LECTURE NOTES FOR VECTOR CALCULUS (CALC 2)

ALBERT M. FISHER

Contents

Date: August 6, 2024.

1. Introduction: Mathematics and Science; On writing and reading mathematics; how to study math

Mathematics is the "language of science". This means that the phenomena, the rules and the results of science experiment and theory are expressed in terms of mathematics. Thus, the laws of physics, for example the laws of gravitation, electricity and magnetism are written by means of mathematical formulas. The basic math needed there is calculus, both one-dimensional (in the real numbers $\mathbb{R}$) and in n-dimensional *Euclidean space* $\mathbb{R}^n$. The latter is *vector* calculus, the subject of this course.

All of this is built upon Linear Algebra, since that is basic to understanding Euclidean space. We emphasize first a proper understanding of basic Linear Algebra, as necessary for a true understanding rather than just a manipulation of seemingly interchangeable formulas.

Once we move to the study of waves, including in Quantum Mechanics, and also of heat flow, we extend from Euclidean space to function spaces, which are infinite dimensional vector spaces. So Linear Algebra is also essential there. The study of Linear Algebra in infinite dimensions is known as *Functional Analysis*.

Also fundamental to science are *differential equations* : *ordinary differential equations* (ODE) which includes *systems* of ODE (this can also be called *vector ODE*), and *partial differential equations* (PDE). PDE is essentially the study of ODE in infinite dimensional spaces.

So far this is just referring to physics, but remarkably, in all branches of science, the same mathematical theories come into play and indeed are essential.

But what are the key differences between math and science? In science one tries to find *fundamental laws*, from which in principle all the behavior can be derived. Examples in physics are Newton's laws of mechanics and of gravity; Einstein's Special Relativity; the foundations of Quantum Mechanics; the models of fundamental particles. But in physics there are also higher level laws like for thermodynamics and fluid flow. In Chemistry there are rules, or laws like the Periodic Table; chemical equations... In Biology one has DNA and Natural Selection; the way cells and viruses interact, the study of the brain and of intelligence, including Arificial Intelligence. One also tries to find useful basic laws in Psychology, Botany, Climatology, Economics and so on.

Probably most physicists believe that it should be possible to find precise mathematical expressions for all the fundamental physical laws, though at present we are still far off from this goal. And even if given a perfect set of laws, to actually derive the consequences of those rules, although "in principle" possible, is practically not achievable because of the immense amount of computation needed even to get approximate predictions. Nevertheless, incredible progress in many areas has been made, since say the mid 1800s. Of course this has been enormously accelerated in recent decades by the exponential advancements in computing power, including AI. However it should be kept in mind that AI is no replacement for a deep understanding by a human being. Just as no one would care to watch a robot Simone Biles perform gymnastics, or a robot soccer match (except as a joke, when they knock each other over), so AI-aided research is at best one more tool which can only be useful in the hands of a researcher who actually deeply understands, for themselves, the necessary math and science. And for that, there is no substitute for actual in-person courses, hard individual study, and interactions with others.

An analogy to math is that the fundamental physical laws are reminiscent of the *axioms* of a mathematical system. There are however big differences here. From the axioms one can in principle derive all the true statements for the system; in other words, one can prove all *theorems*; we are skipping over the caveat here due to the results in Logic of Goedel and others. In the next section, we explain the axiomatic approach. The physical laws say Newton's laws for gravity and mechanics can be taken as axioms for those theories, but they are expressed in terms of Calculus which is based on the axioms of mathematics, and which goes far beyond Physics in its applications.

Mathematical axioms are *assumed*, that is they are defined to be true, and one derives consequences (theorems) via the laws of Logic. By contrast, physical laws must be justified by *experiment*. So experimental evidence in some sense takes the place of proofs, or rather it provides the ultimate argument that both the laws and the consequences derived from them are correct.

Thus when experiments showed Newton's laws to not be quite exact, they were improved by Einstein. But the underlying math did *not* change as once proved, it is true *forever*, and in all possible worlds and universes.

The closest analogy to experiment in math is to study *examples*, including to work out computer graphics simulations. That is how a math researcher comes up with ideas; the next step is to prove them *rigorously*, which means precisely, following at each step the rules of logic. Like experiment, examples in math give a rough idea of what is actually true; the real explanation, and the real understanding, comes from proof.

1.1. **Studying math.** One can say that studying math consists of several aspects:

–the study of examples;

–understanding the definitions and statements of theorems;

–understanding the proofs of the theorems;

–working out exercises. Those are often special cases of the theorems, or simpler examples.

The best way to learn math is to teach it, that is, to explain it to someone else. The next best way is to work out exercises, and the third way is to thoroughly study the theory, in lectures and on one's own. Waching video lectures is, by comparison, of little value. At best, it provides some motivation and inspiration to get back to the actual, hard, work.

1.2. **How books are written; how to read them.** Math books are often written in a "linear" manner, logically presented in the most efficient way possible, with Definitions, then Theorems, the Proofs, followed by Examples and the Exercises.

But that is often not the best way to learn math, starting on Page 1 and preceeding in order. Because, that is *never* the way mathematicians think about, understand, or learn math themselves. Rather, that linear understand comes after years of preparations, as a kind of culmination which solidifies and reinforces the author's own understanding.

Rather, math is usually actually learned much more like a *spiral* rather than a straight line, where we circle around and around and fill in more and more details all the while.

The reason for this is that to learn, it is essential to remain curious, and interested. (It is my own firm belief that math should always be interesting, and fun! So I try to maximize my own enjoyment of learning, by letting my own curiosity run free.)

Thus, for example, if we find the material on Set Theory and Logic at the beginning of these Notes to be intimidating or boring, it may be better to skip ahead to more familiar material.

A reference for this course is a typical Calculus text, and often the best approach there is to first try to do some of the exercises, picking up the necessary definitions and consulting the worked-out examples; only then might we go back and read the whole chapter in order. This has the advantage of challenging us immediately to think, and spurs our curiosity and motivates our desire to master the material.

This text is not written exactly in the traditional linear fashion; rather we mix styles as seems appropriate. But we encourage the reader to jump around! And to then always come back and study what was missed on the first go-round.

1.3. **How to study math.** Learning math is very different from many other subjects, where memorization can play an important role, as in the study of, say, medicine or law. In mathematics we try to for example never memorize equations but rather to learn the concepts so well, to understand the material so deeply, that we can derive the equations freshly each time we need them.

Math is hard and is challenging. And precisely becaue of that it is immensely satisfying! That is one reason why I myself fell in love with it. Also I loved the sudden discovery as a student that everything in math can, in principle, be totally understood, from the bottom up. Furthermore, there are interesting logical or philisophical challenges, like understanding infinity, limits, and paradoxes! Moreover, the kind of rigorous logical training you get from proofs is fantastic preparation for any other intellectual or human endeavor, from stuyng law to economics to politics and History. The satisfaction one can get from finally understanding a difficult result is like no other, and the immense reward from doing original research, or from teaching, has no replacement. Grades are of secondary importance: getting good grades should be a natural side effect of a study aimed at mastery. One should maintain always the attitude that "I can understand it if I work hard enough, no matter what it is", and that attitude will provide incredible rewards. Once you learn something thoroughly, once you really master it, you will never forget it, and even if you forget the details, all that will come back quickly. The goal should never be to memorize, rather to understand, deeply. That is where the true satisfaction lies. Superfical understanding, or leaning only to chase the grade, is a waste of precious time which will never come back, for any of us....Inspiration, curiosity, discipline and the will to take on challenges are the key. Math is infinitely challenging and difficult; no matter how smart you are, you can find problems that will challenge you to your limits. The goal is to get to that point, over and over again. You can do it, we all can. The struggle is the reward, and the best part is when you realize you are really confused, and push yourself just a bit more with the confidence that you *can* do it.

2. Review of Logic

Mathematics is the language of science; and the language of mathematics is Set Theory and Mathematical Logic.

We give here a somewhat informal account of the basics.

A *phrase* is a collection of words (technically an *ordered* collection of words) which make sense. Thus "boy goat dog swim or" is not a phrase as it does not make sense. A *proposition* or a *statement* is a phrase which is a *declarative sentence*; it states or declares something to be true. This can either be true (T) or false (F); if P is a proposition then its *truth value* is either T or F.

Given propositions P, Q, R the we can form more complicated propositions by the use of the *logical symbols* $\sim, \wedge, \vee, \Rightarrow, \forall, \exists$ read "not, and, or, implies, for every, there exists".

We *define* the meaning of these symbols by meas of *truth tables.* The first shows the possible truth values for a single proposition P, the second defines the meaning of *not* ($\sim$). The third shows the possible truth values for two propositions P, Q. The next defines the meaning of *and* ($\wedge$) and *or* ($\vee$).

P
T
F

P	$\sim P$
T	F
F	T

P	Q
T	T
T	F
F	T
F	F

P	Q	$P \wedge Q$	$P \vee Q$
T	T	T	T
T	F	F	T
F	T	F	T
F	F	F	F

Implication $\Rightarrow$ is defined by this truth table:

P	Q	$P \Rightarrow Q$
T	T	T
T	F	F
F	T	T
F	F	T

It is important to note that this meaning of implication does not always accord with usage in common language. However this is *always* the meaning in mathematics! Thus if we start with a false statement, we can reach *any* conclusion: "if Washington, DC is the capital of Brazil then 2+2 = 4" is a true statement, but so is "if Washington, DC is the capital of Brazil then $2+2=5$."

(This is one reason why we don't want to have contradictions in any mathematical theory: for then every proposition, and its opposite, are both true!)

Reverse implication, $(P \Leftarrow Q)$ is defined to mean $(Q \Rightarrow P)$. *If and only if* $(P \Leftrightarrow Q)$ is defined to mean $(P \Rightarrow Q) \wedge (P \Leftarrow Q)$. We also write this as P *iff* Q.

We say two propositions P, Q are *equivalent* iff $P \Leftrightarrow Q$.

We shall now use truth tables to prove the following result: $\sim (P \vee Q) \Leftrightarrow (\sim P \wedge \sim Q)$.

P	Q	$\sim P$	$\sim Q$	$(P \vee Q)$	$\sim (P \vee Q)$	$(\sim P \wedge \sim Q)$	$\sim (P \vee Q) \Leftrightarrow (\sim P \wedge \sim Q)$
T	T	F	F	T	F	F	T
T	F	F	T	T	F	F	T
F	T	T	F	T	F	F	T
F	F	T	T	F	T	T	T

Exercise 2.1. Prove, similarly: $\sim (P \wedge Q) \Leftrightarrow (\sim P \vee \sim Q)$.

Try to think of "practical" examples of these results, with propositions from either mathematics or day-to-day life.

In summary, we have shown that $\sim (P \vee Q) \Leftrightarrow (\sim P \wedge \sim Q)$ is *always* true, no matter what are the truth values of P, Q. This means the two propositions are *equivalent* and that this statement is a *theorem* of logic. Another word for a logical theorem is a *tautology*: precisely, if R is a more complicated proposition defined using P, Q, the R is a *tautaulogy* or *logical theorem* iff it is *always true*, i.e. it has the following truth values:

P	Q	R
T	T	T
T	F	T
F	T	T
F	F	T

It is a *contradiction* iff *always false*, i.e. it has the truth values:

P	Q	R
T	T	F
T	F	F
F	T	F
F	F	F

Exercise 2.2. Prove using truthtables that $P \wedge P$ is a tautology while $P \wedge \sim P$ is a contradiction.

Exercise 2.3.

(i) Prove, using truthtables, the logical theorems

$$P \wedge (Q \vee R) \equiv (P \wedge Q) \vee (P \wedge R)$$

and

$$P \vee (Q \wedge R) \equiv (P \vee Q) \wedge (P \vee R)$$

These are known as *distributive laws* (of $\wedge$ over $\vee$, and of $\vee$ over $\wedge$).
(ii) Prove, using truthtables, the logical theorem that

$$(P \Rightarrow Q) \Leftrightarrow (\sim Q \Rightarrow \sim P)$$

.

Definition 2.1. The *converse* of the statement $(P \Rightarrow Q)$ is $(Q \Rightarrow P)$.

It is *not* usually the case that both are true; indeed that happens iff they are equivalent! However we have the following relationship:

Exercise 2.4.
(ii) Prove, using truthtables, the logical theorem that

$$(P \Rightarrow Q) \Leftrightarrow (\sim Q \Rightarrow \sim P)$$

.

The statement $(\sim Q \Rightarrow \sim P)$ is known as the *contrapositive* of $(P \Rightarrow Q)$
One often proves theorems in this way,

Exercise 2.5.
(iii) Prove, using truthtables, that

$$(P \Rightarrow Q) \Leftrightarrow (\sim (P \wedge \sim Q))$$

.

This is known as *proof by contradiction.*

The explanation is as follows. To prove $(P \Rightarrow Q)$, suppose we find out that, assuming P, the conclusion Q is false. This means we assume $\sim Q$ (is true). If $\sim Q$ together with our hypothesis P leads to a contradiction, this means that $(P \wedge \sim Q)$ is false. Equivalently, $(\sim (P \wedge \sim Q))$ is true, and our logical theorem tells us that, equivalently, $(P \Rightarrow Q)$ is true.

Both proof by contrapositive and proof by contradiction are very often used in mathematics. What is fascinating is that even though these statements are logically equivalent to $(P \Rightarrow Q)$, they may be much easier to understand intuitively or to verify. We shall see examples below.

We will define $\exists$ and $\forall$, once we have a bit of set theory which will provide us with some motivating examples.

3. Review of Set Theory

A *set* is a collection of objects. This means that given some *object* x, and a set A, since we know what objects are in A, we know whether x is in A or not.

That is: given A and x, exactly one of these statements is true:

–$x \in A$ or

–$x \notin A$.

Thus a set is exactly specified by what are its elements.

Another name for an element of A is a *member* of A. Another name for an element of a set is an *object* or a *point* in the set.

We can *define* a set by means of a sentence S containing a variable x, as follows: we write

$A = \{x : S(x)\}$ which is read: "A is the set of all x such that $S(x)$ is true."

For example, the collection of all students in the course Calc 2 could be defined as: $A = \{x : x \text{ is enrolled in Calc 2}\}$.

Here is another example. Given two sets A, B we define the intersection and union as follows: $A \cap B = \{x : x \text{ is is in either set } A \text{ or set } B\}$, in other words, $A \cap B = \{x : (x \in A) \vee (x \in B\}$.

Similarly, $A \cup B = \{x : (x \in A) \wedge (x \in B\}$.

Exercise 3.1. Prove, making use of the logical theorems we proved above, that

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

and

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

These are the *distributive laws*, for intersection over union, and union over intersection, respectively, analogous to the distributive law of multiplication over addition for the real numbers: $a(b + c) = ab + ac$.

Definition 3.1. Given sets A, B we say A is a *subset* of B, written $A \subseteq B$, iff $x \in A \Rightarrow x \in B$.

Sometimes we can define a set by listing its elements. Thus $A = \{0, 1, 2, 3\}$ is a subset of $B = \{0, 1, 2, 3, 4\}$ and also of the infinite set of natural numbers $\mathbb{N} = \{0, 1, 2, \dots\}$ (which we discuss in more detail below).

We write $\#A$ for the number of elements in A. Thus $\#A = 4$, and $\#\mathbb{N} = \infty$. However it will be more correct to write $\#\mathbb{N} = \omega$, as it turns out that there are some infinite sets that are bigger than others! This is one of the most fascinating aspects of Set Theory, which we explain below.

Definition 3.2. The simplest set is the *empty set* $\emptyset$, also written as $\{\}$. This is defined to be the set which contains no elements. Thus $(x \notin A)$ is true for any object x, and $\#\emptyset = 0$.

It is interesting that $\emptyset$ can be defined as follows, since for any x, $x \neq x$ is false:

$$\emptyset = \{x : x \neq x\}.$$

Exercise 3.2. Prove that $\emptyset$ is a subset of every set A.

We define two sets A, B to be *equal* iff they have exactly the same elements. This just agrees with our definition of set as a collection of objects.

Thus: $A = B \Leftrightarrow (x \in A) \Rightarrow (x \in B)$ and also $(x \in B) \Rightarrow (x \in A)$. Equivalently $A = B$ iff $A \subseteq B)$ and $B \subseteq A)$. We often prove two sets are equal in this way.

Exercise 3.3. Prove, making use of the logical theorems we proved above, that

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

and

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

These are the distributive laws, for intersection over union, and union over intersection, respectively.

Definition 3.3. Given a set X, we say X is the *universe* (of discussion) if we shall only be talking for the moment about subsets of X. Thus if we are studying statistics, and choosing points form the set $X = \{0, 1, 2, 3, 4\}$, then this is our universe. If we are studying Calculus, our universe may be the real numbers $\mathbb{R}$ or three-dimensional space $\mathbb{R}^3$.

Such a universe X can also be termed a *relative universe* as we are studying things relative to this restriction.

Given a relative universe X, and a subset $A \subseteq X$, the *complement* of A is $A^c = \{x : (x \in X) \wedge (x \notin A)\}$. This is also called the *complement of A in X*, or the *complement of A relative to X*.

Exercise 3.4. Given a universe X, prove, using the theorems of logic shown above, that for $A, B \subseteq X$,

$$(A \cup B)^c = A^c \cap B^c$$

and

$$(A \cap B)^c = A^c \cup B^c.$$

Definition 3.4. These very important theorems are known as *de Morgan's laws*.

Remark 3.1. Notice how the logical statements and the set theory statements parallel each other! Here is another example:

Exercise 3.5. Prove that

$$(A \subseteq B) \Leftrightarrow (B^c \subseteq A^c).$$

3.1. **What is a "proposition", and what is an "axiom", really?** In mathematics, everything has to be defined *rigorously*, which is math-talk for: "absolutely precisely". For each area of mathematics, we carry this out, always specifying the particular mathematical theory we will be dealing with.

Example 1. For a first example, let us consider Set Theory itself. We begin with the most simple phrase possible, which is $x \in A$. We say this as "x is an element of the set A" but actually as we see below, x can itself be a set, with say an element a, and A can be an element of a further set, say B. In this case we have $a \in x$, $x \in A$, and $A \in B$ which we can summarize as: $a \in x \in A \in B$.

Indeed, in classical Set Theory, *all* elements will in fact themselves be sets. This is the case for example in the excellent book by Halmos, [Hal74].

We describe the axioms for we might call "very very naive Set Theory" as contrasted with the *Zermelo-Frankel* Set Theory presented by Halmos. Another name is *Cantor's Set Theory* as it was proposed by Georg Cantor, the inventor in the period 1874-1884 of Set Theory.

(1) Given a set A and x, then $(x \in A) \vee x \notin A$ but not both.

This tells us when two sets are equal. We then have a second Axiom:

(2) Given a statement S, containing a variable x, whose truth value depends on what x is, then $A = \{x : S(x)\}$ is a set.

From this, as explained above, we can define these concepts: emptyset, union, intersection, subset, complement. Cantor's Set Theory could be used as a foundation for nearly all of mathematics, and greatly clarified all areas of mathematics.

However, as we shall see below, Cantor's Axioms although they apparently make perfect sense, are (to everyone's surprise at the time) *not* coherent, as they lead to a contradiction.

That is why they were replaced by the Zermelo-Frankel and other lists of axioms.

Example 2. For a second example, take the study of vector spaces: Linear Algebra. Here we begin with Set Theory, and then define a vector space to be a set of a special type, with properties given by the axioms of vector spaces.

The axioms are *theorems*. That is, they are statements which we assume to be true for our set. Then, using the laws of logic, we construct *proofs* of all the other true statements of the theory, called the *Theorems*, *Propositions* and *Lemmas* of Linear Algebra.

It is important to know that the axioms of a theory are *coherent*, meaning that you cannot arrive, by proof, at a contradiction.

In principle, it should be possible to *prove* this! However in practice finding such a proof can be difficult, depending on deep ideas of Set Theory and of Logic.

