = \inf G_j = \cap_{j\geq 1} \cup_{t\geq 0} \tau_{-t}(\mathcal{U}_j).$$

We wish to show each of the open sets G_j is dense. Thus we claim that for each $i \geq 1$, $\mathcal{U}_i$ meets G_j.

For part (i), we are given that there exists a transitive point x with $\{\tau_t(x) : t \in \mathbb{R}\}$ dense; for (ii) we know this for $\mathbb{R}^+$. Since the flow space X has no isolated orbit segments, this orbit cannot be periodic.

Let $\mathcal{U}$ be an open set. Since τ_t is jointly continuous, given $w \in X$, the curve $\tau_t(w)$ is a continuous map from $\mathbb{R}$ to X, so the inverse image in $\mathbb{R}$ of $\mathcal{U}$ by this curve is open hence a countable union of disjoint intervals. Since there are no isolated orbit segments, this inverse image is unbounded at $+\infty$ for a semiflow, and also at -∞ for a flow.

Now we argue as for the case of a map. Given $i, j \geq 0$, the orbit of x enters both $\mathcal{U}_i$ and $\mathcal{U}_j$ in an unbounded, infinite number of time intervals. So it enters one of them first. If it is $\mathcal{U}_i$, we are done, exactly as before, as $\mathcal{U}_i$ meets G_j. Now if the flow or semiflow is forward transitive, this is always the case for some point, as the intervals corresponding to $\mathcal{U}_i$ and $\mathcal{U}_j$ each occur infinitely often towards $+\infty$. Lastly, in the flow case, given a biinfinitely transitive point and that $\mathcal{U}_j$ occurs first, then there is an interval of times $J = [a, b]$ such that $\mathcal{U} \equiv \mathcal{U}_j \cap \{\tau_{-t}(\mathcal{U}_i) : \ t \in J\}$ is nonempty. By the property of no isolated orbit segments, there is some $s > b$ such that $\mathcal{U} \cap \tau_{-s}(\mathcal{U})$ is nonempty. Therefore, reasoning as for discrete time, $\mathcal{U}_i$ meets G_j and we are done. □

Remark 51.1. One can extend the notion of transitive point to an action of a group G action on a topological space X by continuous maps (hence by homeomorphisms, since we have inverses). Note that the group itself is not required to have a topology. We shall say the action is **dynamically** transitive iff there exists a transitive point. The reason we have added the modifier "dynamically" is because of this much stronger

property: the action is called **transitive** iff for each $x, y \in X$ there exists $g \in G$ with $g(x) = y$. Thus a transitive transformation or flow is generally not transitive as a $\mathbb{Z}-$ or $\mathbb{R}-$action!

52. Yet more examples

52.1. **The boundary at infinity of the free semigroup and free group.** –use erg thm to show goes to ∞ a.s.

52.2. **Graph-directed IFS.**

53. Substitution dynamical systems and adic transformations

53.1. **The Morse-Thue example.**

53.2. **The golden rotation.**

53.3. **Tiling spaces and nonstationary solenoids.**

53.4. **The space-filling curve of Arnoux and Rauzy.**

53.5. **Penrose tiles ala Robinson.**

54. Nonstationary dynamical systems

55. Infinite ergodic theory

Hopf thm, ∞ meas: r walk, Ch-Erd. ; integer Cantor set

55.1. **The scenery flow for hyperbolic Cantor sets.**

56. Conformal measures

57. Back to the scenery flow

57.1. **Bowen's formula for Hausdorff dimension.**

57.2. **Entropy of the scenery flow.**

57.3. **Geometric models for Gibbs states.**

57.4. **Unique ergodicity for the horocycle flow.**

57.5. **The scenery flow of a hyperbolic Julia set.**

57.6. **Infinite measures.**

58. Móbius transformations and the scenery flow of rotation tilings

59. The frame flow and the scenery flow for Kleinian limit sets

60. The scenery flow of a hyperbolic Julia set

61. The Oseledec Theorem

62. The Oseledec Theorem and the boundary at infinity

63. The boundary of a Gromov hyperbolic group

64. The Stable Manifold Theorem

65. Ideas

–towers: (WANT INVERTIBLE EX WHERE X_A IS NOT INVARIANT SET)
T compressible iff T_A is....??
-amenability
–is weak mixing true for infinite product?
hyp metric: see Pliss Lemma used by Pujals ? Peson ??