In fact, for Cantor's version of Set Theory this is impossible, as one can prove a contradiction! This famous result, due to the British mathematician and philosopher Bertrand Russell, in 1901, threw the world of mathematics into a turmoil and led to much further remarkable work. This is known as *Russell's Paradox* and will be described below.

Furthermore, on an even more fundamental level, a list of axioms should be found for mathematical logic itself, and again, it should be shown that these are coherent. This question is also *very* subtle and interesting, as shown by work of Goedel and others. The study of these matters is known as the *foundations of mathematics*.

3.2. **Russell's Paradox.**

3.3. **Orders of infinity; irrational numbers.**

4. Review of numbers, and of Linear Algebra

Let us recall:

Definition 4.1. In this course we write $\mathbb{N}$ for the for the *natural numbers*, $\mathbb{N} = \{0, 1, 2, \dots\}$ and $\mathbb{Z}$ for the *integers*, $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$. If we want only the positive integers, we denote this by $\mathbb{N}^* = \{1, 2, \dots\}$. (There is a difference of notation here depending on the country and the mathematician). The *prime numbers* are $\{1, 2, 3, 5, 7, 11, 13, 17, \dots\}$; again there is a difference of custom as many number theorists begin the primes with 2, but as someone working in dynamical systems theory I prefer these choices! (*Conventions* like this- meaning mathematical customs - are not really important as long as one is clear about one's definitions when writing a book or article!) $\mathbb{Q}$ denotes the rational numbers, $\mathbb{Q} = \{m/n : \ m, n \in \mathbb{Z}, n \neq 0\}$. $\mathbb{Z}$ and $\mathbb{Q}$ are both *rings*: there are two operations on it, addition and multiplication, denoted by $+$ and $\cdot$, satisfying the following axioms:
(1) There exists a unique number 0 such that $0 + x = x$ *(existence of an identity element for +)*
(2) $x + y = y + x$ *(the commutative law for +)* We note that another name for commutative is *abelian*.
(3) $(x + y) + z = x + (y + z)$ *(the associative law for +)*
(4) $1 \cdot x = x$ *(existence of an identity element for ·)*
(5) $x \cdot (y \cdot z) = (x \cdot y) \cdot z$ *(the associative law for ·)*
(6a) $x \cdot (y + z) = x \cdot y + x \cdot z$
(6b) $(x + y) \cdot z = x \cdot z + y \cdot z$ (the two *distributive laws*).
For a *commutative ring* the multiplication is also commutative.

Remark 4.1. We can either assume the identity element 0 is unique as in (1) and prove from that:
(1') there exists an additive inverse for each x, denoted $-x$, or we can drop the uniqueness assumption and replace it by $(1')$ and then *prove* uniqueness from that.

A *field* is a set K, with two operations $+$ and $\cdot$, which is a commutative ring and such that also:
(8) each $x \neq 0$ has a *multiplicative inverse*: there exists $\tilde{x}$ such that $x \cdot \tilde{x} = 1$. We write x^{-1} for this number.

The basic examples to keep in mind of fields and the only ones we use in this course, are $\mathbb{R}$ (the real numbers) and $\mathbb{C}$ (the complex numbers) Other examples are $\mathbb{Q}$ (the rational numbers) and the finite fields $\mathbb{Z}_p$ for p prime, and *extension fields* of $\mathbb{Q}$ such as $\mathbb{Q}[\sqrt{2}]$ (see any algebra text). There are few fields, and many rings! A *group* G is a set with only one operation which is associative but not necessarily commutative, and with inverses. This is usually called *multiplication* and is written $\cdot$; however if the group is commutative this operation may be written $+$. Thus the fields are commutative groups for $+$ but also their non-zero elements form a commutative group for multiplication. Any ring is a group for addition. $\mathbb{N}$ with addition gives an example of a *semigroup* as it does not have additive inverses.

There are many, many more groups than rings. Other examples are invertible $(n \times n)$ matrices with the operation of matrix multiplication. The *quaternions* are a four-dimensional example of something which is not quite a field, as multiplication is noncommutative, and the *octonions* are an eight-dimensional example which is even worse: multiplication is not even associative! Both of these find uses in Physics, and both can be represented by certian collections of matrices.

Abstract Algebra is the study of all these sorts of things: groups, rings, fields, vector spaces and much, much more. Each of these can be called an *algebraic theory.* For example, *group theory* consists of all the definitions and theorems for groups.

The basic idea of any mathematical theory is that we study certain *objects* (usually sets) satisfying certain properties called *axioms.* The axioms are the most basic true statements. We use calculations and intuition to try to guess at new true statements, and then endeavor to use the rules of logic to prove all these statements, called *theorems.* (Thus the axioms themselves are also theorems!)

One of the most important and useful algebraic theories (in math, as well as in physics) is the theory of *vector spaces*, called *Linear Algebra.*

What is magnificent about the axiomatic approach to math is that any theorem proved is then valid in all examples, which at first sight may seem completely different. So it is a very powerful approach. Also, many things become much clearer when presented this way.

For example, theorems proved for real vector spaces are valid (with the same proofs) for complex vector spaces, and of any dimension, including infinite dimension (although new axioms are introduced in those cases, to handle other issues: Hermitian inner products; convergence of infinite series in the case of infinite dimensions....)

We now give a brief introduction to Linear Algebra. An excellent text is [Axl97].

Definition 4.2. A *vector space* (sometimes called a *linear space*) *over the real numbers* $\mathbb{R}$ (sometimes over a different *field*, such as the complex numbers $\mathbb{C}$ or the rational numbers $\mathbb{Q}$) is a set V with two operations $+, \cdot$ (addition of vectors, multiplication of a vector by a real number) satisfying these properties, called the *vector space axioms.* The elements of the field are also called *scalars.*
(1) There exists a unique vector $\mathbf{0}$ such that $\mathbf{0} + \mathbf{v} = \mathbf{v}$ *(existence of an identity element for +)*
(2) $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$ *(the commutative law for +)*
(3) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ *(the associative law for +)*
(4) $0 \cdot \mathbf{v} = \mathbf{0}$ *(multiplication by* 0*)*
(5) $1 \cdot \mathbf{v} = \mathbf{v}$ *(existence of an identity element for* $\cdot$*)*
(6a) $a \cdot (\mathbf{v} + \mathbf{w}) = a \cdot \mathbf{v} + a \cdot \mathbf{w}$
(6b) $(a + b) \cdot \mathbf{v} = a \cdot \mathbf{v} + b \cdot \mathbf{v}$ *(two distributive laws).*

Note that the associatve law allows us to define the sum of more than two vectors, as $\mathbf{u} + \mathbf{v} + \mathbf{w} + \mathbf{x} = ((\mathbf{u} + \mathbf{v}) + \mathbf{w}) + \mathbf{x} = (\mathbf{u} + \mathbf{v}) + (\mathbf{w} + \mathbf{x})$ and so on.

Exercise 4.1. Show that for each $\mathbf{v} \in V$ there exists a vector $\tilde{\mathbf{v}}$ such that $\tilde{\mathbf{v}} + \mathbf{v} = \mathbf{0}$: *(the existence of an additive inverse.)* Show that this is *unique.* See Remark 4.1 above!

Example 3. *Arrows in the plane, based at a point* $\mathbf{0}$. This is a purely geometrical definition. We are given a geometrical plane; we know what lines are and what length is. We draw a line segment from a point labelled $\mathbf{0}$ to another labelled $\mathbf{v}$. The arrowhead is located at $\mathbf{v}$. The collection of all such arrows is V. We define our operations: to form $\mathbf{u} = \mathbf{v} + \mathbf{w}$ we move in parallel the second vector $\mathbf{w}$ to the tip of $\mathbf{v}$; this gives the tip of $\mathbf{u}$.

Multiplication by the real number $a > 0$ multiplies the length by a, and by -1 gives us the opposite arrow.

Exercise 4.2. A *linear combination* of two vectors $\mathbf{v}, \mathbf{w}$ is a sum $a\mathbf{v} + b\mathbf{w}$. A linear combination of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$. We also write this as $\sum_1^n a_i\mathbf{v}_i$. Draw the picture for arrows, for a linear combination of 2 and of 3 vectors. Verify the axioms for V geometrically (by drawing pictures). No coordinates are used. Note that the same pictures work for arrows in 3 dimensions.

Example 4. *Euclidean space* $\mathbb{R}^n$. The points are n-tuples $\underline{x} = (x_1, x_2, \ldots, x_n)$. We define addition in $\mathbb{R}^n$ by adding coordinates, similarly for multiplication by a scalar. Thus $(a, b) + (c, d) = (a + c, b + d)$ and $t(a, b) = (ta, tb)$.

Example 5. $(m \times n)$ *matrices, the space* $M_{m\times n}$. Addition and multiplication are defined coordinate-by-coordinate.

Exercise 4.3. Show the axioms hold for the last two examples.

Example 6. *A function space.* We define $\mathcal{C}^0([a, b], \mathbb{R})$ to be the collection of all continuous functions from the closed interval $[a, b]$ to $\mathbb{R}$. $\mathcal{C}^n([a, b], \mathbb{R})$ for $n \geq 1$ denotes the function with n derivatives such that the n^{th} derivative $f^{(n)}$ is continuous. We also write $\mathcal{C}$ for $\mathcal{C}^0$.

Exercise 4.4. Define addition and multiplication by a scalar in an appropriate way for the last example, and show the axioms hold.

Definition 4.3. Given two vector spaces V, W, a map (i.e. a function) T from V to W is called a *linear transformation* iff it preserves the operations, that is, iff
(i) $T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$;
(i) $T(a\mathbf{v}) = aT(\mathbf{v})$.

We can summarize this by:
(L) $T(a\mathbf{v} + b\mathbf{w}) = aT(\mathbf{v}) + bT(\mathbf{w})$. We call property (L) *linearity.*

Exercise 4.5. Show that the map from the space of arrows V to $\mathbb{R}^2$ defined by sending an arrow $\mathbf{v}$ with tip at location (a, b) to $(a, b) \in \mathbb{R}^2$ is a linear map. Show it is *bijective* i.e. it is $1 - 1$ and onto. This is called an *isomorphism* of the vector spaces. Isomorphic spaces like this are not the same, but we say they can be *identified* via the isomorphism.

Exercise 4.6. Show that $\mathbb{R}^n$ is isomorphic to the space of *column vectors* $M_{n\times 1}$ and also to the space of *row vectors* $M_{1\times n}$.

Show that a linear transformation $T : V \to W$ always sends $\mathbf{0}_V$ (the zero element in V) to $\mathbf{0}_W$.

Definition 4.4. A *subspace* U of a vector space V is a subset that is also a vector space, with the same operations. Show that a subset U is a subspace iff for all $\mathbf{v}, \mathbf{w} \in U$, then $\mathbf{v} + \mathbf{w} \in U$ and for any $a \in \mathbb{R}$, then $a\mathbf{v} \in U$. In other words, any linear combination of elements of U remains in U.

Exercise 4.7. Show that the intersection of two subspaces of V is a subspace. Show that this can be false for unions.

Definition 4.5. Given a collection of vectors $S \subseteq V$ we say the *vector space generated by* S is the smallest vector space that contains S. We denote this by $\langle S \rangle$. Show that this makes sense (hint: use the previous exercise). Show that $\langle S \rangle$ is the collection of all linear combinations of elements of S. We call $\langle S \rangle$ the *span* of S. Given a subspace $U \subseteq V$, we say $S \subseteq V$ *spans* or *generates* U iff $\langle S \rangle = U$.

Definition 4.6. A collection $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of vectors in V is *linearly independent* iff
(def 1) $\sum_1^n a_i \mathbf{v}_i = \mathbf{0} \Rightarrow a_i = 0$ for all i.
(def 2) If $\mathbf{u} = \sum_1^n a_i \mathbf{v}_i$ and also $\mathbf{u} = \sum_1^n b_i \mathbf{v}_i$, then $a_i = b_i$ for all i. Thus, any vector $\mathbf{u}$ in the span of the vectors $\mathbf{v}_i$ has unique a_i, called the *coefficients* or *coordinates* of $\mathbf{u}$ with respect to these vectors.

A *basis* for $U \subseteq V$ is an ordered collection S of vectors in U which spans U and is linearly independent. Thus, any $\mathbf{u} \in U$ can be expressed as a linear combination of the vectors $(\mathbf{v}_1, \dots, \mathbf{v}_n)$ in S and this expression is unique. In particular $\mathbf{u}$ has well-defined (i.e. uniquely defined) coordinates in the basis S.

Theorem 4.1. *If V is a vector space and there exists a finite number of vectors which generate V, we say V is* finite-dimensional. *A finite dimensional vector space has a basis. The number of element in this basis is well-defined (does not depend on the choice of the basis).*

Proof. This takes some work: see [Axl97] for a nice proof. □

Definition 4.7. This number is defined to be the *dimension* of V.

Proposition 4.2. *Given a vector space V of dimension n, then V is isomorphic to $\mathbb{R}^n$.*

Proof. We express $\mathbf{u} \in V$ as a linear combination of the basis: $\mathbf{u} = \sum_1^n a_i \mathbf{v}_i$, and define $\Phi : V \to \mathbb{R}^n$ by $\Phi(\mathbf{v}) = (a_1, \dots, a_n)$. (Check that this is an isomorphism). □

Definition 4.8. (Definition of matrix products; see Lectures and handwritten notes.)

Exercise 4.8. (function spaces) Show that the collection of all Riemann integrable functions and the collection of all differentiable functions on $[a, b]$ are subspaces of $\mathcal{C}$. Show that differentiation and definite integrals define linear transformations. Show that the polynomials $\mathcal{P}_n$ of degree $\leq n$, for some $n \geq 0$, form a vector space of dimension $n + 1$. (Hint: find a basis!) Find a matrix which represents the derivative map from $\mathcal{P}_{n+1}$ to $\mathcal{P}_n$. Show the space $\mathcal{C}^0$ has infinite dimension.

Norms and inner products.

Definition 4.9. 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).

Example 7. For V the space of arrows in 2 or 3 dimensional space, the *standard norm* is the length of the line segment. For $\mathbb{R}^n$, the *standard norm* of $\mathbf{a} = (a_1, \dots, a_n)$ is $||\mathbf{a}|| = \left(\sum_1^n a_i^2\right)^{1/2}$. This is also called the *l^2-norm.*

Remark 4.2. The isomorphism from V to $\mathbb{R}^2$ or $\mathbb{R}^3$ preserves the norm.

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

Definition 4.10. 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.* Note that a positive definite bilinear form defines a norm, via

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

Example 8. The *standard inner product* for the space of *arrows*, V is

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

where in the plane θ is the angle from $\mathbf{v}$ to $\mathbf{w}$ measured in the counterclockwise direction.

Note that this would give the same number if we went in the opposite direction, since $\cos(\theta) = \cos(-\theta)$. This is lucky since the same definition works in space, where we don't know what "counterclockwise" means!!

The *standard inner product* for $\mathbb{R}^n$ is $(x_1, \dots, x_n) \cdot (y_1, \dots, y_n) = \sum_1^n x_i y_i$.

Exercise 4.9. Verify that these each satisfy the axioms, and also that they are equal for our isomorphism from V to $\mathbb{R}^2$ or $\mathbb{R}^3$.

Example 9. (Hilbert space). This is an example which comes up and also in any study of waves.

For this $L^2 = L^p([a, b])$ is defined to be the vector space of all $f : [a, b] \to \mathbb{R}$ such that $\int_{[a,b]} |f|^2 \, \mathrm{d}x < \infty$. We define the norm

$$||f||_2 = \left(\int_{[a,b]} |f|^2 \, \mathrm{d}x \right)^{\frac{1}{2}}.$$

Thus L^2 is the space of all function with finite norm.

We define an inner product by $f \cdot g = \langle f, g \rangle = \int_{[a,b]} fg \, \mathrm{d}x$.

Note that this gives the norm as above.

Hilbert space is infinite-dimensional; a basis is given by all functions of the form (taking $[a, b] = [0, 2\pi]$ for simplicity) $\sin(nx), \cos(nx)$. The expansion is called Fourier series. Here we have to use *infinite* linear combinations, and the key point is that the series must converge. For this all to work properly, we should use a fancier integral than the Riemann integral: the *Lebesgue* integral, for which we need to study *Real Analysis* and in particular *Measure Theory.*

In applications to Physics (especially Quantum Mechanics) the field is $\mathbb{C}$, then we have to use the complex conjugate to define a *Hermitian inner product*, which has slightly different axioms.

5. Vector Calculus, Part I: Derivatives and the Chain Rule

5.1. **Metrics, open sets, continuity.** Let us recall:

Definition 5.1. 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).

Given a set X, a *metric* on X is a function $d : X \times X \to [0, +\infty]$ satisfying:
(*i*) $d(x, y) = d(y, x)$ (symmetry);
(*ii*) $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality);
(*iii*) $d(x, y) \geq 0$, and $d(x, y) = 0$ iff $x = y$ (positive definiteness).

We then say that $(V, ||\cdot||)$, respectively (X, d), is a *normed space*, respectively a *metric space.*

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

Exercise 5.1. Verify this!

A *topology* on X is a collection $\mathcal{T}$ of subsets of X (these will be called *open sets*) satisfying:
(*i*) $\emptyset, X \in \mathcal{T}$;
(*ii*) an arbitrary union of open sets is open;
(*iii*) a finite intersection of open sets is open.

A set $C \subseteq X$ is *closed* iff its complement $C^c = X \setminus C = \{x \in X : x \notin C\}$ is open.

The collection $\mathcal{F}$ of closed sets satisfies:
(*i*) $\emptyset, X \in \mathcal{F}$;
(*ii*) an arbitrary intersection of closed sets is closed;
(*iii*) a finite union of closed sets is closed.

These properties of $\mathcal{F}$ are equivalent to the corresponding properties for $\mathcal{T}$ via the laws for unions and intersections of complements of sets:

$(A \cap B)^c = A^c \cup B^c$, $(A \cup B)^c = A^c \cap B^c$

and more generally,

$$(\cap_{i\in I}A_i)^c = \cup_{i\in I}(A_i)^c;$$

$$(\cup_{i\in I}A_i)^c = \cap_{i\in I}(A_i)^c.$$

where I is some index set, for example $\mathbb{N} = \{0, 1, 2, \dots\}$ or even an uncountable set like $\mathbb{R}$.

Exercise 5.2. Verify these statements!

For a metric space a *limit point* of $A \subseteq X$ is $x \in X$ such that for each $r > 0$, there is some point of A in $B_r(x)$. For a metric spaces a set C is closed iff it contains all of its limit points. Thus for example (a, b) is not a closed set as a, b are limit points.

Having a metric allows us to define a *topology* on X, as follows.

Definition 5.2. Given a metric space $(X.d)$, the *(open) ball of radius* $r \in [0, \infty]$ *around* $x \in X$ is $B_r(x) = \{y \in X : d(x, y) < r\}$.

A set $\mathcal{U} \subseteq X$ is *open* iff, equivalently,
(i) $\mathcal{U}$ is a union of open balls;
(ii) for every $x \in \mathcal{U}$, $\exists r > 0$ such that $B_r(x) \subseteq \mathcal{U}$.

Exercise 5.3. Verify that this does give a topology.

Convergence and continuity.

Given a sequence $(x_n)_{\mathbb{N}\in\mathbb{N}}$, we say (x_n) *converges to* x, equivalently written $\lim_{n\to\infty} x_n = x$ or $x_n \to x$, iff for every open set $\mathcal{U}$ containing x, then for n sufficiently large, $x_n \in \mathcal{U}$. For a metric space, equivalently given $r > 0$, $\exists N$ such that for all $n > N$, $d(x_n, x) < r$ (since we can use balls of radius r).

Definition 5.3. Given two metric spaces (X_1, d_1) and (X_2, d_2), then $f : X \to Y$ is *coninuous* iff
(i) if $x_n \to x$ then $f(x_n) \to f(x)$. (This is the usual definition for $f : \mathbb{R} \to \mathbb{R}$.)
Equivalently, iff:
(ii) if $f(x_0) = y_0$, then given $\varepsilon > 0$, $\exists\delta > 0$ such that if $d(x, x_0) < \delta$ then for $y = f(x)$, then $d(y, y_0) < \varepsilon$. (This is the famous "$\varepsilon - \delta$"- definition.)
(iii)the inverse image of every open set is open.
This third definition works for the more general situation of two topological spaces, $(X, \mathcal{T})$ and $(Y, \mathcal{S})$.

Exercise 5.4.
(i) Use each of the three definitions to prove:

Proposition 5.1. *A composition of continuous functions is continuous.*

(ii) Show that (for metric spaces), all three definitions are equivalent. Hint: first prove $(i) \Leftrightarrow (ii)$, then $(ii) \Leftrightarrow (iii)$.

Remark 5.1. For some unusual non- metric topological spaces one has to replace sequences by so-called *nets* or *filters*.

5.2. **Curves.**

Definition 5.4. A *(parametrized) curve* in a vector space V is a function $\gamma : [a, b] \to V$.

The *image* of the curve is the image of this function, i.e. the collection of all values: $\{\gamma(t) : t \in [a, b]\}$. Thus the *parameter* t parametrizes the image.

The simplest example is a *parametrized line*; the curve $l(t) = \mathbf{p} + t\mathbf{v}$ where $\mathbf{p}, \mathbf{v}$ are elements of some vector space V. The image of l is a straight line in V; the parametrized line passes through the point $\mathbf{p}$ going in the direction $\mathbf{v}$.

Note that the image of a curve is different from the *graph*. Here we recall that by definition the *graph* of a function $f : X \to Y$ (where X, Y can be any sets) is $\text{graph}(f) \equiv \{(x, y) \in X \times Y : y = f(x)\} = \{(x, f(x) : x \in X\}$. (Here $X \times Y$ is the *product space*, defined to be the collection of all ordered pairs.)

Thus the graph of the curve γ in V is $\{(t, \gamma(t)) : t \in [a, b]\}$, a subset of $[a, b] \times V$. The image shows where you go on the curve, but not how fast or in what direction you go along this image. We see shortly how you can change the parametrization of a curve, keeping the same image.

Let us suppose that V has a norm defined on it. Then V is a metric space with $d(\mathbf{v}, \mathbf{w}) = ||\mathbf{v} - \mathbf{w}||$, so we know what it means for a curve to be *continuous*. We define the derivative of γ to be

$$\gamma'(t) = \lim_{h \to 0} \frac{\gamma(t + h) - \gamma(t)}{h}$$

if the limit exists; this is also called the *tangent vector* to γ at time t. See Fig. 5.

Lemma 5.2. *We have in coordinates:* $\gamma'(t) = (x_1'(t), \dots, x_m'(t))$.

Proof. This is immediate from the definition. For example, in $\mathbb{R}^2$, for $\gamma(t) = (x, y)(t) = (x(t), y(t))$ then

$$\frac{\gamma(t + h) - \gamma(t)}{h} = \left(\frac{x(t + h) - x(t)}{h}, \frac{y(t + h) - y(t)}{h} \right) \to (x'(t), y'(t)).$$

□

Note that given a differentiable curve $\gamma : [a, b] \to \mathbb{R}^m$ then γ' is a second curve in $\mathbb{R}^m$. We can keep going and define the higher derivatives $\gamma'' = (\gamma')'$ and so on, all curves in $\mathbb{R}^m$.

The most common interpretation of the tangent vector of a curve comes from physics. There we interpret t as time and γ as position:

Definition 5.5. If $\gamma(t)$ represents the position of a particle at time t, then the derivative γ' (the tangent vector) gives the *velocity* of the particle $\mathbf{v}(t) = \gamma'(t)$, and the *acceleration* at time t is the vector $\mathbf{a}(t) = \mathbf{v}'(t) = \gamma''(t)$. Note that all of these are *vector quantities*, having both a magnitude and a direction. The *speed* is the magnitude of the velocity vector, the *scalar quantity* $||\mathbf{v}||$.

Now we prove some basic facts about curves and their derivatives:

Proposition 5.3. *(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 5.4. *Let* γ *be a differentiable curve in* $\mathbb{R}^m$ *such that* $||\gamma|| = c$ *for some constant* c*. Then* $\gamma \perp \gamma'$*.*

Proof.
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.

□

The meaning of $||\gamma|| = c$ is intuitively clear: for $\mathbb{R}^2$ this says that the curve is in a circle; for $\mathbb{R}^3$ that the image of the curve is in a sphere, and the statement is that the tangent vector to the curve is tangent to the sphere as it is perpendicular to the position vector. See Fig. 5.

Corollary 5.5. *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'. □

Corollary 5.6. *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 you apply the brakes or the accelerator pedal, a component vector of acceleration tangent to the curve will be added to this.

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: we explain this next.

Proposition 5.4 allows us to make the following definition.

Definition 5.6. 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.

5.3. **Arc length of a curve.** Given a curve $\gamma_1 : [c, d] \to \mathbb{R}^n$, by a *reparametrization* γ_2 of the curve we mean the following: we have an invertible differentiable function $h : [a, b] \to [c, d]$ with $h'(t) \neq 0$ for all t, 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.

Definition 5.7. We define the *arc length* of γ to be:

$$\int_a^b ||\gamma'(t)||\mathrm{d}t.$$

We introduce a special formula for this:

$$\int_\gamma \mathrm{d}s = \int_a^b ||\gamma'(t)||\mathrm{d}t.$$

As we shall explain below, ds is interpreted to mean *integration with respect to arc length*, and "$\int_\gamma$" is read "the integral over the curve γ", so all together this is read "the integral over the curve γ with respect to arc length".

For an example we already know from Calculus I, consider a function $g : [a, b] \to \mathbb{R}$, we consider its graph $\{(x, g(x)) : a \le x \le b\}$. We know 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 the arc length is $\int_\gamma \mathrm{d}s = \int_a^b \sqrt{1 + (g'(t))^2}\mathrm{d}x$ as claimed.

Proposition 5.7.
(i) The arc length of a curve is unchanged for any change of parametrization, independent of orientation. That is,

$$\int_{\gamma_1} ds = \int_{\gamma_2} ds.$$

Proof. (i) We have $h : [a, b] \to [c, d]$ so writing $u = h(t)$, then $\gamma_2 = \gamma_1 \circ h$ with $\gamma_1 : [c, d] \to \mathbb{R}^n$ and $\gamma_2 : [a, b] \to \mathbb{R}^n$. So $\gamma_2 \equiv \gamma_1 \circ h$ with $\gamma_2(t) = (\gamma_1 \circ h)(t) = \gamma_1(h(t)) = \gamma_1(u)$, so $\gamma_2'(t) = (\gamma_1 \circ h)'(t) = \gamma_1'(h(t))h'(t)$. Then d$u = h'(t)dt$, and using the Chain Rule, we have:

$$\int_{\gamma_2} \mathrm{d}s \equiv \int_{t=a}^{t=b} ||\gamma_2'(t)||\,\mathrm{d}t = \int_{t=a}^{t=b} ||(\gamma_1'(h(t))h'(t)||\,\mathrm{d}t$$

Assuming first that $h' > 0$, this equals (1)

$$= \int_{t=a}^{t=b} ||(\gamma_1'(u)||\,h'(t)\mathrm{d}t = \int_{u=c}^{u=d} ||(\gamma_1'(u)||\,\mathrm{d}u = \int_{\gamma_1} \mathrm{d}s.$$

If instead $h' < 0$, then we have as before

$$\int_{\gamma_2} \mathrm{d}s \equiv \int_{t=a}^{t=b} ||\gamma_2'(t)||\,\mathrm{d}t = \int_{t=a}^{t=b} ||(\gamma_1'(h(t))h'(t)||\,\mathrm{d}t$$

and now because, since $h' < 0$, $h(b) = c, h(a) = d$,

$$= -\int_{t=a}^{t=b} ||(\gamma_1'(u)||\,h'(t)\mathrm{d}t = -\int_{u=d}^{u=c} ||(\gamma_1'(u)||\,\mathrm{d}u = \int_{u=c}^{u=d} ||(\gamma_1'(u)||\,\mathrm{d}u = \int_{\gamma_1} \mathrm{d}s. \tag{2}$$

□

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. Let us denote the length of γ by L. Thus function l maps $[a,b]$ to $[0,L]]$. Note that l is a primitive (antiderivative) for $||\gamma'||$ so $l'(t) = ||\gamma'(t)||$. We shall assume that $||\gamma'(t)|| > 0$ for all t; in this case, the function l is invertible. Our parameter change will be given by its inverse, $h(t) = l^{-1}(t)$; then h' is also positive.

Proposition 5.8. *Assume that $||\gamma'(t)|| > 0$ for all t. Then the reparametrized curve $\widehat{\gamma} = \gamma \circ h$ has speed one.*

Proof. Now $(l \circ h)(t) = t$ so $1 = (l \circ h)'(t) = l'(h(t)) \cdot h'(t)$. We have (by the Fundamental Theorem of Calculus) that $l'(t) = ||\gamma'(t)||$ so $l'(h(t)) = ||\gamma'||(h(t)) = ||\gamma'(h(t))||$ since $h'(t) > 0$. Thus

$$||\widehat{\gamma}'(t)|| = ||(\gamma \circ h)'(t)|| = ||\gamma'(h(t)) \cdot h'(t)|| = ||\gamma'(h(t))|| \cdot h'(t) = l'(h(t)h'(t) = 1.$$

□

The function l maps $[a,b]$ to $[0,L]]$ whence the parameter-change function h maps $[0,L]$ 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 $[0,L]$ 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

$$\int_{\gamma} \mathrm{d}s \equiv \int_a^b ||\gamma'(t)||\,\mathrm{d}t = \int_0^{l(b)} ||\widehat{\gamma}'(s)||\mathrm{d}s \equiv \int_{\widehat{\gamma}} \mathrm{d}s$$

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 $\int_\gamma \mathrm{d}s$ for the total arc length.

5.4. **Level curves of a function.** We would like to visualize a function $F : \mathbb{R}^2 \to \mathbb{R}$. We do this in two main ways, by drawing the **graph** of the function (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 $\mathbb{R}^2$:

$$\{(x,y) : F(x,y) = c\}.$$

Remark 5.2. In geography, a *topographic map* of a region X shows the level curves of the altitude function $F(x,y)$ with $F : X \to \mathbb{R}$. See Fig. 1.

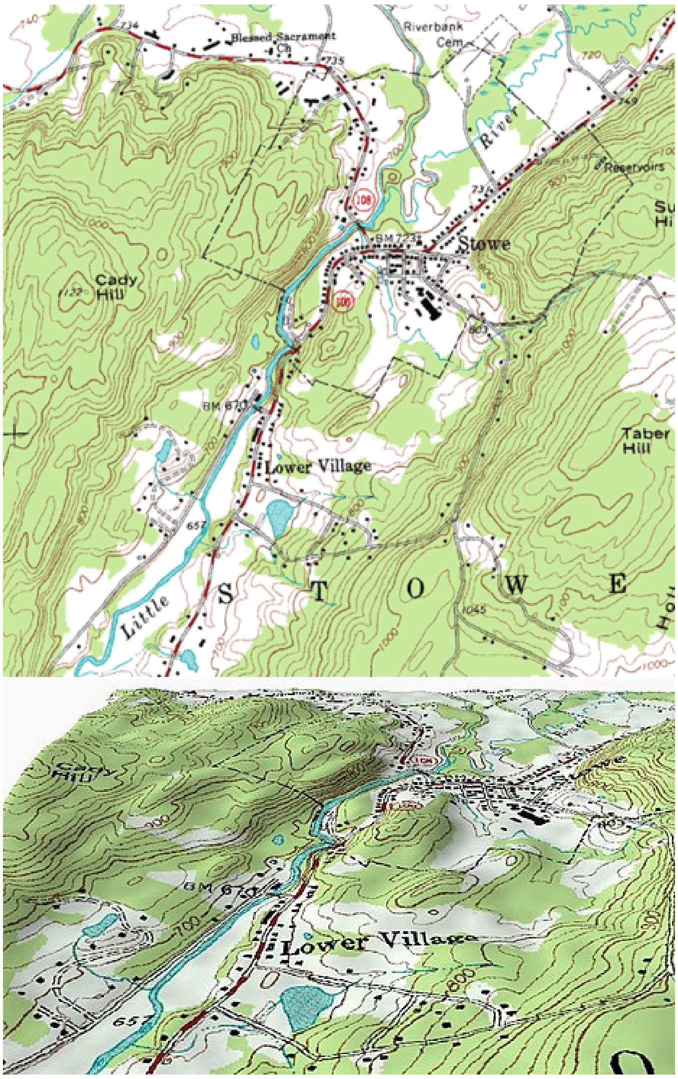


FIGURE 1. From Wikipedia, Topographic Map: a topographic map of the ski area of Stowe, Vermont and a shaded version of the map which helps to visualize the landscape.

For a function $F : \mathbb{R}^3 \to \mathbb{R}$ we can no longer draw the graph (we would need four dimensions!) but can still draw the analogue of the level curves. These are the level *surfaces*. An example is $F(x, y, z) = x^2 + y^2 + z^2$ for which the level surfaces of level c^2 are the spheres of radius c. See §5.13.

Remark 5.3. In weather maps we see curves which could indicate constant pressure or temperature. These actual functions are defined on space (since height above the ground is also a variable) so are the part close to earth of these level surfaces; if the Earth were perfectly flat, these would be the level curves of G defined by $G(x, y) = F(x, y, 0)$, in other words, the level surfaces for F meet sea level $z = 0$ in the level curves for G.

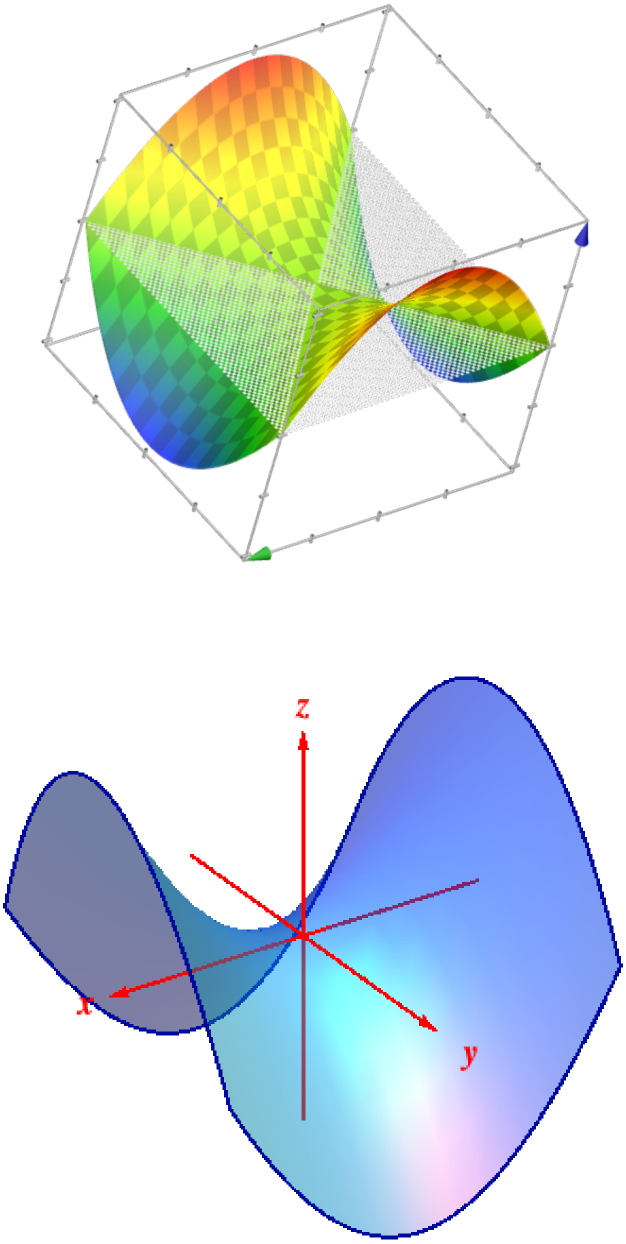


FIGURE 2. Graph of the function $F(x, y) = x^2 - y^2$ (a *parabolic hyperboloid*). From Google, search "$x^2 - y^2$" (rotatable image there) and from https://web.ma.utexas.edu/users/m408m/Display12-6-2.shtml. Horizontal slices (these project to the *level curves*) give a family of hyperbolas in the plane. Sliced vertically parallel to the x and y axes gives parabolas, sliced parallel to the lines $x = \pm y$ one way gives hyperbolas.

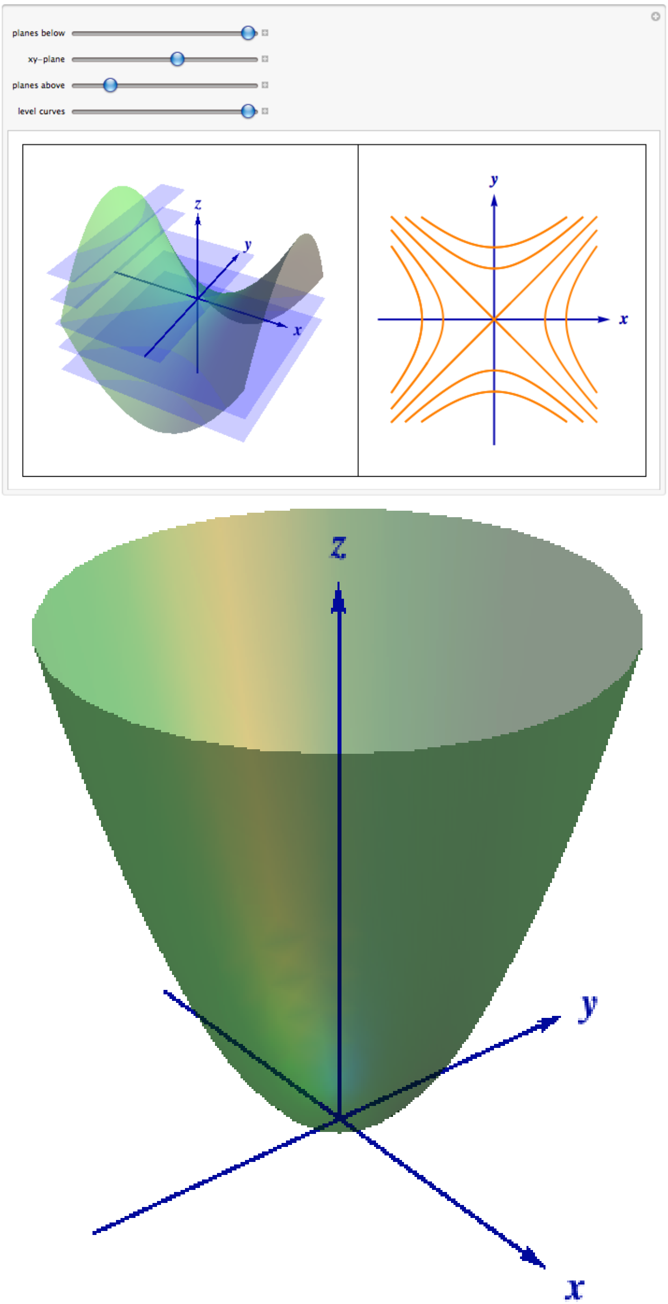


FIGURE 3. Graphs of parabolic hyperboloid with level curves (a family of hyperbolas), and of the paraboloid $F(x, y) = x^2 + y^2$.
From https://web.ma.utexas.edu/users/m408m/Display12-6-2.shtml.

5.5. **Partial derivatives; directional derivative.** Given a map $F : \mathbb{R}^2 \to \mathbb{R}$, choosing a point $\mathbf{p} \in \mathbb{R}^n$ then the *directional derivative* of F at $\mathbf{p}$ in the direction $\mathbf{u}$ is the following. Here we assume $||\mathbf{u}|| = 1$, i.e. $\mathbf{u}$ is a *unit vector*.

Above (Def. 5.4) we have defined the parametrized line $l(t) = \mathbf{p} + t\mathbf{u}$: the curve which starts at $\mathbf{p}$ and moves in the direction $\mathbf{u}$ at unit speed.