Example 66. *other non-compact groups*

–Coudene Banach-Saks, Hopf argument, von Neumann Erg Thm proof, mixing.
–new proof of Caratheodory extension;
—for primitive matrix sequence, the other eigenvalues show up as finitely additive charges like Boyland/Bufetov; by Alexandroff's thm these signed measures can't be regular?!
—this passes over to the upper triangular situation;
—giving a version of Oseledec splitting like with Simon;
—check out Kaimanovich as quoted by Margulis/K: do they really handle all sequences???
–version for Gromov hyperbolic/ for Teich space.
–Harmonic functions when not independent? Harmonic projectin????
–averages ala Manuel in that case??
–Hermitian version of Oseledec???
–when PF condition fails what happens to the other exponents? e.g. two positive then fewer nonpositive; two countably additve.
–asip like Freedman; joining like Dudley
–stable limit thm.
–Martingale CLT/ Gordin
–subadditve erg thm.
–vel changes in flows
—nonst dyns driven not by stat dyns but by something else (global warming; universe expansion)...
–use Polish spaces for nat'l extn
–tightness/Fomin
–Dye via adic
–microsets

–Ehrenfest urn
–RW on $\mathbb{Z}, \mathbb{Z}^2$
–note : dual stmt for Parry is related to Birkhoff comp subspace

66. Self-similar groups

Theorem 66.1. *$B_0 \in \mathcal{B}_0$ to this cross-section are $B_1, \ldots, B_i$ with $B_{i+1} = A_i B_i D_i$ where $B_i = \begin{bmatrix} a_i & c_i \\ -b_i & d_i \end{bmatrix}$, $D_i = \begin{bmatrix} \lambda_i & 0 \\ 0 & \lambda_i^{-1} \end{bmatrix}$ and*
for parity **0**:
$a_i = [n_i n_{i+1} \ldots], b_i = 1, d_i/c_i = [n_{i-1} n_{i-2} \ldots]$, *and* $\lambda_i = 1/a_i$, *and* $A_i = \begin{bmatrix} 1 & 0 \\ n_i & 1 \end{bmatrix}$,
for parity **1**:
$b_i = [n_i n_{i+1} \ldots], a_i = 1, c_i/d_i = [n_{i-1} n_{i-2} \ldots]$, *and* $\lambda_i = 1/b_i$ *and* $A_i = \begin{bmatrix} 1 & n_i \\ 0 & 1 \end{bmatrix}$.

Fractal geometry, log averages, and infinite measures in ergodic theory

A measure- preserving transformation is called *recurrent* if the set of return times to a finite measure subset is infinite for a.e. initial point.

For transformations which preserve a finite measure, recurrence always holds, by Poincare's Recurrence Theorem. A stronger statement is Kac' theorem that the expected return time, in the ergodic case (meaning there are no nontrivial invariant subsets) is inversely proportional to the subset measure.

Birkhoff's ergodic theorem gives the much deeper statement that the frequency of returns a.s. converges to this measure. This leads to the statement "time average equals space average", extending the Strong Law of Large Numbers far beyond the original setting of independent coin-tosses.

Let us call an integer set *sparse* if it is infinite and of density zero. For example, in the conservative (equivalently, recurrent) ergodic infinite measure setting, the return times to a finite measure subset are always sparse. Thus the statement "time average equals space average" still holds in a trival sense as both are zero.

However for certain cases a nontrivial meaning for this statement can be discovered. This is when the return sets have a fractal-like structure. Then we can define the "Hausdorff dimension" of the integer subset to be $d = \lim \log N_n^A / \log n$ where $N_n^A(x)$ is the number of returns of x to the set A up to time n. We then normalize by n^d followed by application of a log average. The result is an *order-two ergodic theorem.*

Cases where this has been shown to work are of three basic types: adic examples, incuding certain infinite measure-preserving interval exchange transformations; renewal- type examples, including certain maps of the interval with a neutral fixed point; and horocycle flows of a geometrically finite Riemann surface of second type.

The analysis of all of these is based on the idea of constructing a related *scenery flow* which shows the fractal-like behavior.

In this talk we survey some of these ideas, developed together with coauthors Tim Bedford, Mariusz Urbanski, Jon Aaronson, Manfred Denker, Pierre Arnoux, and Marina Talet.

References

[AA68] Vladimir Igorevîč Arnold and André Avez. *Ergodic Problems of Classical Mechanics*, volume 9 of *Mathematical Physics Monograph Series*. Benjamin, 1968.

[Ace18a] Jeovanny de Jesus Muentes Acevedo. On the continuity of the topological entropy of non-autonomous dynamical systems. *Bulletin of the Brazilian Mathematical Society, New Series*, 49(1):89–106, 2018.

[Ace18b] Jeovanny de Jesus Muentes Acevedo. Openness of anosov families. *Journal of the Korean Mathematical Society*, 55(3):575–591, 2018.

[Adl98] Roy L. Adler. Symbolic dynamics and Markov partitions. *Bulletin of the American Mathematical Society*, 35(1):1–56, 1998.

[AF01] P. Arnoux and A. M. Fisher. The scenery flow for geometric structures on the torus: the linear setting. *Chinese Ann. of Math.*, 4:427–470, 2001.