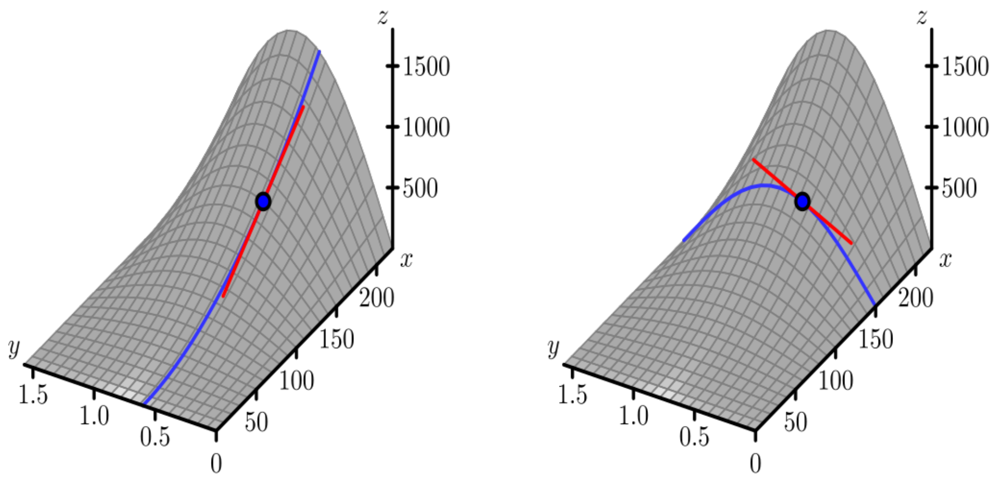


FIGURE 4. Partial derivatives; figure from https://activecalculus.org/vector/

Now $f(t) = F(l(t))$ is a function from $\mathbb{R}$ to $\mathbb{R}$. We define

$$D_{\mathbf{u}}(F)|_{\mathbf{p}} = f'(0).$$

This gives the amount of increase of the function F in the direction $\mathbf{u}$ at that point. A special case is for $\mathbf{u} = (1, 0)$. We define

$$\frac{\partial F}{\partial x}(\mathbf{p}) = D_{\mathbf{u}}(F)|_{\mathbf{p}}.$$

Similarly for $\mathbf{u} = (0, 1)$ we define

$$\frac{\partial F}{\partial y}(\mathbf{p}) = D_{\mathbf{u}}(F)|_{\mathbf{p}}.$$

See Fig. 4.

It is very easy to calculate the partial derivatives. For the partial with respect to x, we fix the variable y and find the derivative with respect to x alone.

For example, when $F(x, y) = x^2y^3$, then $\frac{\partial F}{\partial x} = 2xy^3$ while $\frac{\partial F}{\partial y} = 3x^2y^2$.

For $F : \mathbb{R}^n \to \mathbb{R}$ the definitions are similar.

5.6. **Properties of the gradient of a function.** Given a map $F : \mathbb{R}^m \to \mathbb{R}$ we define a vector at each point $\mathbf{p} \in \mathbb{R}^m$, $\nabla F = (\frac{\partial F}{\partial x_1}, \ldots, \frac{\partial F}{\partial x_m})$, called the *gradient* of F at $\mathbf{p}$.

The term may be related to, for example, a road going up a steep *grade*.

As we shall see in the next sections, the gradient has the following important properties:

(1) This defines a *vector field*, called the *gradient vector field* of F.

(2) The gradient vector field is everywhere orthogonal to the *level sets* of F. These are level *curves* for $F : \mathbb{R}^2 \to \mathbb{R}$ and level *surfaces* for $F : \mathbb{R}^3 \to \mathbb{R}$. We prove this via the Chain Rule, see §5.12. In general, the level sets are *submanifolds*, i.e. differentiable subsets, of $\mathbb{R}^n$, of dimension $(n - 1)$; this is a consequence of the *Implicit Function Theorem*, §5.19. (Here we have to assume that $\mathbf{n} \neq \mathbf{0}$).

(3) The gradient vector points in the direction of steepest increase of the function F at the point $\mathbf{p}$; its magnitude is the amount of increase in that direction.
(4) The directional derivative of F at $\mathbf{p}$, in the direction of the unit vector $\mathbf{u}$, is given simply by the inner product:

$$D_{\mathbf{u}}(F)|_{\mathbf{p}} = \nabla F|_{\mathbf{p}} \cdot \mathbf{u}.$$

(5) The gradient ∇F is the vector form of the *derivative* DF of the function F.
(6) The gradient can be used to easily write the equation of the *tangent line* to a level curve of $F : \mathbb{R}^2 \to \mathbb{R}$ at the point $\mathbf{p} = (x, y)$. For $F : \mathbb{R}^3 \to \mathbb{R}$, the gradient can be used to write the equation of the *tangent plane* to a level *surface* at a point $\mathbf{p} = (x_0, y_0, z_0)$. We explain this below in §5.12.

5.7. **Three types of curves and surfaces.** In the course we actually encounter three different (but related) types of curves and surfaces. First, recall:

Definition 5.8. For $f : [a, b] \to \mathbb{R}$ then its *graph* is

$$\text{graph}(f) = \{(x, f(x)) : x \in [a, b]\}$$

which is the subset of the plane we usually draw for this. Similarly, for $F : \mathbb{R}^2 \to \mathbb{R}$ then $\text{graph}(F) = \{(\mathbf{v}, F(\mathbf{v})) : \mathbf{v} = (x, y) \in \mathbb{R}^2\}$. Thus $\text{graph}(F) = \{(x, y, z) : z = F(x, y)\}$.

The different types of curves are:
(i) the graph of function $f : \mathbb{R} \to \mathbb{R}$;
(ii) a level curve of a function $F : \mathbb{R}^2 \to \mathbb{R}$;
(iii) a parametrized curve $\gamma : \mathbb{R} \to \mathbb{R}^m$.
Note that the first two are curves in the plane, while the second is a curve in $\mathbb{R}^m$ for any dimension m.

For surfaces we have, similarly:
(i) the graph of function $F : \mathbb{R}^2 \to \mathbb{R}$;
(ii) a level surface of a function $F : \mathbb{R}^3 \to \mathbb{R}$;
(iii) a parametrized surface $S : \mathbb{R}^2 \to \mathbb{R}^m$.
Now the first two are surfaces in $\mathbb{R}^3$, while the last is a surface inside of m-dimensional space so is much more general. Usually we will require that these functions be continuous and in the case of (iii) that the derivative DS exists, is continuous, and is surjective at every point. This guarantees the the image surface does not have creases or folds.

We also have *non*-parametrized curves and surfaces: the image of one of the functions in parts *(iii)*.

In both situations, curves and surfaces, these are all related, with (i) being a special case of (ii) and (ii) a special case of (iii). Regarding the passage from (ii) to (iii) see Exercise 5.7.

In all three cases it is important to first consider the linear (or affine) situation. That is because, first, it is the simplest case, and secondly, because these will describe the tangent line and the tangent plane, of a curve or surface, in all cases. These are exactly the affine lines or planes which bext approximate the curve or surface at a chosen point.

Here are the affine versions in all cases:
(i) An affine function $f : \mathbb{R} \to \mathbb{R}$ is $f(x) = y$ where $y = ax + b$.
An affine function $F : \mathbb{R}^2 \to \mathbb{R}$ has the form $F(x, y) = z$ where $z = ax + by + c$.
Note that the graph of f is a line in the plane, while the graph of F is a plane in $\mathbb{R}^3$.
(ii) The level curve of an affine function $F : \mathbb{R}^2 \to \mathbb{R}$: given $F(x, y) = ax + by$ then the level curve of level c is the line in the plane,

$$ax + by = c,$$

equivalently

$$Ax + By + C = 0,$$

for $A = a, B = b, C = -c$. This is now in the form of the *general equation of a line in the plane in* $\mathbb{R}^3$.
Given the affine function $F : \mathbb{R}^3 \to \mathbb{R}$ defined by given $F(x, y, z) = ax + by + cz$ then the level surface of of level d is the plane in $\mathbb{R}^3$:

$$ax + by + cz = d$$

or equivaelntly the *general equation of a plane*

$$Ax + By + Cz + D = 0,$$

for $A = a, B = b, C = c, D = -d$.
(iii) A parametrized affine curve $l : \mathbb{R} \to \mathbb{R}^m$: this is a parametrized line,

$$l(t) = \mathbf{p} + t\mathbf{v}.$$

A parametrized affine surface is $L : \mathbb{R}^2 \to \mathbb{R}^m$: this is a parametrized plane:

$$L(s, t) = \mathbf{p} + s\mathbf{v} + t\mathbf{w}.$$

Exercise 5.5.
$(\widetilde{i})$ Which lines in the plane, or planes in $\mathbb{R}^3$, can (or cannot) be written as the graph of an affine function as in (i)?
$(\widetilde{ii})$ For which values of a, b or a, b, c in (ii) do you get a line or plane?
$(\widetilde{iii})$ For which vectors $\mathbf{v}$ and $\mathbf{v}, \mathbf{w}$ in (iii) do you get a line or plane?
$(\widetilde{iv})$ Make sure you know how to go from one type of line (or plane) to the other, whenever possible (see the Linear Algebra lecture notes and exercises)!
$(\widetilde{v})$ Write each type of line or plane in *matrix form.*
Solution: We explain $(\widetilde{ii})$. We claim that the equation $Ax + By + C = 0$ gives a line in the plane $\mathbb{R}^2$ exactly when not both A, B are 0. Here we have to understand the meaning of "gives the equation of a line in the plane."
There are two important points:
(1) This means that we are *in the plane*, this is our *Universe of Discourse* (we are talking only about points in the plane $\mathbb{R}^2$, not about $\mathbb{R}$ or $\mathbb{R}^3$).
(2) "gives the equation of a line" means that the collection of all solutions to the equation forms a line.
That is,

$$\{(x, y) \in \mathbb{R}^2 : Ax + By + C = 0\}$$

is a geometrical line in $\mathbb{R}^2$.

It makes a huge difference what is our Universe of Discourse (i.e. what we are talking about). For example, the equation $x = 2$ in $\mathbb{R}$ is a point, in $\mathbb{R}^2$ it is a vertical line, $\{(x, y) : x = 2\}$, in $\mathbb{R}^3$ it is a vertical *plane.*

Now for $Ax + By + C = 0$ to be a line means that

$$\{(x, y) \in \mathbb{R}^2 : Ax + By + C = 0\}$$

is a line. Let us consider the case where $B \neq 0$. Then this equation is *equivalent*, i.e. it has the same solutions:

$$y = -A/By - C/B = ax + b$$

which we know is a line.

Next suppose A, B are both 0. Then we have

$$\{(x, y) \in \mathbb{R}^2 : 0x + 0y + C = 0\}$$

equivalently

$$\{(x, y) \in \mathbb{R}^2 : C = 0\}$$

and there are two cases:
(i) $C = 0$: this statement is *true*, hence is true for all (x, y), so the solution set is all of $\mathbb{R}^2$;
(i) $C \neq 0$: this statement is *false*, hence is false for all (x, y), so there are no solutions, and the solution set is the empty set.

This proves the Claim.

Planes are handled similarly.

Exercise 5.6.
(i) Given vector spaces V, W and a linear transformation $T : V \to W$, prove that:

Proposition 5.9. *The image of T, $Im(T)$ and the kernel (null space) of T,* $\ker(T)$ *are (vector) subspaces of W, V respectively.*

(ii) Interpret these statements for $T : \mathbb{R}^2 \to \mathbb{R}^m$ (for the image), respectively $T : \mathbb{R}^3 \to \mathbb{R}$ (for the kernel) as matrix equations, and describe the connection to the parametric and general equations of a plane.

Finding the tangent line or plane: as the graph of a function.

Next, for each of the three cases of curves and surfaces, we show how to find the tangent line, respectively the tangent plane. These are exactly the affine lines or planes which bext approximate the curve or surface at the point.

First, given a function $f : \mathbb{R} \to \mathbb{R}$ the formula for the tangent line to its graph is

$$l(x) = f(p) + f'(p)(x - p).$$

(Draw a picture!)

Note that $l(t)$ is itself a curve, and that it satisfies:
(i) it is an affine function, that is, linear plus a vector;
(ii)$l(p) = f(p)$;
(iii) $l'(p) = f'(p)$.

The formula for the tangent line to a (parametrized) curve $\gamma : \mathbb{R} \to \mathbb{R}^m$ is nearly identical:

$$l(t) = \gamma(p) + (t - p)\gamma'(p).$$

This now satisfies similar properties:
(i) it is an affine function;
(ii)$l(p) = \gamma(p)$;
(iii) $l'(p) = \gamma'(p)$.

For a level curve there are two ways to approach finding the tangent line. The first is to parametrize the level curve somehow and apply the previous case of a parametrized curve.

Exercise 5.7. For $F(x, y) = x^2 + y^2$, the curve of level 1 is the unit circle, the solutions of the equation (i.e. all pairs (x, y) which satisfy the equation)

$$x^2 + y^2 = 1.$$

Find parametrizations for this level curve, and use that to find the tangent line at a point.
Solution: We can parametrize this for example by the variable x. Then

$$y = \pm\sqrt{1 - x^2}$$

and we have two parametrized curves, with $t = x$, so $\gamma(t) = \pm\sqrt{1 - t^2}$. This works at all points except where $y'(x) = \infty$, that is, $x = \pm 1$. If we instead parametrize it by y then this works except where $x'(y) = 0$, that is, for $y = \pm 1$. We can also parametrize the entire curve at once, by the angle θ, with

$$\gamma(t) = (\cos t, \sin t)$$

and $t = \theta$.

When we parametrize by the variable x, we say the functions $f(x) = \sqrt{1 - x^2}$, $\widetilde{f}(x) = -\sqrt{1 - x^2}$, are *defined implicity* by the equation $x^2 + y^2 = 1$.

That is, they are *explicit* functions which are "implied" by the equation.

The *Implicit Function Theorem*, §5.19 describes when this can be done, basically when (partial) derivatives become infinite as above.

Given this, we can apply the formulas for the graph of a function, or for a curve in the plane, to find the tangent line of the level curve.

Finding the tangent line or plane: using the normal vector to find the tangent space. The second way to find the tangent line to a level curve is to find a normal vector to the curve. We explain this in §5.12.

Definition 5.9. Given $F : \mathbb{R}^n \to \mathbb{R}$ the *tangent space* $T_{\mathbf{p}}$ to the level set at the point $\mathbf{p} = (x_1, x_2, \ldots, x_n)$, for level $c = F(\mathbf{p})$, is an affine subset of $\mathbb{R}^n$, all vectors $\mathbf{v}$ such that $(\mathbf{v} - \mathbf{p})$ is othogonal to the gradient, $\mathbf{n} = \nabla F_{\mathbf{p}}$.

We consider first the case of $n = 2$. We write the equation of the tangent line, recalling that given a point $\mathbf{p}$ and a normal vector $\mathbf{n} = (A, B)$ then the line passing through $\mathbf{p}$ and perpendicular to $\mathbf{n}$ is the collection of all $\mathbf{x} = (x, y)$ such that

$$\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = 0.$$

Thus for $\mathbf{n} = (A, B)$ and $\mathbf{x} = (x, y)$ and $\mathbf{p} = (x_0, y_0)$ then

$$(A, B) \cdot (x - x_0, y - y_0) = 0$$

giving the general equation for the line,

$$Ax + By + C = 0$$

where $C = -\mathbf{n} \cdot \mathbf{p} = -(Ax_0 + By_0)$.

Since $\nabla F = \mathbf{n} = (\frac{\partial F}{\partial x}|_{\mathbf{p}}, \frac{\partial F}{\partial y}|_{\mathbf{p}})$ this gives the formula for the tangent line as

$$z_0 + \frac{\partial F}{\partial x}|_{\mathbf{p}}(x - x_0) + \frac{\partial F}{\partial y}|_{\mathbf{p}}(y - y_0) = 0 \tag{3}$$

We know the formula for the tangent line to the graph of a function $f : \mathbb{R} \to \mathbb{R}$ is $l(x) = f(p) + f'(p)(x - p)$. We can also use the normal vector method to find this formula in a second way. To do this we define $F : \mathbb{R}^2 \to \mathbb{R}$ by

$$F(x, y) = f(x) - y.$$

Then the level curve of level 0 gives $f(x) - y = 0$, so $y = f(x)$ which is the graph of f.

(Consider a simple example like $f(x) = x^2$ to understand what is going on!)

Note that at the point $\mathbf{p} = (p, f(p)$ is $\nabla F_{\mathbf{p}} = (f'(p), -1)$ so the formula (3) gives

$$\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = 0$$

where $\mathbf{p} = (p, f(p))$ so we have

$$(f'(p), -1) \cdot ((x - p, y - f(p)) = 0$$

so

$$f'(p)(x - p) - (y - f(p)) = 0$$

so

$$y = f(p) + f'(p)(x - p)$$

as claimed.

Exercise 5.8. See Exercise 5.9 below.

5.8. **The gradient vector field; the matrix form of the tangent vector and of the gradient.** The gradient ∇F of a function $F : \mathbb{R} \to \mathbb{R}^m$ gives an important example of a *vector field*. In general, a *vector field V on* $\mathbb{R}^m$ is a function V from $\mathbb{R}^m$ to $\mathbb{R}^m$.

As we mentioned above and shall prove in Proposition 5.15, *the level curves of a function F are orthogonal to the gradient vector field*, so the gradient can help us understand the level curves of F.

We draw the vector $\mathbf{w}_{\mathbf{v}} = V(\mathbf{v})$ *based at each point* $\mathbf{v}$. See Fig. 6.