[AF02] P. Arnoux and A. M. Fisher. A simple proof of the Perron-Frobenius theorem, 2002. *Preprint.*

[AF05] P. Arnoux and A. M. Fisher. Anosov families, renormalization and nonstationary subshifts. *Erg. Th. and Dyn. Sys.*, 25:661–709, 2005.

[Ahl66] Lars V. Ahlfors. *Complex Analysis*. McGraw-Hill, second edition, 1966.

[Aki13] Ethan Akin. *Recurrence in topological dynamics: Furstenberg families and Ellis actions*. Springer Science & Business Media, 2013.

[Arm83] M.A. Armstrong. *Basic Topology*, volume 8 of *Undergraduate Texts in Mathematics*. Springer, 1983.

[Arn94] Pierre Arnoux. Le codage du flot géodésique sur la surface modulaire. *l'enseignement Mathématique*, 40:29–48, 1994.

[Arn04] Vladimir Igorevc Arnold. *Lectures on partial differential equations*. Springer, 2004.

[Arn12] Vladimir Igorevich Arnold. *Geometrical methods in the theory of ordinary differential equations*, volume 250. Springer Science & Business Media, 2012.

[AW70] Roy L. Adler and Benjamin Weiss. Similarity of automorphisms of the torus. *Memoirs of the American Mathematical Society*, 98:1–43, 1970.

[AW73] Roy L. Adler and Benjamin Weiss. The ergodic infinite measure preserving transformation of Boole. *Israel J. Math.*, 16:263–278, 1973.

[Axl97] Sheldon Jay Axler. *Linear Algebra Done Right*, volume 2. Springer, 1997.

[Bar66] Bartle. *The Elements of Integration*. Wiley, 1966.

[Bea83] Alan F. Beardon. *The Geometry of Discrete Groups*. Springer-Verlag, 1983.

[Bed86a] Tim Bedford. Dimension and dynamics for fractal recurrent sets. *Journal of the London Mathematical Society*, 2(1):89–100, 1986.

[Bed86b] Tim Bedford. Generating special Markov partitions for hyperbolic toral automorphisms using fractals. *Ergodic Theory and Dynamical Systems*, 6(3):325333, 1986.

[BF92] T. Bedford and A. M. Fisher. Analogues of the Lebesgue density theorem for fractal sets of reals and integers. *Proc. London Math. Soc.*, 64:95–124, 1992.

[BF96] T. Bedford and A. M. Fisher. On the magnification of Cantor sets and their limit models. *Monatsh. Math.*, 121:11–40, 1996.

[BF97] T. Bedford and A. M. Fisher. Ratio geometry, rigidity and the scenery process for hyperbolic Cantor sets. *Erg. Th. and Dyn. Sys.*, 17:531–564, 1997.

[BFU02] T. Bedford, A. M. Fisher, and M. Urbański. The scenery flow for hyperbolic Julia sets. *Proc. London Math. Soc.*, 2(85):467–492, 2002.

[BG95] T. Bogenschütz and M. Gundlach. Ruelle's transfer operator for random subshifts of finite type. *Ergodic Th. and Dynam. Sys.*, 15:413–447, 1995.

[BH92] Mladen Bestvina and Michael Handel. Train tracks and automorphisms of free groups. *Annals of Mathematics*, 135(1):1–51, 1992.

[Bil65] Patrick Billingsley. *Ergodic Theory and Information*. John Wiley and Sons (republished 1978 by Krieger), 1965.

[Bil68] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.

[Bil78] P. Billingsley. *Ergodic Theory and Information*. Wiley, 1965 (reprinted by Krieger, New York, 1978).

[Bir31] George D Birkhoff. Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences*, 17(12):656–660, 1931.

[Bir57] Garrett Birkhoff. Extensions of Jentzsch's theorem. *Trans. AMS*, 85:219–227, 1957.

[Bir67] G. Birkhoff. *Lattice Theory*, volume XXV of *AMS Colloq. Publ.* AMS, 3rd edition, 1967. Chapter XVI.

[BM77] Rufus Bowen and Brian Marcus. Unique ergodicity for horocycle foliations. *Israel Jour. Math.*, 26(1):43–67, 1977.

[Bor95] Vivek S Borkar. *Probability theory. Universitext*. Springer-Verlag, New York. An advanced course, 1995.

[Bou13] Nicolas Bourbaki. *Topological vector spaces: Chapters 1–5*. Springer Science & Business Media, 2013.

[Bow75] Rufus Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms, SLN 470*. Springer Verlag, 1975.

[Bow77] Rufus Bowen. On Axiom A diffeomorphisms. *Conference Board Math. Sciences*, 35, 1977.

[BS02] M. Brin and G. Stuck. *Introduction to Dynamical Systems*. Cambridge University Press, 2002.

[BT82] Raoul Bott and Loring W Tu. *Differential Forms in Algebraic Topology*, volume 82 of *Graduate Texts in Mathematics*. Springer Verlag, 1982.

[Bus73] P. Bushell. Hilbert's metric and positive contraction mappings in Banach spaces. *Arch. Mech. Rat. Anal.*, 52:330–338, 1973.

[CB14] Ruel Churchill and James Brown. *Ebook: Complex Variables and Applications*. McGraw Hill, 2014.

[CFS82] I.P. Cornfeld, S.V. Fomin, and Ya. G. Sinai. *Ergodic Theory*, volume 245 of *Grundlehren der Matematischen Wissenschaften*. Springer, 1982.

[DC16] Manfredo P Do Carmo. *Differential geometry of curves and surfaces: revised and updated second edition*. Courier Dover Publications, 2016.

[Dek82] Michel Dekking. Recurrent sets. *Adv. in Math.*, 44:78–104, 1982.

[dJMA17] Jeovanny de Jesus Muentes Acevedo. *Anosov families: Structural Stability, Invariant Manifolds and Entropy for Non-Stationary Dynamical Systems*. PhD thesis, Universidade de Sao Paulo, 2017.

[DLH93] Pierre De La Harpe. On Hilbert's metric for simplices. *Geometric Group Theory*, 1:97–119, 1993.

[dMvS93] Welington de Melo and Sebastien van Strien. *One-Dimensional Dynamics*. Number 3. Folge- Band 25 in Ergebnisse der Matematik und Ihrer Grenzgebiete. Springer, 1993.

[Doo12] Joseph L Doob. *Measure theory*, volume 143. Springer Science & Business Media, 2012.

[DS57] N. Dunford and J. T. Schwartz. *Linear Operators, Part I: General Theory*, volume VIII of *Pure and Applied Mathematics*. Interscience Publishers (John Wiley and Sons), 1957.

[ES80] Edward G. Effros and Chao-Liang Shen. Approximately finite C^*-algebras and continued fractions. *Indiana University Math. Journal*, 29(2):191–204, 1980.

[Ete81] Nasrollah Etemadi. An elementary proof of the strong law of large numbers. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 55(1):119–122, 1981.

[EW10] Manfred Einsiedler and Thomas Ward. *Ergodic Theory*, volume 259 of *Graduate Text in Mathematics*. Springer, 2010.

[FB79] P. Ferrero and B.Schmitt. Ruelle's Perron-Frobenius Theorem and projective metrics. *Colloquia Math. Soc. János Bolyai (Random Fields)*, 27:333–336, 1979.

[FB88] P. Ferrero and B.Schmitt. Produits aléatoires d'operateurs matrices de transfert. *Probability Theory*, 79:227–248, 1988.

[Fer95] Sebastien Ferenczi. Les transformations de Chacon: combinatiore, structure géoétrique, lien aves les systèmes de complexité $2n + 1$. *Bull. SMF*, 123(2):272–292, 1995.

[Fer02] Sebastien Ferenczi. Substitutions and symbolic dynamical systems. In A. Siegel V. Berthe, P. Arnoux, editor, *Substitutions in Dynamics, Arithmetics and Combinatorics*, number 1794 in Lecture Notes in Math. Springer, 2002.

[Fis87] A. M. Fisher. Convex-invariant means and a pathwise central limit theorem. *Adv. Math.*, 63:213–246, 1987.

[Fis92] A. M. Fisher. Integer Cantor sets and an order-two ergodic theorem. *Ergod. Th. and Dynam. Sys.*, 13:45–64, 1992.

[Fis04] A. M. Fisher. Small-scale structure via flows. In *Progress in Probability*, pages 59–78. Birkhäuser, 2004. Conference Proceedings, Fractal Geometry and Stochastics III, Friedrichroda, Germany, March 2003.

[Fis09] A. M. Fisher. Nonstationary mixing and the unique ergodicity of adic transformations. *Stochastics and Dynamics*, 9(3):335–391, 2009.

[Fla63] Harley Flanders. *Differential forms with applications to the physical sciences*, volume 11. Courier Corporation, 1963.

[FLS64] RP Feynman, RB Leighton, and M Sands. The feynman lectures on physics, ii, addison-rwesley. *Reading, Massachusetts*, 1964.

[Fol99] Gerald B Folland. *Real analysis: modern techniques and their applications*, volume 40. John Wiley & Sons, 1999.

[Fom43] S. Fomin. Finite invariant measures in the flows (Russian). *Rec. Math. (Math. Sbornik) N.S.*, 12(54):99– 108, 1943.

[Fra69] John Franks. Anosov diffeomorphisms on tori. *Trans. AMS*, 145:117–124, 1969.

[Fra70] John Franks. Anosov diffeomorphisms. *Global Analysis: Proceedings of the Symposia in pure mathematics*, 14:61–93, 1970.