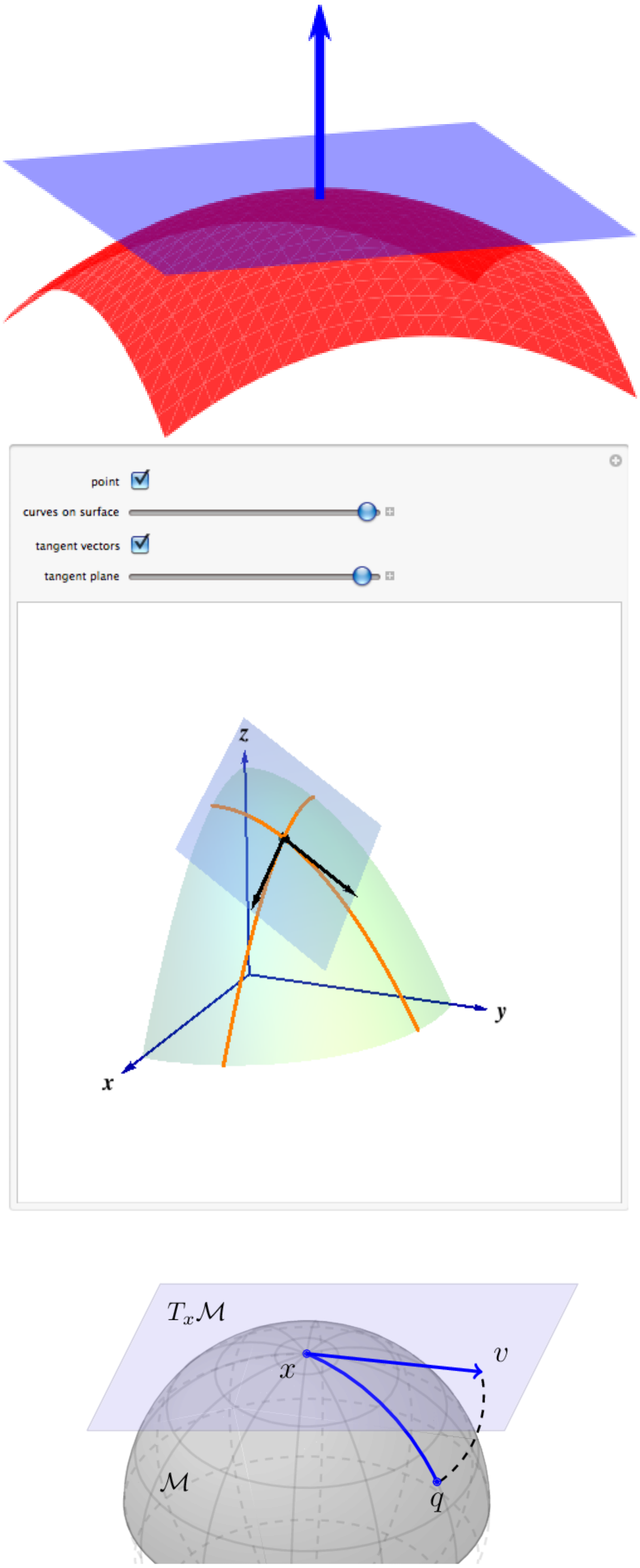


FIGURE 5. The normal vector to the surface is normal (orthogonal) to the tangent plane at that point. Tangent plane to the graph of a function defined on the plane, showing the meaning of the partial derivatives at the point. A tangent vetor to a curve in a surface is in the tangent plane. From https://web.ma.utexas.edu/users/m408m/Display14-4-2.shtml From https://www.researchgate.net/figure/Figure-S3-Geometric-illustration-of-tangent-vector-tangent-space-curve-and... From: Wikipedia, Normal (geometry)

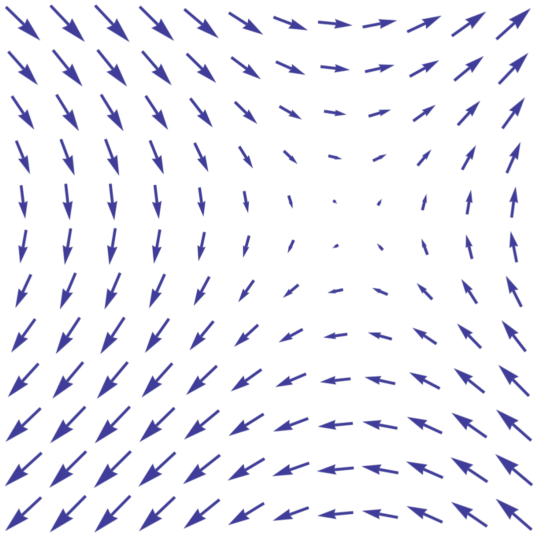


FIGURE 6. A vector field in the plane, from Wikipedia. Compare to the pictures of curves below!

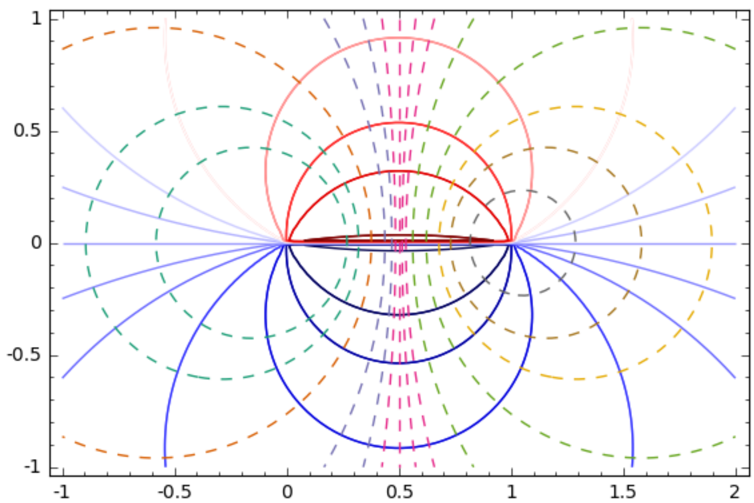


FIGURE 7. Equipotential curves for the electrostatic field of two opposite charges in the plane. Colors indicate different levels of the potential. This can also be interpreted as a gravitational field, where the potential function is height above sea level, and the positive charge is a mountain top while the negative charge is a valley. Orthogonal to the equipotentials are the lines of force; these are tangent to the gradient vector field of the potential function. One can imagine flowing along the lines of force from positive to negative charge, as in a fluid, although this is a force not a velocity field. (Because of this analogy with fluids, they are also called the *lines of flux* of the electrostatic field.

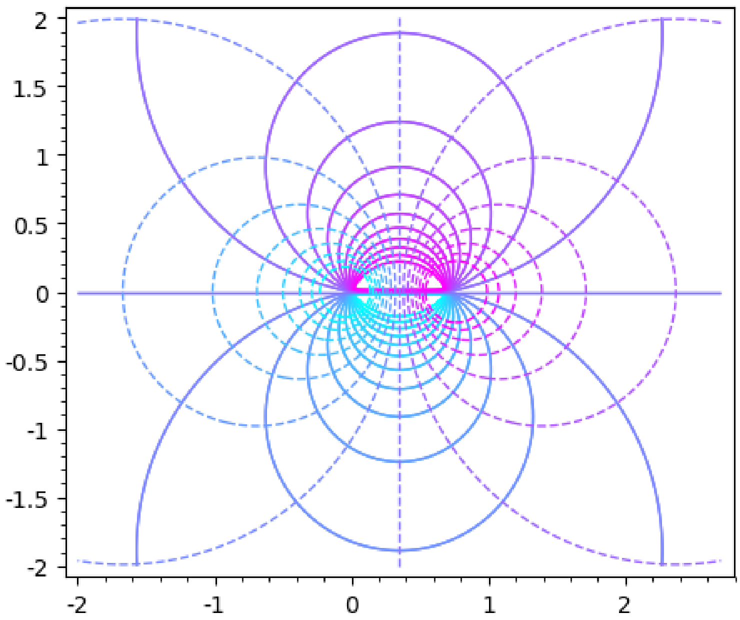


FIGURE 8. Equipotential curves and lines of force for the electrostatic field of two opposite charges in the plane, now closer together.

The tangent vector gives the first definition of derivative of a curve $\gamma : \mathbb{R} \to \mathbb{R}^m$, the vector form of the derivative; the second definition, the *matrix form* of the derivative, is the $(n \times 1)$ matrix, i.e. the column vector with those same entries:

$$D\gamma = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t)) \end{bmatrix}.$$

Thus $D\gamma : \mathbb{R} \to \mathcal{M}_{n\times 1}$.

For a function $F : \mathbb{R}^n \to \mathbb{R}$, the vector form of its derivative is the gradient ∇F. This has a matrix form, the row vector i.e. $(n \times 1)$ matrix with the same entries:

$$DF|_{\mathbf{x}} = \begin{bmatrix} \frac{\partial F}{\partial x_1} & \dots & \frac{\partial F}{\partial x_n} \end{bmatrix}.$$

Given $\gamma : \mathbb{R} \to \mathbb{R}^m$ and $F : \mathbb{R}^m \to \mathbb{R}$ then the composition is $F \circ \gamma : \mathbb{R} \to \mathbb{R}$, so we can take its derivative $(F \circ \gamma)'(t)$. The *Chain Rule* says we can compute this in a second way. In vector notation it states:

$$(F \circ \gamma)'(t) = \nabla F_{\gamma(t)} \cdot \gamma'(t).$$

This is even simpler to remember in matrix notation, as we have the product of a row vector and a column vector. For example, with $\gamma : \mathbb{R} \to \mathbb{R}^3$ and $F : \mathbb{R}^3 \to \mathbb{R}$, we have

$$D(F \circ \gamma(t)) = \begin{bmatrix} F_x \, F_y \, F_z \end{bmatrix} |_{\gamma(t)} \begin{bmatrix} x_1'(t) \\ x_2'(t) \\ x_3'(t)) \end{bmatrix}.$$

Exercise 5.9. $F(x,y) = x^2y^3$; $\gamma(t) = (e^t, e^{t^2})$.
(1)Find $F \circ \gamma'(0)$.

First method (directly): $f(t) = F \circ \gamma(t) = e^{2t}e^{3t^2}$. $f'(t) = 2e^{2t}e^{3t^2} + 6t \cdot e^{2t}e^{3t^2}$. $f'(0) = 2$.

Second method (Chain Rule): $\nabla F = (2xy^3, x^2 3y^2)$. $\gamma'(t) = (e^t, t^2 e^{t^2})$.

$\gamma(0) = (1,1)$. $\nabla F(1,1) = (2,3)$. $\gamma'(0) = (1,0)$.

So $F \circ \gamma'(0) = (2,3) \cdot (1,0) = 2$.

Which is easier depends on the problem!
(2)Find ∇F at $\mathbf{p} = (2,5)$.
(3) Find the equation of the tangent plane to the graph of F at that point.

5.9. **General definition of derivative of a map.** Both the matrix version of tangent vector and of gradient are special cases of the general notion of derivative of a map from $F : \mathbb{R}^n \to \mathbb{R}^m$. We state this more generally for normed vector spaces.

Definition 5.10. Let V, W be Banach spaces (a vector space, possibly infinite-dimensional, on which we have a complete norm defined; *complete* here means that there are "no holes" as Cauchy sequences converge; this only can be an issue in infinite dimensions. The reader can think of $\mathbb{R}^n$ with the standard inner product and norm to get the basic idea). We say 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 write $DF_{\mathbf{p}} = L$ for this transformation, called the *derivative* of F at $\mathbf{p}$.

The idea here is that the derivative DF should give the "best linear approximation" at each point.

What this actually means is the linear part of the best *first-order approximation.* The best 0^{th}-order approximation at $\mathbf{x} \in \mathbb{R}^n$ is the constant map with the value at that point, thus the map $\mathbf{x} \mapsto \mathbf{p} = F(\mathbf{x})$. If $L = DF|_{\mathbf{x}}$, then the best first-order approximation will be the linear map shifted by this value, thus the affine map $(\mathbf{x}+\mathbf{v}) \mapsto \mathbf{p} + L\mathbf{v}$. (Literally what the term "first-order approximation" is "making use of the first derivative"; 0^{th}-order means approximating the function near $\mathbf{x}$ by its *value*, which is the stupidest approximation; note that this only helps at all for *continuous* functions. Similarly when the function has k continuous derivatives, we can use the derivatives at $\mathbf{x}$ from 0 to k to make a k^{th}-order approximation near that point. This gets better and better as k increases, as guaranteed by Taylor series at the point. See SS5.10 and §5.23.

Let us relate the above formula to the usual definition for a function $f : \mathbb{R} \to \mathbb{R}$, that is,

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = c$$

This definition still works for curves, giving us the tangent vector. However for V of dimension larger than 1 this makes no sense, as *we cannot take the ratio of two vectors.*

Remark 5.4. 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 $L\mathbf{v} = \mathbf{w}$; it is only well-defined if V has dimension 1. What the definition of derivative requires is that L works for *all* directions $\mathbf{h}$, and this does make L well-defined.

Remark 5.5. Let us see what happens to the general definition for $f : \mathbb{R} \to \mathbb{R}$. Then

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = c$$

iff 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

$$\frac{|f(x+h) - f(x) - ch|}{|h|} < \varepsilon$$

And this is now a special case of the general formula.

We introduce the notation $\mathcal{L}(V, W)$ for the collection of all linear transformations from V to W. If we choose a basis for $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$, then $L : \mathbb{R}^n \to \mathbb{R}^m$ is represented by an $(m \times n)$ matrix. Then $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ can be identified with the matrices $\mathcal{M}_{mn} \sim \mathbb{R}^{mn}$, so $DF : \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \sim \mathbb{R}^{mn}$.

When considering $F : \mathbb{R}^n \to \mathbb{R}^m$ then both $\mathbf{x} \in \mathbb{R}^n$ and $F(\mathbf{x}) \in \mathbb{R}^m$ can be written in components, with $\mathbf{x} = (x_1, \dots, x_n)$ and $F(\mathbf{x}) = (F_1(\mathbf{x}), \dots, F_m(\mathbf{x}))$. We write the components of F as a column vector, so

$$F(\mathbf{x}) = \begin{bmatrix} F_1(\mathbf{x}) \\ \vdots \\ F_m(\mathbf{x}) \end{bmatrix}.$$

Each component F_k is a function from $\mathbb{R}^n$ to $\mathbb{R}$ so has a gradient, $\nabla F_k = (\frac{\partial F_k}{\partial x_1}, \dots, \frac{\partial F_k}{\partial x_n})$. Now DF_k is by definition is the corresponding row vector so

$$DF_k = \begin{bmatrix} \frac{\partial F_k}{\partial x_1} & \dots & \frac{\partial F_k}{\partial x_n} \end{bmatrix}.$$

We define the *matrix of partials* of $F : \mathbb{R}^n \to \mathbb{R}^m$ to be the $(m \times n)$ matrix with the k^{th} row this gradient vector. Let us write $[\nabla F_k]$ for the row vector DF_k.

Then the ij^{th}-matrix entry is the partial derivative

$$(DF)_{ij} = \frac{\partial F_i}{\partial x_j}$$

and so we have the $(m \times n)$ matrix

$$DF|_{\mathbf{x}} = \begin{bmatrix} [\nabla F_1] \\ \vdots \\ [\nabla F_m] \end{bmatrix} = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \dots & \frac{\partial F_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial F_m}{\partial x_1} & \dots & \frac{\partial F_m}{\partial x_n} \end{bmatrix}.$$

The most basic cases are $f : \mathbb{R}^1 \to \mathbb{R}^m$ and $\mathbb{R}^n \to \mathbb{R}^1$. The first is a curve, discussed above, and usually written $\gamma : \mathbb{R} \to \mathbb{R}^m$. The general formula then gives the matrix form of the tangent vector; since γ is a column vector, with

$$\gamma(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_m(t)) \end{bmatrix}$$

then $D\gamma$ is the $(m \times 1)$ matrix with the same entries as the tangent vector $\gamma'(t) = (x_1',, \dots, x_m')(t)$, so

$$D\gamma|_t = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_m'(t)) \end{bmatrix}.$$

The second type of map $F : \mathbb{R}^n \to \mathbb{R}$ we call simply a *function*. The general formula above then gives the $(1 \times n)$ matrix:

$$DF|_{\mathbf{x}} = \begin{bmatrix} \frac{\partial F}{\partial x_1} & \dots & \frac{\partial F}{\partial x_n} \end{bmatrix}_{\mathbf{x}}.$$

This row vector is the *matrix form* of the gradient ∇F, since as explained above, $\nabla F = (\frac{\partial F}{\partial x_1}, \dots, \frac{\partial F}{\partial x_n})$.

As we shall see in Proposition 5.15, *the level curves of a function F are orthogonal to the gradient vector field.*

An example of level curves is seen in Fig. 9.

One can think of the matrix of partials for a function $F : \mathbb{R}^n \to \mathbb{R}^m$ as consisting of lined-up column vectors (tangent vectors) or row vectors (gradients) respectively. We have explained this regarding the rows. To understand this for the columns, writing a vector in the domain as $\mathbf{x} = (x_1, \dots, x_n)$ then fixing say $x_2, \dots, x_n$ and setting $t = x_1$ we have a curve $\gamma(t) = F(t, x_2, \dots, x_n)$; note that the first column of the derivative matrix DF is the derivative of this curve, the column tangent vector $D\gamma$.

We have described how the derivative at a point defines a matrix of partial derivatives. The converse is:

Lemma 5.10. *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.*

For proof we refer to e.g. §11.2 of [Gui02].

Another basic theorem regarding derivatives is the relation to the matrix of partials:

Theorem 5.11. *If for $F : \mathbb{R}^n \to \mathbb{R}^m$, all the partial derivatives $\partial F_i / \partial x_j$ exists and is continuous at* $\mathbf{p}$*, then F is differentiable at* $\mathbf{p}$*, and its derivative is the linear map given by the matrix of partials.*

For a proof see Theorem 6.4 of Marsden's book [Mar74]. The of derivatives is very clearly carried out on pp. 158-185 of Marsden.

5.10. **Best affine approximation: tangent line and plane.** Given a map $F : \mathbb{R}^n \to \mathbb{R}^m$, the terminology "$k^{\text{th}}$-order approximation" to F at a point $\mathbf{x} \in \mathbb{R}^n$ comes from the Taylor polynomials and Taylor series. For a function $f : \mathbb{R}^n \to \mathbb{R}^m$, the best k^{th}-order approximation at $\mathbf{x} \in \mathbb{R}^n$ is the polynomial of degree k which best fits the map near that point. This is the polynomial (in k variables!) which has all the same partial derivatives at that point, up to order k. See §5.23.

Thus the best 0^{th}-order approximationof F at $\mathbf{p} \in \mathbb{R}^n$ is the constant map with the value at that point: the map $\mathbf{x} \mapsto F(\mathbf{p})$. To get the best first-order approximation we add on the linear map given by the derivative matrix $DF|_{\mathbf{p}}$.

This is the affine map

$$\mathbf{x} \mapsto F(\mathbf{p}) + DF|_{\mathbf{p}}(\mathbf{x} - \mathbf{p}),$$

which we mentioned above for the case of $\mathbb{R}^2$, where this gives the equation of the tangent plane to the graph of F.

Example 10. **(Derivative of a linear or affine map)** What is the best linear approximation to a linear map? Answer: it should be the linear map itself!

Let us understand this precisely.

Consider the matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Acting on column vectors, this defines the linear transformation

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$$

or written as vectors,

$$T(x, y) = (ax + by, cx + dy).$$

We want to compute the matrix of partials $DT_{\mathbf{p}}$ at a point $\mathbf{p} = (x_0, y_0)$. Now the components of T are $T = (T_1, T_2)$ where

$$T_1(x, y) = ax + by$$
$$T_2(x, y) = cx + dy.$$

Then

$$\nabla T_1 = (a, b)$$
$$\nabla T_2 = (c, d)$$

for each point $\mathbf{p}$. Thus at each point $\mathbf{p}$,

$$DT_{\mathbf{p}} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

This shows:

Proposition 5.12. *For an affine map $F : \mathbb{R}^n \to \mathbb{R}^m$, defined by $F(\mathbf{v}) = \mathbf{v}_0 + T(\mathbf{v})$ where $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear, the derivative is*

$$DF_{\mathbf{p}} = T.$$

That is to say, the derivative is constant, *is constantly equal to the linear part of the map.*

Remark 5.6. To really understand this, consider the case of $\mathbb{R}_\theta$, the rotation counterclockwise of the plane by angle θ.

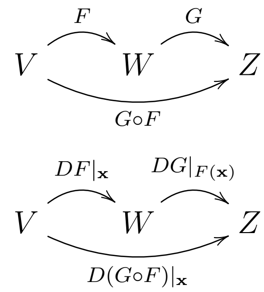
5.11. **The general Chain Rule.** The main theorem involving derivatives is the:

Proposition 5.13. (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 have seen the Chain Rule above; in matrix notation it is:

$$D(F \circ \gamma(t)) = \begin{bmatrix} F_x \, F_y \, F_z \end{bmatrix} |_{\gamma(t)} \begin{bmatrix} x_1'(t) \\ x_2'(t) \\ x_3'(t)) \end{bmatrix}$$

The product gives a (1×1) matrix, whose entry is a number.

In vector notation the Chain Rule is:

$$(F \circ \gamma)'(t) = \nabla F_{\gamma(t)} \cdot \gamma'(t).$$

This number is the same as the entry of the (1×1) matrix above.

Now we can give a second proof of Proposition 5.4 above, which we repeat here:

Proposition 5.14. *Let γ be a differentiable curve in $\mathbb{R}^n$ such that $||\gamma|| = c$ for some constant c. Then $\gamma \perp \gamma'$.*

Proof. (Second Proof, using gradient) We define a function $F : \mathbb{R}^n \to \mathbb{R}$ by $F(\mathbf{x}) = ||\mathbf{x}||^2 = \mathbf{x} \cdot \mathbf{x} = \sum_{i=1}^n x_i^2$. Then since $||\gamma|| = c$ is constant, $c^2 = ||F \circ \gamma||$ whence by the Chain Rule,

$$0 = (F \circ \gamma)'(t) = (\nabla F(\gamma(t)) \cdot \gamma'(t)$$

but $F(\mathbf{x}) = F(x_1, \dots, x_n) = x_1^2 + \dots + x_n^2$ whence $\nabla F(\mathbf{x}) = 2(x_1, \dots, x_n) = 2\mathbf{x}$. Thus $0 = 2\gamma(t) \cdot \gamma'(t)$, as claimed.