[FT12] A. M. Fisher and M. Talet. Dynamical attraction to stable processes. *Annales de l' Institut H. Poincaré, Probabilités et Statistiques*, 48(2):551–578, 2012.

[FT15] A. M. Fisher and M. Talet. Asymptotic self-similarity and order-two ergodic theorems for renewal flows. *Journal d'Analyse Mathématique*, 127:1–45, 2015.

[FT23] A. M. Fisher and M. Talet. Nonstationary solenoids and adic transformations, 2023. In preparation.

[Fur60] Harry Furstenberg. *Stationary Processes and Prediction Theory*. Number 44 in Annals of Mathematics Studies. Princeton University Press, 1960.

[Fur61] Hillel Furstenberg. Strict ergodicity and transformation of the torus. *Amer Jour Math*, 83:573–601, 1961.

[Fur70] Hillel Furstenberg. Intersections of Cantor sets and transversality of semigroups. In R. C. Gunning, editor, *Problems in Analysis, a symposium in honor of S. Bochner*. Princeton Univ. Press, 1970.

[Fur71] Hillel Furstenberg. *Random walks and discrete subgroups of Lie groups*, volume 1, pages 3–63. Dekker, New York, 1971.

[Fur80] Hillel Furstenberg. Random walks on Lie groups. In M. De Wilde J.Wolf, M. Cohen, editor, *Harmonic Analysis and Representations of Lie groups, Math. Phys. and Appl.*, volume 5, pages 467–489. D. Reidel, Dordrecht, Holland, 1980.

[Fur81] H. Furstenberg. *Recurrence in Ergodic Theory and Combinatorial Number Theory*. Princeton University Press, Princeton, 1981.

[Gil82] W. J. Gilbert. Fractal geometry derived from complex bases. *Math. Intelligencer*, 4:78–86, 1982.

[GP74] Victor Guillemin and Alan Pollack. *Differential topology*. Prentice-Hall, 1974.

[Gre69] Frederick P. Greenleaf. *Invariant Means on Topological Groups and Their Applications*, volume 16 of *Van Nostrand mathematical studies*. Van Nostrand Reinhold Co., New York, 1969. University of Michigan.

[GS92] G. Grimett and D. Stirzaker. *Probability and Random Processes*. Oxford Univ. Press, 1992.

[Gui02] Hamilton Luiz Guidorizzi. Um Curso de Cálculo, Vols I- III, ltc, 5a, 2002.

[Hal50] Paul R Halmos. *Measure Theory*. Van Nostrand Reinhold Company, 1950.
[Hal60] P. R. Halmos. *Lectures on Ergodic Theory*. Chelsea Publishing Co., New York, 1960.
[Hal74] P. R. Halmos. *Naive Set Theory*. Undergraduate Texts in Mathematics. Springer, 1974.
[Hal76] Paul R. Halmos. *Measure Theory*. Number 18 in Graduate Texts in Mathematics. Springer New York, 1976.
[Hal15] Brian Hall. *Lie groups, Lie algebras, and representations: an elementary introduction*, volume 222. Springer, 2015.
[Hat02] Allen Hatcher. Algebraic Topology, 2002.
[HH15] John H Hubbard and Barbara Burke Hubbard. *Vector Calculus, Linear Algebra, and Differential Forms: a unified approach*. Matrix Editions, 2015.
[HK03] Boris Hasselblatt and Anatole Katok. *A first course in dynamics: with a panorama of recent developments*. Cambridge University Press, 2003.
[Hop39] Eberhard Hopf. Statistik der geodätischen linien in mannigfaltigkeiten negativer krümmung. *Ber. Verh. Sächs. Akad. Wiss. Leipzig*, 91:261–304, 1939.
[Hop71] Eberhard Hopf. Ergodic theory and the geodesic flow on surfaces of constant negative curvature. *Bulletin of the American Mathematical Society*, 77:863–877, 1971.
[HS74] Morris W. Hirsch and Stephen Smale. *Differential Equations, Dynamical Systems, and Linear Algebra*. Number 60 in Pure and Applied Mathematics. Academic Press, 1974.
[Irw80] Michael C. Irwin. *Smooth Dynamical Systems*. Academic Press, London, 1980.
[Jac99] John David Jackson. Classical electrodynamics, 1999.
[JS87] Gareth A. Jones and David Singerman. *Complex Functions*. Cambridge, 1987.
[Kai97] Vadim A Kaimanovich. Harmonic functions on discrete subgroups of semi-simple Lie groups. *Contemporary Mathematics*, 206:133–136, 1997.
[Kal82] Steven Arthur Kalikow. T, T^{-1} transformation is not loosely Bernoulli. *Annals of Mathematics, Second Series*, 115(2):393–409, 1982.
[Kam82] Tetsuro Kamae. A simple proof of the ergodic theorem using nonstandard analysis. *Israel Journal of Math.*, 42(4):284–290, 1982.
[Kea72] M. Keane. Strongly mixing g-measures. *Invent. Math.*, 16:309–324, 1972.
[Kea91] Michael Keane. *Ergodic theory and subshifts of finite type*, chapter 2. Oxford, 1991.
[Kel75] John L. Kelley. *General Topology*. Graduate Texts in Mathematics. Springer, 1975.
[Kel98] Gerhard Keller. *Equilibrium States in Ergodic Theory*, volume 42 of *Student Texts*. London Math. Soc., 1998.
[KH95] A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems*, volume 54 of *Encyclopedia of Mathematics and its Applications*. Cambridge, 1995.
[KK97] Tetsuro Kamae and Michael Keane. A simple proof of the ratio ergodic theorem. *Osaka Journal of Math.*, 34:653–657, 1997.
[KN32] B.O. Koopman and J. von Neumann. Dynamical systems of continuous spectra. *Proceedings of the National Academy of Sciences*, 18(3):255–263, 1932.
[KN$^+$63] JL Kelley, I Namioka, et al. Linear topological spaces, 1963.
[Kno45] Konrad Knopp. *Theory of Functions, parts 1-2*. Dover Publications, 1945.
[KP82] Elon Kohlberg and John W. Pratt. The contraction mapping approach to the Perron-Frobenius theory: why Hilbert's metric? *Mathematics of Operations Research*, 7(2):198–210, May 1982.
[KP06] Michael Keane and Karl Petersen. Easy and nearly simultaneous proofs of the ergodic theorem and maximal ergodic theorem. *Lecture Notes-Monograph Series*, pages 248–251, 2006.
[KP08] Steven G Krantz and Harold R Parks. *Geometric integration theory*. Springer Science & Business Media, 2008.
[KU07] Svetlana Katok and Ilie Ugarcovici. Symbolic dynamics for the modular surface and beyond. *Bull AMS*, 44(1):87–132, January 2007.
[KW82] Yithak Katznelson and Benjamin Weiss. A simple proof of some ergodic theorems. *Israel Journal of Math.*, 42(4):291–296, 1982.

[Lam66] J. Lamperti. *Probability*. Benjamin-Cummings, 1966.
[Lan99] Serge Lang. *Complex Analysis*, volume 103 of *Graduate Texts in Mathematics*. Spriger, 1999.
[Lan01] S Lang. Real and functional analysis third edition springer verlag 1993, 2001.
[Lan02] Serge Lang. *Introduction to differentiable manifolds*. Springer Science & Business Media, 2002.
[Led74] F. Ledrappier. Principe variationnel et systemes dynamiques symboliques. *Z. Wahr.*, 30:185–202, 1974.
[Liv71] A Livsic. Some homology properties of usystems. *Math Notes USSR, Acad Sci (Mat Zametki)*, 10:758–763, 1971.
[Liv72] A Livsic. Cohomology of dynamical systems. *Math USSR, Isvestiya*, 6(6):1278–1301, 1972.
[Liv96] C. Liverani. Central Limit Theorem for deterministic systems. In S.Newhouse F.Ledrappier, J.Levovicz, editor, *International Conference on Dynamical Systems, Montevideo 1995, a tribute to Ricardo Mañe*, volume 362 of *Pitman Research Notes in Mathemaics Series*, 1996.
[LM95] Douglas Lind and Brian Marcus. *Symbolic Dynamics and Coding*. Cambridge University Press, 1995.
[Loè77] M Loève. Probability Theory I, 1977.
[Los93] Jérôme E Los. Pseudo-Anosov maps and invariant train tracks in the disc: a finite algorithm. *Proceedings of the London Mathematical Society*, 3(2):400–430, 1993.
[LR04] Chao-hui Lin and Daniel Rudolph. Sections for semiflows and kakutani shift equivalence. *Modern Dynamical Systems and Applications*, page 145, 2004.
[Mag74] W. Magnus. *Noneuclidean tesselations and their groups*. Academic Press, New York, 1974.
[Man74] Anthony Manning. There are no new Anosov diffeomorphisms on tori. *American Jour. of Math.*, 96(3):422–429, 1974.
[Mañ87] Ricardo Mañé. *Ergodic theory and differentiable systems*. Springer Verlag, 1987.
[Man02] Anthony Manning. A Markov partition that reflects the geometry of a hyperbolic toral automorphism. *Trans. AMS*, 354:2865–2895, 2002.
[Mar74] Jerrold E. Marsden. *Elementary Classical Analysis*. W. H. Freeman, 1974.
[Mas88] Bernard Maskit. *Kleinian Groups*. Number 287 in Grundlehren der Matematischen Wissenschaften. Springer, 1988.
[Mas91] William S Massey. *A Basic Course in Algebraic Topology*, volume 127 of *Graduate Texts in Mathematics*. Springer Verlag, 1991.
[MH87] Jerrold E. Marsden and Michael J. Hoffman. *Basic Complex Analysis*. W. H. Freeman, second edition, 1987.
[MH98] JE Marsden and JM Hoffman. Basic complex analysis, 3ª edição, 1998.
[MI87] Masahiro Mizutani and Shunji Ito. Dynamical systems on dragon domains. *Japan Journal of Applied Mathematics*, 4(1):23–46, 1987.
[MR20] Jeovanny Muentes and Raquel Ribeiro. Expansiveness, shadowing and markov partition for anosov families. *arXiv preprint arXiv:2007.07424*, 2020.
[MU03] Dan Mauldin and Mariusz Urbański. *Graph Directed Markov Systems: Geometry and Dynamics of Limit Sets*. Cambridge University Press, 2003.
[Nel59] Edward Nelson. Regular probability measures on function space. *Annals of Mathematics*, pages 630–643, 1959.
[Neu32a] J v Neumann. Physical applications of the ergodic hypothesis. *Proceedings of the National Academy of Sciences*, 18(3):263–266, 1932.
[Neu32b] J v Neumann. Proof of the quasi-ergodic hypothesis. *Proceedings of the National Academy of Sciences*, 18(1):70–82, 1932.
[Nit71] Zbigniew Nitecki. *Differentiable Dynamics*. MIT Press, 1971.
[O'N06] Barrett O'Neill. *Elementary differential geometry*. Elsevier, 2006.
[Orn73] D. Ornstein. *Ergodic Theory, Randomness and Dynamical Systems*. Yale Mathematical Monographs. Yale Univ. Press, New Haven, 1973.