□

Directional derivative and the gradient.

The gradient gives us a simple way of calculating the directional derivative. Given $F : \mathbb{R}^n \to \mathbb{R}$, with gradient vector field ∇F, and given a unit vector $\mathbf{u}$, then the directional derivative of F in direction $\mathbf{u}$ is given simply by the inner product:

$$D_{\mathbf{u}}(F)|_{\mathbf{p}} = (\nabla F(\mathbf{p})) \cdot \mathbf{u}.$$

Exercise 5.10. Check this on the standard basis vectors and compare to the partial derivatives! What is the direction of steepest increase of F at a point $\mathbf{p}$? Of steepest decline? What is the rate of increase of F if in a direction tangent to a level curve?

5.12. **Level curves and parametrized curves.** There are two very distinct types of curves we encounter here: the curves of this section, which are *parametrized curves* (with *parameter* $t =$ time), and the *level curves* of a function. Next we describe a link between the two:

Proposition 5.15. *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 all t. Hence $G(\gamma(t))' = 0$ for all t. Then by the chain rule, this equals $0 = D(G \circ \gamma)(t) = DG|_{\gamma(t)} D\gamma|_t$. 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 gives $0 = (G \circ \gamma)'(t) = (\nabla G)(\gamma(t)) \cdot \gamma'(t)$, so $\nabla G|_{\gamma(t)} \cdot \gamma'(t) = 0$, telling us that the gradient is perpendicular to the tangent vector of the curve, as claimed. □

Example 11. (Dual hyperbolas) See Fig. 9, depicting level curves of the functions $F(x, y) = (x^2 - y^2)$ and $G(x, y) = 2xy$.

Exercise 5.11. Plot the level curves of F for levels $0, 1, -1$ and for G of levels $0, 2, -2$. Compute the gradient vector fields and find their matrices (they are linear!) Compare to the earlier examples of linear vector fields.

These functions are related algebraically by a change of variables, $u = \frac{1}{\sqrt{2}}(x - y)$, $v = \frac{1}{\sqrt{2}}(x + y)$ and geometrically by a rotation $R_{\pi/4}$. To verify this we define the function $H : \mathbb{R}^2 \to \mathbb{R}^2$ by $H(x, y) = (u, v) = \frac{\sqrt{2}}{2}(x - y, x + y)$ then

$$G \circ H(x, y) = 2 \cdot \frac{1}{2}(x - y)(x + y) = x^2 - y^2 = F(x, y)$$

so $F = G \circ H$.

Now H is a linear transformation of $\mathbb{R}^2$ given by

$$\frac{\sqrt{2}}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix}$$

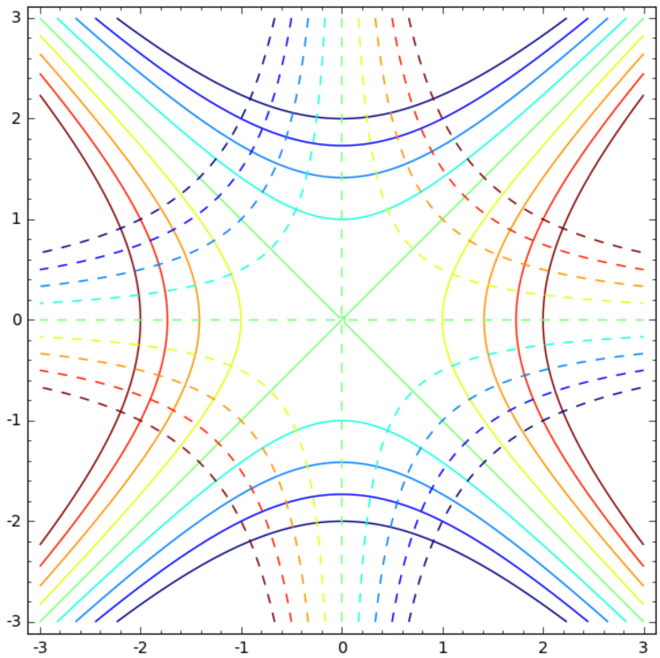


FIGURE 9. Dual families of hyperbolas: Level curves for the functions $F(x, y) = (x^2-y^2)$ and $G(x, y) = 2xy$. Note that in this special example the level curves of F are orthogonal to the level curves of G. In fact, the gradient vector field of F is orthogonal to the level curves of F, and is tangent to the level curves of G, and vice-versa!

and so by the matrix

$$\frac{\sqrt{2}}{2}\begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix} = \begin{bmatrix}\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2}\\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2}\end{bmatrix}$$

which is indeed rotation counterclockwise by $\pi/4$.

We next check Proposition 5.15 for this example.

The gradient of F is $\nabla F = (\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}) = (2x, -2y)$ and of G is $\nabla G = (\frac{\partial G}{\partial x}, \frac{\partial G}{\partial y}) = (y, x)$. Note that $(2x, -2y) \cdot (y, x) = 0$ so these vector fields are orthogonal. Furthermore we can find tangent vectors to the level curves as follows. Let us parametrize the level curve $F(x, y) = x^2 - y^2 = c$ by the variable x. Then $y = y(x)$, so the curve is $\gamma(x) = (x, y(x)$ with tangent vector $(1, y'(x))$taking the derivative of the equation with respect to x gives $2x - 2yy' = 0$, so $y' = x/y$. Thus $\gamma'(x) = (1, x/y)$. This is proportional to the vector $(y, x) = \nabla G$ which as we have already noted is orthogonal to the gradient of F at that point.

5.13. **Level surfaces, the gradient and the tangent plane.** In Proposition 5.15 of §5.12 we showed that the gradient vector field of $F : \mathbb{R}^2 \to \mathbb{R}$ is orthogonal to the level curves of F. In fact something similar is true for any dimension. For the case of $\mathbb{R}^3$ we get a new formula for the tangent plane, as we now explain.

Proposition 5.16. *Let* $G : \mathbb{R}^3 \to \mathbb{R}$ *be differentiable and suppose* $\gamma : [a, b] \to \mathbb{R}^3$ *is a curve such that the image of* γ *remains inside the level surface of level* c, $\{(x, y, z) : G(x, y, z) = c\}$. *That is, for all* t, $G(\gamma(t)) = c$. *Then* $\gamma'(t)$ *is perpendicular to the gradient of* G.

More generally this is true for higher dimensions, $G : \mathbb{R}^n \to \mathbb{R}$.

Proof. We have that $G(\gamma(t)) = c$ for all t. Hence $G(\gamma(t))' = 0$ for all t. Now by the chain rule, $0 = D(G \circ \gamma)(t) = DG(\gamma(t)D\gamma(t)$. DG is now a $(1 \times n)$ matrix and $D\gamma$ a $(n \times 1)$ column vector; in vector notation, these are the gradient and tangent vector, so this gives $0 = \frac{\mathrm{d}}{\mathrm{d}t}c = (G \circ \gamma)'(t) = (\nabla G)(\gamma(t)) \cdot \gamma'(t) = 0$. □

Exercise 5.12. First we have a review problem from Linear Algebra: Recall that the general equation for a plane in $\mathbb{R}^3$ is:

$$Ax + By + Cz + D = 0$$

where not all three of A, B, C are 0. Given a point $\mathbf{p} = (x_0, y_0, z_0)$ and a vector $\mathbf{n} = (A, B, C)$ then find the general equation of the plane through $\mathbf{p}$ and perpendicular to $\mathbf{n}$.

Solution: We know that the plane is the collection of all $\mathbf{x} = (x, y, z)$ such that

$$\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = 0,$$

so for $\mathbf{n} = (A, B, C)$ and $\mathbf{x} = (x, y, z)$ and $\mathbf{p} = (x_0, y_0, z_0)$ then

$$(A, B, C) \cdot (x - x_0, y - y_0, z - z_0) = 0$$

giving the general equation for the plane,

$$Ax + By + Cz + D = 0$$

where $D = -\mathbf{n} \cdot \mathbf{p} = -(Ax_0 + By_0 + Cz_0)$.

See also Exercise 5.5.

Exercise 5.13. Given the function $F(x, y, z) = x^2 + y^2 + z^2$, find the tangent plane to this sphere at the point $(1, 2, 3)$.

Solution: Note that $F(1, 2, 3) = 14$. Therefore this point is on the level surface of F of level 14. (This is the sphere about the origin of radius $\sqrt{14}$).

Now the gradient of F is $\nabla F(x, y, z) = (2x, 2y, 2z)$. We know the gradient is orthogonal to the sphere hence to the tangent plane. This *normal vector* (to both) is $\mathbf{n} = \nabla F(1, 2, 3) = (2, 4, 6)$. We are in the situation of the previous exercise: the equation of the plane is

$$Ax + By + Cz + D = 0$$

where the normal vector is $\mathbf{n} = (A, B, C) = (2, 4, 6)$ and the plane passes through the point $\mathbf{p} = (1, 2, 3)$.

The equation of the plane is therefore

$$\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = 0$$

or

$$\mathbf{n} \cdot ((x, y, z) - \mathbf{p}) = 0$$

so

$$(A, B, C) \cdot (x - 1, y - 2, z - 3) = 0$$

giving

$$2x + 4y + 6z + D = 0$$

where

$$D = -\mathbf{n} \cdot \mathbf{p} = -(2, 4, 6) \cdot (1, 2, 3) = -(2 + 8 + 18) = -28,$$

so we have the plane with general equation

$$2x + 4y + 6z + -28 = 0.$$

Exercise 5.14. We solve an exercise requested by a student.

Guidorizzi Vol 2 # 2 of §11.3, p. 204. [Gui02],

Find the equation of a plane which passes through the ponts $(1, 1, 2)$ and $(-1, 1, 1)$and which is tangent to the graph of the function $f(x, y) = xy$.

Solution. We use normal vectors, as follows. The graph of f is

$$\{(x, y, z) : z = f(x, y)\}$$

which equals

$$\{(x, y, z) : z = xy\}$$

equivalently written

$$\{(x, y, z) : xy - z = 0\}.$$

This is the level surface of $F : \mathbb{R}^3 \to \mathbb{R}$ defined by $F(x, y, z) = xy - z$. This has gradient vector $\nabla F = (y, x, -1)$. Let $\mathbf{p} = (x_0, y_0, z_0)$ denote the point where the plane meets the graph. Then at the point $\mathbf{p}$ we have $\nabla F_{\mathbf{p}} = (y_0, x_0, -1)$. We know that the gradient is orthogonal to the level surfaces, in other words it is orthogonal to the tangent plane to the surface at that point. So $\mathbf{n} = \nabla F_{\mathbf{p}}$ is a normal vector to the tangent plane of the level surface at $\mathbf{p}$. This gives us the equation for the tangent plane

$$\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = 0$$

so

$$(y_0, x_0, -1) \cdot ((x, y, z) - (x_0, y_0, z_0)) = 0$$

$$(y_0, x_0, -1) \cdot (x - x_0, y - y_0, z - z_0) = 0$$

so

$$y_0x - x_0y - z + z_0 = 0$$

Now $z_0 = x_0y_0$ since $\mathbf{p}$ is also on the graph of the function. This gives

$$y_0x - x_0y - z + x_0y_0 = 0$$

We need to find x_0, y_0. The two points are on this plane so satisfy the equation.

Substituting $(x, y, z) = (1, 1, 2)$ and $(-1, 1, 1)$ gives us the equations

$$y_0 - x_0 - 2 + x_0y_0 = 0$$

$$-y_0 - x_0 - 1 + x_0y_0 = 0$$

Subtracting,

$$2y_0 - 1 = 0$$
$$y_0 = 1/2$$

We now have from the first equation,

$$y_0 - x_0 - 2 + x_0 y_0 = 0$$

so

$$1/2 - x_0 - 2 + x_0 1/2 = 0$$

multiplying by 2,

$$1 - 2x_0 - 4 + x_0 = 0$$
$$x_0 = 3$$

Thus $z_0 = 3/2$ giving the equation of the plane:

$$\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = 0$$

with $\mathbf{n} = (y_0, x_0, -1) = (1/2, 3, -1)$ and $\mathbf{p} = (x_0, y_0, z_0) = (3, 1/2, 3/2)$. Finally in the form

$$Ax + By + Cz + D = 0$$

we have

$$1/2x + 3y - z - 3/2 = 0$$

or equivalently

$$x + 6y - 2z - 3 = 0.$$

To check our numbers we can verify that the three points are indeed on this plane.

Remark 5.7. In these notes we have emphasized the role of three distinct ways of presenting, or of viewing, the same object: for example a curve may be the graph of a function, a level curve, or a parametrized curve. We wish to indicate how this fits into a larger context, in other parts of mathematics.

First, here is a solution to part of Exercise 5.6: to write the image and kernel in matrix form.

Consider

$$\begin{bmatrix} 2 & 3 \\ 1 & 2 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} s \\ t \end{bmatrix} = s \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} + t \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix} \tag{4}$$

Thus if we write the columns of a (3×2) matrix as $\mathbf{v} = (v_1, v_2, v_3), \mathbf{w} = (w_1, w_2, w_3)$ we have more generally

$$\begin{bmatrix} v_1 & w_1 \\ v_2 & w_2 \\ v_3 & w_3 \end{bmatrix} \begin{bmatrix} s \\ t \end{bmatrix} = s \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} + t \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \tag{5}$$

defining the map $T : \mathbb{R}^2 \to \mathbb{R}^3$ where $T(s,t) = s\mathbf{v} + t\mathbf{w}$. This is a parametrized plane, which in this case passes through $\mathbf{0}$.

Given a parametrized plane in $\mathbb{R}^3$, we should be able to find the general equation. To do this we bring in the *vector product*, which we next explain. But first, a few words about the *determinant*!

5.14. **Elementary row and column operations.** A good reference on matrix operations in Linear Algebra is Strang's text [Str12].

For the next section we recall:

Definition 5.11. Given an $(n \times m)$ matrix M, a *basic elementary row operation* has two types:
(1) We exchange two rows;
(2) We replace a chosen row by itself added to a multiple of a different row.

In many texts, a third operation is permitted:
(3) We multiply a row by a constant.

But we don't need this and it will be more convenient for us to not include this one.

An *elementary row operation* is the result of applying the basic operations 1 and 2 finitely many times.

Elementary column operations are defined similarly.

A *basic elementary matrix* is a $(d \times d)$ matrix formed by carrying out an elementary row operation on the identity matrix I.

Lemma 5.17.
(i) A basic elementary matrix can also be formed by carrying out an elementary column *operation on I.*
(ii) Given an $(n \times m)$ matrix M, let $\widetilde{M}$ be the matrix which results after carrying out an elementary row operation. Then $\widetilde{M} = EM$ where E is the $(n \times n)$ elementary matrix formed by carrying out the same elementary row operation on I.

The same is true for column operations except now E is $(m \times m)$ and $\widetilde{M} = ME$.

Proof. First we consider the basic elementary matrices. (i) If we form E by exchanging row i and row k, then E has a 1 in place i of row k and place k of row i. Thus $E_{ki} = 1$ and $E_{ik} = 1$. If we exchange column i and column k then E has a 1 in place i of column k and place k of column i. Thus $E_{ik} = 1$ and $E_{ki} = 1$, the same thing.

If we form E by replacing row i with a times row k, then E is the identity matrix except for $E_{ik} = a$. If we form E by replacing column k with a times row i, then E is the identity except for $E_{ik} = a$, the same thing!

(ii) For $\mathbf{e}_k^t$ the standard basis row vector, so it is $(1 \times n)$ with $(\mathbf{e}_k)_k = 1$, then $\mathbf{e}_k^t M$ gives the $(1 \times m)$ row vector which is the k^{th} row of M. It follows from this that if an $(n \times n)$ matrix N has $\mathbf{e}_k^t$ as its j^{th} row, then row j of NM is the k^{th} row of M. This implies the claim, but it is much better to see this by trying out some examples; see (8), (6), (7) below.

The case of operation (2) is similar.

General elementary matrices: by what we have just shown, a general elementary matrix E is a product $E = E_l \dots E_2 E_1$, proving (i). Part (ii) follows.

For column operations, we take the transpose, using the fact that $(EM)^t = M^t E^t$. □

Pictorially,

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \longrightarrow & \mathbf{u} & \longrightarrow \\ \longrightarrow & \mathbf{v} & \longrightarrow \\ \longrightarrow & \mathbf{w} & \longrightarrow \end{bmatrix} = \begin{bmatrix} \longrightarrow & \mathbf{v} & \longrightarrow \\ \longrightarrow & \mathbf{u} & \longrightarrow \\ \longrightarrow & \mathbf{w} & \longrightarrow \end{bmatrix} \tag{6}$$

and

$$\begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \longrightarrow & \mathbf{u} & \longrightarrow \\ \longrightarrow & \mathbf{v} & \longrightarrow \\ \longrightarrow & \mathbf{w} & \longrightarrow \end{bmatrix} = \begin{bmatrix} \longrightarrow & \mathbf{u} & \longrightarrow \\ \longrightarrow & a\mathbf{u}+\mathbf{v} & \longrightarrow \\ \longrightarrow & \mathbf{w} & \longrightarrow \end{bmatrix} \tag{7}$$

5.15. **Geometrical meaning of elementary matrices: reflection and sliding.** For this, we first consider elementary row operations. Let us consider the (2×2) basic elementary matrix $E = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. In its action on column vectors it interchanges $\mathbf{i}$ and $\mathbf{j}$. We have

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

Geometrically, this is a reflection in the line $y = x$.

Next consider (3×3) matrix $E = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Now we have

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} y \\ x \\ z \end{bmatrix}$$

which is reflection in the plane $y = x$.

So these elementary matrices of the first type give reflections. Next, the matrix given by the second type of basic elementary operation gives what we shall call a *sliding transformation*; a more common name is a *skewing transformation*. We explain why we call this "sliding": For an example note that

$$\begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x + ty \\ y \end{bmatrix}$$

which moves the unit square to a parallelogram, sliding the top horizontally by distance t.

From (7) we see that left multiplication by $E = \begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ transforms the parallelopiped $P(\mathbf{i}, a\mathbf{i} + \mathbf{j}, \mathbf{k})$ to $P(\mathbf{i}, \mathbf{j}, \mathbf{k})$ which slides the top of it over so it is the unit cube.

By applying successively these operations we can turn any parallelopiped in $\mathbb{R}^3$ into a rectangular solid. In algebraic terms, this says (equivalently) that we can apply the elementary row operations on a (3×3) invertible matrix M and end up with a diagonal matrix D. We note that by the way we have defined elementary matrices, they all have determinant 1, explained in the next section.

5.16. **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, say, the i^{th} row to be $\sum_{j=1}^{n} (-1)^{i+j} \det A(ij)$ or indeed along any column.

Lemma 5.18. *These are equal, giving the same number whatever row or column chosen!*

Proof. The proof depends on the following two facts:
Fact 1: If $\widetilde{A}$ is the matrix formed from A by exchanging two rows, then $\det \widetilde{A} = (-1)\det A$; and
Fact 2: $\det A^t = \det A$.

Now if we expand along the second row of A, the formula we get is that for expanding along the first row of $\widetilde{A}$, times (-1), which is $-\det A$, and the two sign changes cancel.

If we expand along the first column of A, this is the same formula as expanding along the first row of A^t. This completes the proof.

□

Note that this algorithm also works for the (2×2) case!

Geometric definition:

Definition 5.12. 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. (Here this is *n-dimensional volume* and so is length, area in dimensions $1, 2$).

Theorem 5.19. *The algebraic and geometric definitions are equivalent.*

Proof. For a (2×2) matrix A, 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 algebraically is ad. But the parallelogram area is (base)(height)= ad as well.