[OW83] Donald Ornstein and Benjamin Weiss. The Shannon-McMillan-Breiman theorem for amenable groups. *Israel Journal of Math.*, 44:53–60, 1983.
[Oxt80] John C. Oxtoby. *Measure and Category*. Springer Verlag, second edition, 1980.
[Par64] W. Parry. Intrinsic Markov chains. *Trans. AMS*, 112:55–66, 1964.
[PdM82] Jaco Palis and Wellington de Melo. *Geometric Theory of Dynamical Systems*. Springer Verlag, 1982.
[Pet89] Karl E Petersen. *Ergodic Theory*, volume 2. Cambridge University Press, 1989.
[Phe01] Robert R. Phelps. *Lectures on Choquet's theorem*, volume 1757. Springer Science & Business Media, 2001.
[PP90] William Parry and Mark Pollicott. Zeta Functions and the Periodic Orbit Structure of Hyperbolic Dynamics. *Astérisque*, 187-188:1–268, 1990.
[PP97] William Parry and Mark Pollicott. The Livsic cocycle equation for compact lie group extensions of hyperbolic systems. *J. London Math. Soc.*, 56(2):405–416, 1997.
[PUZ89] Feliks Przytycki, Mariusz Urbanski, and Anna Zdunik. Harmonic, Gibbs and Hausdorff measures on repellers for holomorphic maps, I. *Annals of mathematics*, pages 1–40, 1989.
[Rat73] M. Ratner. The Central Limit Theorem for geodesic flows on n-dimensional manifolds. *Is. Jour. Math.*, 16:181–197, 1973.
[Ros68] Maxwell Rosenlicht. Liouvilles theorem on functions with elementary integrals. *Pacific Journal of Mathematics*, 24(1):153–161, 1968.
[Roy68] H.L. Royden. *Real Analysis*. Macmillan, second edition, 1968.
[RS72] Mike Reed and Barry Simon. *Functional Analysis*, volume 1 of *Methods of Mathematical Physics*. Academic Press, 1972.
[Rud70] Walter Rudin. *Real and Complex Analysis*. McGraw-Hill, 1970.
[Rud73] W. Rudin. *Functional Analysis*. McGraw-Hill, New York, 1973.
[Rud87] Walter Rudin. *Real and Complex Analysis*. Tata McGraw-Hill, 1987.
[Sal] Dietmar Salamon. Notes on compact lie groups. Lecture Nots 2013, ETH.
[Sam56] H. Samelson. On the Perron-Frobenius Theorem. *Mich. Math. J.*, 4:57–59, 1956.
[Sam12] Hans Samelson. *Notes on Lie algebras*. Springer Science & Business Media, 2012.
[Sen81] E. Seneta. *Non-negative Matrices and Markov Chains*. Springer, second edition, 1981.
[Shi73] P. Shields. *The Theory of Bernoulli Shifts*. Univ. of Chicago Press, Chicago, 1973.
[Shi87] Paul Shields. The ergodic and entropy theorems revisited. *IEEE Trans. Inform. Th.*, IT-33:263–266, 1987.
[Shi96] Paul Shields. *The Ergodic Theory of Discrete Sample Paths*, volume 13 of *Graduate Studies in Mathematics*. AMS, Providence, RI, 1996.
[Shu85] Michael Shub. Endomorphisms of compact differentiable manifolds. *Amer. Jour. Math.*, 91:175–199, 1985.
[Shu87] Michael Shub. *Global Stability of Dynamical Systems*. Springer Verlag, 1987.
[Sig66] Laurence E Sigler. Exercises in set theory. 1966.
[Sil17] Ricardo Ramos Silva. *Existencia de uma particao de Markov nao estacionariia do tipo Manning para familias Anosov no toro*. PhD thesis, Universidade de Sao Paulo, 2017.
[Sin76] Ya. G. Sinai. *Introduction to Ergodic Theory*. Number 18 in Mathematical Notes. Princeton, 1976.
[Sot79] J Sotomayor. Liçoes de equaçoes diferenciais ordinárias projeto euclides impa, 1979.
[Spi65] Michael Spivak. *Calculus on Manifolds*, volume 1. WA Benjamin New York, 1965.
[Spi79] Michael Spivak. *A Comprehensive Introduction to Differential Geometry vol 1*. Publish or Perish, 1979.
[SR73] Arthur A Sagle and WALDE RE. Introduction to lie groups and lie algebras. 1973.
[SS85] Michael Shub and Dennis P. Sullivan. Expanding endomorphisms of the circle revisited. *Ergod. Th. and Dynam. Sys.*, 5:285–289, 1985.
[Sul87] Dennis P. Sullivan. Differentiable structures on fractal-like sets, determined by intrinsic scaling functions on dual Cantor sets. *AMS Proc. Symp. Pure Math.*, 48:15–23, 1987.

[SW63] Claude E. Shannon and Warren Weaver. *The Mathematical Theory of Communication*. Univ. of Illinois Press, Urbana, Chicago, London, 1963. Appendix 4.
[Tay96] Michael Eugene Taylor. *Partial differential equations. 1, Basic theory*. Springer, 1996.
[Tes09] Gerald Teschl. Mathematical methods in quantum mechanics. *Graduate Studies in Mathematics*, 99:106, 2009.
[Thu97] William P. Thurston. *Three-Dimensional Geometry and Topology*, volume 1. Princeton, 1997. editor Silvio Levy.
[VE21] Marcelo Viana and José M Espinar. *Differential Equations: A Dynamical Systems Approach to Theory and Practice*, volume 212. American Mathematical Society, 2021. Portuguese version and course web page.
[Ver94] Anatoly M. Vershik. Locally transversal symbolic dynamics. *Algebra i Analiz*, 6:94–106, 1994. in Russian.
[Ver95] Anatoly M. Vershik. Locally transversal symbolic dynamics. *St. Petersburg Math Journal*, 6:529–540, 1995. Translation of 1994 Russian version.
[Via97] M. Viana. Stochastic behaviour of deterministic attractors. *Colóquio Brasileiro de Matemática (IMPA)*, 1997.
[Via06] Marcelo Viana. Dynamics of interval exchange transformations and Teichmuller flows. lecture notes, August 2006. http://w3.impa.br/ viana/out/ietf.pdf.
[Wal75] P. Walters. Ruelle's operator theorem and g-measures. *Trans. AMS*, 214:375–387, 1975.
[Wal82] P. Walters. *An Introduction to Ergodic Theory*. Springer Verlag, New York/Berlin, 1982.
[Wal89] Peter Walters. Book review, ergodic theory of random transformations by Yuri Kifer. *Bull. Amer. Math. Soc.*, 21:113–117, 1989. DOI: https://doi.org/10.1090/S0273-0979-1989-15782-0.
[War71] Frank W Warner. *Foundations of differentiable manifolds and Lie groups*, volume 94 of *Graduate Texts in Mathematics*. Springer Verlag, 1971.
[Whi15] Hassler Whitney. Geometric integration theory. In *Geometric Integration Theory*. Princeton university press, 2015.
[Wic02] AW Wickstead. Vector and banach lattices. *Pure Mathematics Research Centre, Queens University Belfast*, 2002.
[Wil74] Robert F Williams. Expanding attractors. *Publications Mathématiques de l'IHÉS*, 43:169–203, 1974.
[Woj83] Marek Wojtkowski. Invariant families of cones and Lyapunov exponents. *Ergod. Th. and Dynam. Sys.*, 5:145–161, 1983.
[Woj86] Marek Wojtkowski. On uniform contraction generated by positive matrices. *Contemporary Mathematics*, 50:109–118, 1986.
[Zim84] Robert J Zimmer. *Ergodic Theory and Semisimple Groups*, volume 81. Birkhäuser, 1984.

Albert M. Fisher, Dept Mat IME-USP, Caixa Postal 66281, CEP 05315-970 São Paulo, Brazil
URL: `http://ime.usp.br/∼afisher`
E-mail address: `afisher@ime.usp.br`