The formula area(parallelogram)= (base)(height) is usually proved by cutting off a triangle vertically and shifting it to the other side, thus forming a rectangle of the same base and height. Here is a different way to picture this: imagine the parallelogram is a pile of horizontal layers, like a stack of cards, and straighten the pile to a vertical pile by *sliding* the cards, ending up with the same $(a \times d)$ rectangle. See §5.15 where we define sliding transformations.
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.

We observe that the same procedure works in $\mathbb{R}^3$ and beyond.

We can visualize that a parallelopiped $P(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ in $\mathbb{R}^3$ can be transformed by sliding parallel to these three vectors into a rectangular solid, with sides parallel to the axes. To prove this we again note that the operation of sliding is given algebraically by an elementary column operation, or equivalently by right multiplication by an elementary matrix (of determinant ± 1). And we know that column-reduction (just like row-reduction except the transpose) of an invertible matrix can always be done to end up with a diagonal matrix.

□

5.17. **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 *opposite* orientation. As we shall explain, there are only two choices for this.

A basis $\mathcal{B}_1$ of $\mathbb{R}^n$ is an n-tuple $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ of linearly independent vectors which generate. (Recall that in fact, n L.I. vectors will always generate). From this we form the matrix $M_{\mathcal{B}}$ with those columns. This is an invertible matrix. (Exercise: verify this!)

We let $\widehat{\mathcal{B}}$ denote the collection of all bases of $\mathbb{R}^n$.

The change from one basis $\mathcal{B}_1$ to another $\mathcal{B}_2$ is given by an invertible matrix A. What we mean by this is:

$$M_{\mathcal{B}_1} A = M_{\mathcal{B}_2}.$$

An example is to simply change the order of the basis, defining $\mathcal{B}_2 = (\mathbf{v}_2, \mathbf{v}_1, \mathbf{v}_3, \ldots, \mathbf{v}_n)$. Thus to get the matrix $M_{\mathcal{B}_2}$ we have performed on $M_{\mathcal{B}_1}$ the elementary column operation of switching the first two columns. This is given by right multiplication by the elementary matrix E which is I with the first two columns switched: (!!!)

$$\begin{bmatrix} a & e & * & * \\ b & f & * & * \\ c & g & * & * \\ d & h & * & * \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} e & a & * & * \\ f & b & * & * \\ g & c & * & * \\ h & d & * & * \end{bmatrix} \tag{8}$$

where $*$ stands for any number.

Given our invertible matrix A, we know it is a product of elementary matrices E_i times a diagonal matrix D , since by column reduction we have: $AE_1E_2 \dots E_m = D$. Thus the change of basis can always be given in this way.

Now by definition $GL(n, \mathbb{R})$ is the collection of invertible $(n \times n)$ matrices. The collection of those with $\det A > 0$ is called GL^+; these are the *orientation-preserving* matrices. (From the point of view of Group Theory, GL^+ is a subgroup of index 2 of GL, and its coset is GL^-, the collection (not a subgroup!) of *orientation-reversing* 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, and 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 image of this by applying all elements of GL^+ 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^+-image of any $\mathcal{B}_2$ not in $\widehat{\mathcal{B}^+}$.

Example 12. We give some examples. Suppose $\mathcal{B}_1 = (\mathbf{v}_1, \mathbf{v}_2 \dots, \mathbf{v}_n)$ has a certain orientation (say, positive). Then switching the order of two of them, $(\mathbf{v}_2, \mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_4 \dots, \mathbf{v}_n)$ has the opposite orientation. Also, as just shown above for the (4×4) case, this change is give by multiplication by the matrix A with columns $(\mathbf{e}_2, \mathbf{e}_1, \mathbf{e}_3, \dots, \mathbf{e}_n)$ in other words I with the first two columns interchanged. Note A has determinant -1.

For $\mathbb{R}^3$, this gives: $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ has the opposite orientation from $(\mathbf{j}, \mathbf{i}, \mathbf{k})$.

For a second example, $(-\mathbf{v}_1, \mathbf{v}_2 \dots, \mathbf{v}_n)$ also has opposite orientation from $\mathcal{B}_1$. (Exercise: what is A in this case?)

Geometrically, this is a *reflection* in the subspace generated by the vectors $\{\mathbf{v}_2 \dots, \mathbf{v}_n\}$.

For example, reflection in the xy-plane is the map $R(a, b, c) = (a, b, -c)$ which is given by the matrix $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ (times column vectors), and that sends the standard basis $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ to $(\mathbf{i}, \mathbf{j}, -\mathbf{k})$.

Exercise: write the formula for reflection in a *line*, for example in the $x-$ axis!

How can you find from this the formula for reflection in a general plane, or in a general line? For reflection in the *point* $\mathbf{0}$?

Definition 5.13. The *standard orientation* for $\mathbb{R}^n$ is defined by the choice that $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$ has positve orientation. Thus for $\mathbb{R}^1$ this is $(\mathbf{e}_1)$, for $\mathbb{R}^2$ is $\mathbf{e}_1, \mathbf{e}_2) = (\mathbf{i}, \mathbf{j})$ and for $\mathbb{R}^3$ this is $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = (\mathbf{i}, \mathbf{j}, \mathbf{k})$.

Theorem 5.20.

(i) $det(AB) = det(A)det(B)$.
(ii) $det(B^{-1}AB) = det(A)$.

Proof. Part (i) can be proved algebraically, but it is much easier to use the geometric definition of determinant, that $\det(A) = (\pm 1) \cdot$ (factor of change of volume). (Since $(AB)\mathbf{v} = A(B\mathbf{v})$- this is a multiplication of matrices, and we have the associative law- multiplication of the volume by b and then by a changes it by the factor ab).

Now we have the factor of 1 if A preserves orientation, -1 if not. This again works for the product; changing the orientation twice leaves it fixed, and $(-1)(-1) = 1$.

Part (ii) follows from this. □

5.18. **Three definitions of the vector product.** The vector product $\mathbf{v} \wedge \mathbf{w}$ is defined only on $\mathbb{R}^3$, and gives a vector in $\mathbb{R}^3$. (In $\mathbb{R}^2$, we make the special definition $(a, b) \wedge (c, d) = \mathbf{v} \wedge \mathbf{w} \in \mathbb{R}^3$ where $\mathbf{v} = (a, b, 0)$ and $\mathbf{w} = (c, d, 0)$). Here we present three equivalent definitions of the vector product. We write $\mathbf{i}, \mathbf{j}, \mathbf{k}$ for the standard basis vectors in $\mathbb{R}^3$. We write $P(\mathbf{v}, \mathbf{w}, \mathbf{z})$ for the paralelopiped spanned by $\mathbf{v}, \mathbf{w}, \mathbf{z} \in \mathbb{R}^3$, that is, $P(\mathbf{v}, \mathbf{w}, \mathbf{z}) \equiv \{a\mathbf{v} + b\mathbf{w} + c\mathbf{z} \;:\; a, b, c \in [0, 1]\}$, and $P(\mathbf{v}, \mathbf{w})$ for the parallelogram spanned by $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$, so $P(\mathbf{v}, \mathbf{w}) \equiv \{a\mathbf{v} + b\mathbf{w} : a, b \in [0, 1]\}$.

Theorem 5.21. *The following definitions are equivalent.*
(1) (Via the "determinant" formula)*:*

$$\mathbf{v} \wedge \mathbf{w} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = \mathbf{i} \begin{vmatrix} v_2 & v_3 \\ w_2 & w_3 \end{vmatrix} - \mathbf{j} \begin{vmatrix} v_1 & v_3 \\ w_1 & w_3 \end{vmatrix} + \mathbf{k} \begin{vmatrix} v_1 & v_2 \\ w_1 & w_2 \end{vmatrix}$$

(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* $P(\mathbf{v}, \mathbf{w})$*; thus*

$$||\mathbf{v} \wedge \mathbf{w}|| = ||\mathbf{v}||||\mathbf{w}|| \cdot |\sin(\theta)|.$$

(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* $\mathbf{i} \wedge \mathbf{j} = \mathbf{k}$*,* $\mathbf{j} \wedge \mathbf{k} = \mathbf{i}$*,* $\mathbf{k} \wedge \mathbf{i} = \mathbf{j}$*.*

Remark 5.8. (1) is the usual definition given in texts.

Regarding (2), 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.

The formula in (3) is easy to remember as it follows a circle from $\mathbf{i}$ to $\mathbf{j}$ to $\mathbf{k}$.

Proof. To prove that (1) $\Rightarrow$ (2) we note that for any vector $\mathbf{u}$, we used the *mixed product*, a mixture of the inner and vector products $\mathbf{u} \cdot (\mathbf{v} \wedge \mathbf{w})$, and note that:

$$\mathbf{u}\cdot(\mathbf{v}\wedge\mathbf{w}) = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} \tag{9}$$

Taking $\mathbf{u}=\mathbf{v}$ in (9) 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 parallelopiped is the image of the unit cube, and since from Theorem 5.21 we know the determinant gives $\pm$ (factor of change of volume).

Now taking in (9) $\mathbf{u}=\mathbf{z}=\mathbf{v}\wedge\mathbf{w}$, then $||\mathbf{z}||^2=\mathbf{z}\cdot\mathbf{z}=\begin{vmatrix} z_1 & z_2 & z_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}\geq 0$ so the orientation of $(\mathbf{z},\mathbf{v},\mathbf{w})$ is positive. Using this, from the geometric definition of the determinant,

$$\begin{vmatrix} z_1 & z_2 & z_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = \mathrm{vol}(\mathbf{z},\mathbf{v},\mathbf{w})$$

where this means the volume of the parallelopiped spanned by the basis (if linearly independent) $(\mathbf{z},\mathbf{v},\mathbf{w})$. Here we use the fact that we can exchange rows for columns as $\det A = \det A^t$. But since $\mathbf{z}$ is orthogonal to the base parallelogram, this volume is (base area)(height).

This gives

$$||\mathbf{z}||^2 = \text{(base area)(height)} = \text{(base area)}||\mathbf{z}||$$

so $||\mathbf{z}|| =$ (base area) as claimed. This concludes the proof that Def. (1) implies Def. 2.

It is clear that both Defs. (1), (2) imply Def. (3), but knowing Def. (3) for the basis vectors $\mathbf{i},\mathbf{j},\mathbf{k}$ determines $\mathbf{v}\wedge\mathbf{w}$ for all $\mathbf{v},\mathbf{w}$, by bilinearity. Hence all three are equivalent.

□

Corollary 5.22. *We have the nice (and useful!) formula*

$$||\mathbf{v}\wedge\mathbf{w}||^2 = (\mathbf{v}\cdot\mathbf{v})(\mathbf{w}\cdot\mathbf{w}) - (\mathbf{v}\cdot\mathbf{w})^2.$$

Proof. From Theorem 5.21 we know that

$$||\mathbf{v}\wedge\mathbf{w}||^2 = (\text{area})^2 = (||\mathbf{v}||||\mathbf{w}||| \sin\theta|)^2$$

and this is

$$||\mathbf{v}||^2||\mathbf{w}||^2(1-\cos^2\theta) = ||\mathbf{v}||^2||\mathbf{w}||^2 - (||\mathbf{v}||||\mathbf{w}||\cos\theta)^2 = ||\mathbf{v}||^2||\mathbf{w}||^2 - (\mathbf{v}\cdot\mathbf{w})^2.$$

□

We shall next see how the vector product satisfies three important properties, the first two of which we have already proved:

Definition 5.14. 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*

$$[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = \mathbf{0}$$

Proposition 5.23. *The vector product* $\mathbf{v} \wedge \mathbf{w}$ *on* $\mathbb{R}^3$ *is a Lie bracket, setting* $[\mathbf{v}, \mathbf{w}] = \mathbf{v} \wedge \mathbf{w}$.

Proof. We have shown the first two properties.

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. □

General equation of a plane; matrix form. Going back to the parametric equation of a plane in (10), we had the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ where $T(s, t) = s\mathbf{v} + t\mathbf{w}$. Writing H for the (3×2) matrix and $\mathbf{z}$ for the (2×1) column vector $\mathbf{z} = \begin{bmatrix} s \\ t \end{bmatrix}$, then in matrix form this is

$$H\mathbf{z} = \begin{bmatrix} v_1 & w_1 \\ v_2 & w_2 \\ v_3 & w_3 \end{bmatrix} \begin{bmatrix} s \\ t \end{bmatrix} = s \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} + t \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \tag{10}$$

The image of the map H is a plane which passes through $\mathbf{0}$. The plane parallel to this which passes through some point $\mathbf{p}$ is the image of the function $H_{\mathbf{p}} : (s, t) \mapsto T(s, t) = s\mathbf{v} + t\mathbf{w} + \mathbf{p}$. Note that H is a linear transformation, while $H_{\mathbf{p}}$ is affine but not linear (unless $\mathbf{p}$ happens to lie on the plane $\text{Im}(H)$).

Given this parametric equation we can find the general equation of the plane $\text{Im}(H_{\mathbf{p}})$ as follows: we take our normal vector to be $\mathbf{n} = (A, B, C)$ where $\mathbf{n} = \mathbf{v} \wedge \mathbf{w}$. Then points (x, y, z) in the plane $T(s, t) = s\mathbf{v} + t\mathbf{w} + \mathbf{p} = (x, y, z)$ satisfy the equation

$$(A, B, C) \cdot ((x, y, z) - \mathbf{p}) = 0$$

so giving

$$Ax + By + Cz + D = 0$$

where $D = -\mathbf{n} \cdot \mathbf{p}$.

We have explained this above, in Exercise 5.12.

In matrix form this is

$$M\mathbf{v} = \begin{bmatrix} A & B & C \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -D \end{bmatrix}. \tag{11}$$

Defining $S : \mathbb{R}^3 \to R$ to be the function $S(x, y, z) = Ax + By + Cz$ then the plane is the level surface of level $-D$ of S.

Putting these two maps together, we have the composition of maps, with the two matrices acting on column vectors:

$$\mathcal{M}_{2,1} \xrightarrow{H} \mathcal{M}_{3,1} \xrightarrow{M} \mathcal{M}_{1,1}$$

or as linear transformations:

$$\mathbb{R}^2 \xrightarrow{T} \mathbb{R}^3 \xrightarrow{S} \mathbb{R}^1 \tag{12}$$

Restricting to the image of T, the geometrical plane $P \subseteq \mathbb{R}^3$, we have:

$$\mathbb{R}^2 \xrightarrow{T} P \xrightarrow{S} \{-D\} \tag{13}$$

The plane P is a set of points (x, y, z), which on the one hand is the image of the map T, and on the other is a translate of the kernel of the map S by the vector $\mathbf{p}$.

Level surfaces of different levels (that is, planes which are parallel, with different constants D) fit together as described by Equation (12).

Remark 5.9. **The important point in this is the following:** The plane P is, by itself, simply a subset of points, a two-dimensional subspace of $\mathbb{R}^3$. However, Equation (13) gives us two very different ways of viewing P: via the map T or the map S.

Summarizing, P is the *image* of $\mathbb{R}^2$ via the map T. That is, the map T *parametrizes* P; thus via this map P becomes the parametrized plane $L_{\mathbf{p}}(s,t) = s\mathbf{v} + t\mathbf{w} + \mathbf{p}$.

On the other hand, via the map S, P is the *preimage* (inverse image) of a constant value $-D$. Thus it is is seen to be a *level surface* of the map (of level $-D$). Thus it is only one of a family of parallel planes, of different levels.

This also gives us insight as to the meaning of the diagram: it says something about the object (in this case the plane P) in the middle, from two different perspectives, given by the two maps.

Again, this just reflects the difference between our two ways of understanding a plane, as a parametrized plane, see Equation (10), or as the solution set of its general equation. And this latter is, geometrically, a plane which passes through a point and has a certain normal vector, $\mathbf{n} = (A, B, C)$.

This is the simplest case, of a line in the plane or a plane in space. The general situation comes from these fundamental results of Linear Algebra:

Theorem 5.24. *Given finite-dimensional vector spaces V, W let $T : V \to W$ be a linear transformation. Then:*
i) the null space $\mathcal{N}(T)$ is a vector subspace of V;
ii) the image $Im(T)$ is also; and
(iii) $dim(\mathcal{N}(T)) + dim(Im(T)) = dim(V)$.

Exercise 5.15. Prove (i), (ii)! See Exercise 5.6.

Corollary 5.25. *If T above is surjective, then $dim(\mathcal{N}(T)) = dim(V) - dim(W)$.*

Before we describe the proof, we write it as a diagram, of linear transformations on vector spaces:

$$K \xrightarrow{I} V \xrightarrow{T} W$$

Here the first map I is an *injection* $I(\mathbf{v}) = \mathbf{v}$, which just means that it is a $1-1$ function (just the identity map in this case). Its image is the subspace $\mathrm{Im}(I) = K$ which is the kernel of T, and the image of T is W. That is, the map T is onto.

For the previous example, the map I represents the plane K as a parametrized subspace, while the map T gives its general equation.

In Algebra, a diagram of maps where the image of one map is the kernel of the following map is called an *exact sequence*. In fact, the above diagram of vector spaces extends to

$$\{\mathbf{0}\} \xrightarrow{I} K \xrightarrow{T} V \xrightarrow{S} W \xrightarrow{\pi} \{\mathbf{0}\}$$

where I is the *injection* and π is the projection $\pi(\mathbf{v}) = \mathbf{0}$. This extended diagram is also exact: exactness of the first part

$$\{\mathbf{0}\} \xrightarrow{I} K \xrightarrow{T} V$$

says that I is injective ($1-1$ to its image) since the kernel of T is then $\{\mathbf{0}\}$, while exactness of the second part

$$V \xrightarrow{S} W \xrightarrow{\pi} \{\mathbf{0}\}$$

tells us that the map S is onto (surjective) as the kernel of π is all of W, which by exactness is the image of S.

Back to the proof of the theorem, part (*iii*) can be proved by writing the map as a matrix and solving the system of linear equations.

For example when $m = 3$ and $n = 2$, we have the following.

Given a matrix

$$M = \begin{bmatrix} A & B & C \\ D & E & F \end{bmatrix}$$

we have the matrix equation

$$M\mathbf{v} = \mathbf{w}$$

where $\mathbf{q} = \mathbf{w}$ is fixed, and M is fixed, and by the *solution set* of this equation we mean the collection of all $\mathbf{v}$ which satisfy this equation. Writing $\mathbf{w} = (s, t)$ we have

$$\begin{bmatrix} A & B & C \\ D & E & F \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} s \\ t \end{bmatrix}. \tag{14}$$

The multiplication $\mathbf{v} \mapsto M\mathbf{v}$ defines a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$. Note that $\mathrm{Im}(T)$ is equal to the *column space* of M, the subspace of $\mathbb{R}^3$ generated by the columns of the matrix. This is simply because for a standard column basis vector $\mathbf{e}_k$, $M\mathbf{e}_k$ gives the k^{th} column of M.

Note that the matrix equation (14) is equivalent to the "system of two linear equations in three unknowns":

$$\begin{cases} Ax + By + Cz = s \\ Dx + Ey + Fz = t \end{cases}$$

This system has *full rank* iff the rows are linearly independent, iff the dimension of the image $\text{Im}(T)$ is the maximum possible, in this case 2.

From Linear Algebra we can find the solution set explicitly by *row reduction.*

Example 13. For a concrete example, after row reduction we may have the matrix equation

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} s \\ t \end{bmatrix}. \tag{15}$$

which gives the system

$$\begin{cases} x + y = s \\ 2z = t \end{cases}$$

and we are free to choose y (for this reason known as a "free variable") but then no longer free to choose x or z as these are determined, since $x = -y + s$ and $z = t/2$.

Thus we have for a solution

$$(x, y, z) = (-y + s, y, t/2) = (s, 0, t/2) + y(-1, 1, 0) = \mathbf{p} + y\mathbf{v} = l(y)$$

which is a parametrized line passing through the point $\mathbf{p}$ in the direction of $\mathbf{v}$.

If we change s, t we get lines parallel to this one. In particular, if $\mathbf{p} = (0, 0)$ then the solution set is the parametric line $l(y) = y\mathbf{v}$, and this is the kernel of the map T, of dimension 1.

In conclusion, the dimension of the solution set is the number of free variables, so in this case of full rank this is, by Cor. 5.25, $3 - 2 = 1$, indeed a line.

Fixing the constants in the system to be $s = 1, t = 4$, the parametric line has the matrix form

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = y \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}. \tag{16}$$

The composition of the two maps is

$$\begin{aligned} [y] \mapsto & \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \left(y \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \right) \\ &= y \left(\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right) + \left(\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \right) \\ &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \end{bmatrix} \end{aligned} \tag{17}$$

which is correct, since the every point on the parametric line is in the solution set of the system.

The geometrical way to think of this is that each of the equations in the system gives a plane, having as normal vector the corrsponding row of the matrix, so the solutions for the pair of equations is the intersection of two planes which is a line, which is perpendicular to both row vectors. The full rank condition means that these planes are not parallel, since their normal vectors are the rows of M, which are linearly independent.

5.19. **The Inverse and Implicit Function Theorems.** What we will see next is how this same point of view applies for the much more general situation of a differentiable, but *nonlinear*, map.

We know that given $F : \mathbb{R}^m \to \mathbb{R}^n$ continuously differentiable, the derivative matrix $DF|_{\mathbf{p}}$ well–approximates the function at the point $\mathbf{p}$. That means that certain properties of F near $\mathbf{p}$ should be reflected in the linear map $DF|_{\mathbf{p}}$ and vice-versa.

Important examples are given by these two theorems, which are closely related. Indeed one can choose to prove either one first, then deducing the other from that.

The *Inverse Function Theorem* states that $F : \mathbb{R}^m \to \mathbb{R}^m$ is invertible near $\mathbf{p}$ iff the matrix $DF|_{\mathbf{p}}$ is invertible, which is true iff $\det(DF|_{\mathbf{p}}) \neq 0$. First we need:

Definition 5.15. $F : \mathbb{R}^m \to \mathbb{R}^n$ is *continuously differentiable* (of class $\mathcal{C}^1$) iff the derivative DF at each point $\mathbf{p}$ exists and the matrix $DF_{\mathbf{p}}$ is a continuous function of $\mathbf{p}$).

Given an open set $\mathcal{U} \subseteq \mathbb{R}^m$, a function $F : \mathcal{U} \to \mathcal{V} = F(\mathcal{U})$ is *invertible* iff there exists $\widetilde{F}$ defined on $\mathcal{V}$ such that $\widetilde{F} \circ F$ is the identity on $\mathcal{U}$ and $F \circ \widetilde{F}$ is the identity on $\mathcal{V}$.

A *parametrized submanifold* $M \subseteq \mathbb{R}^m$ of dimension $d < m$ is the following: there exists $\mathcal{U} \subseteq \mathbb{R}^d$ and $\Phi : \mathcal{U} \to \mathbb{R}^d$ such that Φ is $\mathcal{C}^1$, M is the image $\Phi(\mathcal{U})$, and such that at each point of $\mathcal{U}$,$D\Phi$ is injective to a linear subspace of $\mathbb{R}^m$ of dimension d.

Example 14. A $\mathcal{C}^1$ curve γ in $\mathbb{R}^m$ is a parametrized submanifold of $\mathbb{R}^m$ of dimension one. A parametrized surface in $\mathbb{R}^m$ is a $\mathcal{C}^1$ function S from an open subset $\mathcal{U} \subseteq \mathbb{R}^2$ to its image $\mathbb{S} \subseteq \mathbb{R}^m$ such that the derivative matrix $DS|_{\mathbf{p}}$ is injective for every $\mathbf{p} \in \mathcal{U}$.

Examples are given by cylindrical and spherical coordinates, for a cylinder and sphere respectively:

Cylindrical coordinates on $\mathbb{R}^3$ are given by the map $\alpha : \mathbb{R}^3 \to \mathbb{R}^3$ with $\alpha(r, \theta, z) = (x, y, z)$ where

$$\begin{aligned} x &= r\cos(\theta) \\ y &= r\sin(\theta) \\ z &= z \end{aligned} \tag{18}$$

The cylinder of radius a and height c is the image by α of the subset $r = a, \theta \in [0, 2\pi), z \in [0, c]$. Defining the map S on $[0, 2\pi) \times [0, c]$ by $S(\theta, z) = (a\cos\theta, a\sin\theta, z)$ gives the cylinder as a parametrized surface.

Exercise 5.16. Check that the derivative matrix is:

$$D\alpha = \begin{bmatrix} x_r & x_\theta & x_z \\ y_r & y_\theta & y_z \\ z_r & z_\theta & z_z \end{bmatrix} = \begin{bmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{19}$$

Calculate the determinate of this matrix.

Spherical coordinates on $\mathbb{R}^3$ are given by the map $\beta : \mathbb{R}^3 \to \mathbb{R}^3$ with $\widetilde{S}(\rho,\theta,\varphi) = (x,y,z)$ where $r = \rho\sin(\varphi)$ and so

$$\begin{aligned} x &= r\cos(\theta) = \rho\sin(\varphi)\cos(\theta)\\ y &= r\sin(\theta) = \rho\sin(\varphi)\sin(\theta)\\ z &= \rho\sin(\varphi) \end{aligned} \tag{20}$$

The sphere of radius a is the image by α of the subset $\rho = a, \theta \in [0,2\pi), \varphi \in [0,\pi]$. Defining the map S on $[0,2\pi) \times [0,\pi]$ by $\widetilde{S}(\theta,\varphi) = (a,\theta,\varphi)$ gives the sphere as a parametrized surface.

Exercise 5.17. Calculate the derivative matrix and its determinant.

Theorem 5.26. *(Inverse Function Theorem) Let $F : \mathbb{R}^m \to \mathbb{R}^m$ be $\mathcal{C}^1$. Suppose the matrix $DF_{\mathbf{p}}$ is invertible. Then there exists an open set $\mathcal{U}$ containing $\mathbf{p}$ such that F is $\mathcal{C}^1$ and invertibl e on $\mathcal{U}$.*

Proof. See [Mar74], p. 206 and p. 230 or (for a stronger statement, with estimates) [HH15] p. 264 ff. □

Remark 5.10. Note that by the Chain Rule, we then know that for all points $\mathbf{x} \in \mathcal{U}$, with $\mathbf{y} = F(\mathbf{x})$, $\widetilde{F} \circ F(\mathbf{x}) = \mathbf{x}$ so $I = D(F \circ \widetilde{F})(\mathbf{x}) = (DF)_{\widetilde{F}(\mathbf{x})} DF_{\mathbf{x}}$ whence the inverse of the matrix $DF_{\mathbf{x}}$ is $(DF_{\mathbf{x}})^{-1} = D\widetilde{F}_{\mathbf{y}}$.

The *Implicit Function Theorem* states that for a $\mathcal{C}^2$ function $F : \mathbb{R}^m \to \mathbb{R}^n$ with $m \geq n$, then if the derivative matrix at a point $\mathbf{p}$ is surjective (onto; of full rank) then the inverse image set $F^{-1}(\mathbf{q})$ for $F(\mathbf{p}) = \mathbf{q}$ behaves like the inverse image of a point by the matrix: it is a *submanifold* of dimension $m - n$. For the linear (matrix) case see Cor. 5.25.

A submanifold of dimension 1 is a parametrized curve; of dimension 2 is a *parametrized surface*. Note that the case $m = n$ says the following: the inverse image of a point $F^{-1}(\mathbf{q})$ is a manifold of dimension $m - n = 0$, in other words a single point. Thus F is an invertible function, which is just the statement of the Inverse Function Theorem!

In the case $n = 1$, the Implicit Function Theorem moreover gives conditions when given an equation

$$F(x_1, \dots, x_n) = 0$$

we can solve for one of the variables, and use the rest as our parameters.

Example 15. For the simplest example, $F : \mathbb{R}^2 \to \mathbb{R}$ with $F(x,y) = x^2 + y^2$, the curve of level 1 is the unit circle, the solutions of the equation (i.e. all pairs (x,y) which satisfy the equation)

$$x^2 + y^2 = 1.$$

Solving for y gives $y = \pm\sqrt{1 - x^2}$. See Exercise 5.7.

Theorem 5.27. *(Implicit Function Theorem for $\mathbb{R}^m$) Let $F : \mathbb{R}^m \to \mathbb{R}^n$ be $\mathcal{C}^1$, with $m \geq n$. Suppose the matrix $DF_{\mathbf{p}}$ is surjective. Then for $\mathbf{q} = F(\mathbf{p})$, the set $F^{-1}(\mathbf{q})$ is a submanifold of $\mathbb{R}^m$ of dimension $m - n$. That is, for $\mathbf{p} \in F^{-1}(\mathbf{q})$, there exists an open subset $\mathcal{U} \subseteq \mathbb{R}^{m-n}$ and $H : \mathcal{U} \to \mathbb{R}^m$ $\mathbb{C}^1$ such that $F \circ G(\mathbf{p}) = \mathbf{q}$ for all $\mathbf{p} \in \mathcal{U}$.*

Proof. See [War71] Theorem 1.38, p. 31. □

For example, if $F : \mathbb{R}^3 \to \mathbb{R}$, then the level surface $F^{-1}(q)$ is a parametrized surface: its parametrization is given near the point $\mathbf{p}$ by the map H.

Remark 5.11. Similarly to the linear case as explained in Remark 5.9, in the smooth case, the same set (an embedded manifold) is viewed in two different ways, by means the two maps, one where it is the image, one the domain. The first parametrizes the manifold, the second places it as a level curve, surface or manifold of a map on the higher-dimensional space, and thus shows how it is but one of a family of such "parallel" manifolds. This is a special mathematical object known as a *foliation.* Thus the level surfaces of a function $F : \mathbb{R}^3 \to \mathbb{R}$ foliate $\mathbb{R}^3$, and in the special case of F linear, F is given by the inner product with a normal vector, and the foliation consists of all those parallel planes.

Thus a *parametrized m-dimensional manifold in $\mathbb{R}^m$* is a map $\alpha : \mathcal{U} \to M \subset \mathbb{R}^m$ where $\mathcal{U}$ is a connected open subset of $\mathbb{R}^m$, α is differentiable and invertible with image M. (For the definition of connectedness see Def. **??**).

The higher dimensional version of level curves and surfaces can be stated as follows. Given $f : \mathbb{R}^{m+1} \to \mathbb{R}$ then if f is differentiable and surjective and Df is everywhere onto (one says Df is of *maximal rank*) then for any point $q \in \mathbb{R}$ the set $M = f^{-1}(q)$ is locally a parametrized m-dimensional manifold. Moreover this holds when that condition holds at some point (not necessarily all points): for any $y = f(\mathbf{p})$ such that $Df|_{\mathbf{p}}$ is of maximal rank; q is then called a *regular value.*

Here is a statement of the

Proposition 5.28. *(Implicit Function Theorem for differentiable manifolds) (Lemmas 1,2 of Chapter 2 of* [MW97]*) If $f : M \to N$ is a smooth map between manifolds of dimension $m \geq n$, and if $\mathbf{q} \in N$ is a regular value, such that it is a value of the map (i.e. $f^{-1}(\mathbf{q})$ is nonempty) then the set $f^{-1}(\mathbf{q})$ is a smooth manifold of dimension $m-n$. The null space of $Df_{\mathbf{p}} : TM_{\mathbf{p}} \to TN_{\mathbf{q}}$ is the tangent space of this submanifold, and its orthogonal complement is mapped onto $TN_{\mathbf{q}}$.*

This says that we have a diagram

$$K \xrightarrow{\alpha} M \xrightarrow{f} N$$

where the first map is injective and the second is surjective. When one considers the derivative maps then one gets the exact diagram for the linear case of Example 13; here $\alpha(\mathbf{x}) = \mathbf{p}$. See Cor. 5.25:

$$\{\mathbf{0}\} \longrightarrow K \xrightarrow{D\alpha|_{\mathbf{x}}} M \xrightarrow{Df_{\mathbf{p}}} N \longrightarrow \{\mathbf{0}\}$$

There are many versions of these theorems. For an introduction see Lemmas 1,2 of Chapter 2 of [MW97] and for surfaces Proposition 2 of Chapter 2, p. 59 of [DC16]. For a simple and beautiful general statement see Theorem 1.39 of [War71]. More on the Implicit and related Inverse Function Theorems are given e.g. in §7.2-4 of [Mar74], and in Chapter 2.10 and on p. 729 of [HH15].

Theorem 5.27 raises implicitly the question of, for a given $F : \mathbb{R}^m \to \mathbb{R}^n$, *which* of the variables $x_1, \dots, \mathbf{x}_m$ can be used to parametrize the submanifold. The basic principle is that the answer is the same as for the linear algebra; that is, we look at the matrix $DF_{\mathbf{p}}$. For a modification of Example 13 above,

$$\begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \mathbf{q} \tag{21}$$

we have $3 - 2 = 1$ for the dimension of the solution set $K = F^{-1}(\mathbf{q})$, and could take x or y as the free variable, but not z. The geometrical reason is that the solution set is $l(y) = \mathbf{p} + y\mathbf{v}$ where $\mathbf{p} = (1, 0, 1)$ and $\mathbf{v} = (-1, 1, 0)$. This is a line in the plane $z = 0$ so altough the solution set is one-dimensional, it can't be parametrized by z, but only (in terms of the standard coordinates) by x or y.

Proposition 5.29. *If the* (2×3) *matrix*

$$A = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \tag{22}$$

has rank 2*, then the solution set has dimension* $3-2 = 1$*; and moreover if the* (2×2) *submatrix*

$$\begin{bmatrix} b & c \\ e & f \end{bmatrix}$$

is invertible, then we can take x *as the free variable (the parameter), and similarly for the other* (2×2) *submatrices.*

A nice simpler version of the Implicit Function Theorem, with examples like this, is given on pp. 239-240 of Vol. II of [Gui02]. See also my handwritten *Notas de Aula.* For a general statement like this example, see Hubbard's book Thm. 2.10.14.

5.20. **Higher derivatives.** So far, since §5.9, we have been studying the derivative of a map $F : \mathbb{R}^m \to \mathbb{R}^n$, using the matrix of partial derivatives. This gives the best linear (or affine) approximation to the map, which is also called the best *first-order* approximation. Thus we have the tangent line to the graph of $f : \mathbb{R} \to \mathbb{R}$, the tangent line to a curve $\gamma : \mathbb{R} \to \mathbb{R}^n$, and the tangent plane to the graph of $F : \mathcal{F}^2 \to \mathbb{R}$.

For the best *second-order* approximation we add a term involving the second-order partial derivatives and so on. This gets more and more complicated as we describe next.

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. For this we need to define higher derivatives.

Writing $L(V, W)$ for the collection of linear transformations from V to W, then this is a Banach space with the operator norm. Now for $F : V \to W$, the derivative DF is, at each point $\mathbf{x}$, a linear map from V to W, thus an element of the vector space $L(V, W)$. In other words, given $F : V \to W$, then DF is a map (*nonlinear* in general) from V to $L(V, W)$. Since $DF : V \to L(V, W)$, then the second derivative at the point $\mathbf{x}$ must be a *linear map* $D^2F_{\mathbf{x}} : V \to L(V, W)$. This means that $D^2F_{\mathbf{x}}$ is in the collection of linear maps from V to $L(V, W)$, in other words it is a nonlinear map $D^2F : V \to L(V, L(V, W))$.

Since we can represent the derivative at a point by a matrix, we see that these are increasing in size: If $V = \mathbb{R}^m$ and $W = \mathbb{R}^n$, then $DF_{\mathbf{x}}$ is $(n \times m)$, $D^2F_{\mathbf{x}}$ is $(n \times (nm))$, $D^3F_{\mathbf{x}}$ is $(n \times (n^2m))$; so $D^kF_{\mathbf{x}}$ is a matrix with n^km entries, getting more and more complicated quickly.

The only exception is when $n = 1$, for a curve $\gamma : [a,b] \to \mathbb{R}^m$: in this case (as noted above) then γ' is also curve in $\mathbb{R}^m$, thus so is $\gamma'' = (\gamma')'$ etcetera. By contrast for a function $F : \mathbb{R}^n \to \mathbb{R}$ then the gradient $\nabla F : \mathbb{R}^n \to \mathbb{R}^n$ is a vector field, so $DF : \mathbb{R}^n \to \mathbb{R}^{n^2}$, and then the second derivative is no longer a vector field as $D^2F : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^{n^2} \sim \mathbb{R}^{n^3}$, and so on, with the matrices getting larger and larger.

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^kF exists and is continuous, so $D : \mathcal{C}^{k+1} \to \mathcal{C}^k$).

5.21. **Higher order partials.** Given $F : \mathbb{R}^2 \to \mathbb{R}$ then $F_x = \frac{\partial F}{\partial x}$ is a function defined on the plane. Then setting $G(x,y) = F_x(x,y)$ we take its partial derivatives. We write $\frac{\partial}{\partial y}G$ in these equivalent ways:

$$G_y = \frac{\partial}{\partial y}(G) = \frac{\partial}{\partial y}(F_x) = \frac{\partial F_x}{\partial y} = (F_x)_y = F_{yx}.$$

(This notation can be confusing since $F_{yx} = (F_x)_y$!)

Now for $G = F_x$ then $G : \mathbb{R}^2 \to \mathbb{R}$. This has as its gradient

$$\nabla G = (G_x, G_y) = (F_{xx}, F_{yx}).$$

Similarly for $\widetilde{G} = F_y$ then its gradient is

$$\nabla \widetilde{G} = (\widetilde{G}_x, \widetilde{G}_y) = (F_{xy}, F_{yy}).$$

When a function $L : \mathbb{R}^2 \to \mathbb{R}^2$ then writing it in components $L = (L_1, L_2)$ or in matrix form

$$[L] = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix}$$

we know its derivative matrix is

$$[DL] = \begin{bmatrix} \nabla L_1 \\ \nabla L_2 \end{bmatrix} = \begin{bmatrix} (L_1)_x & (L_1)_y \\ (Ł_2)_x & (L_2)_y \end{bmatrix}$$

So for $DF : \mathbb{R}^2 \to \mathbb{R}^2$ we have

$$D^2F = D(DF) = \begin{bmatrix} (F_x)_x & (F_x)_y \\ (F_y)_x & (F_y)_y \end{bmatrix} = \begin{bmatrix} F_{xx} & F_{yx} \\ F_{xy} & F_{yy} \end{bmatrix}.$$

Now in fact it is a bit simpler than this because of the following:

Proposition 5.30. *For $F : \mathbb{R}^2 \to \mathbb{R}$ then $F_{xy} = F_{yx}$.*

This important fact is often called the *equality of mixed partials.* This can be proved using just derivatives, but we prefer the "Fubini's Theorem argument" given below when we treat double integration. See Lemma 6.1.

Corollary 5.31. *The above matrix D^2F is* symmetric*: it has the form* $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$.

In the case of $F : \mathbb{R}^n \to \mathbb{R}$, all of this makes sense: D^2F is a a symmetric $(n \times n)$ matrix. For $F : \mathbb{R}^n \to \mathbb{R}$ the matrix D^2F is sometimes called the *Hessian matrix.* Its determinant is called the *Hessian determinant* or simply the *Hessian.* In Guidorizzi Vol 2 §16.3 this is written $H(x, y)$. See [Gui02], and §3.6 of [HH15].

The meaning of this symmetric matrix becomes clear when discussing Taylor polynomials of order 2, and finding maximums and minimums.

5.22. **Finding maximums and minimums.** We note that:
(1) Given $F : \mathbb{R}^2 \to \mathbb{R}$, if a minimum or maximum value occurs at $\mathbf{p} = (x_0, y_0)$ then the tangent plane must be horizontal. Equivalently, if F is differentiable, then the partial derivatives F_x, F_y at $\mathbf{p}$ are 0.

Definition 5.16. In this case, $\mathbf{p}$ is a *critical point (ponto critico)* of F. Equivalently, $\nabla F(\mathbf{p}) = \mathbf{0}$.

(2) If it is a maximum then it must be a maximum for the function restricted to the line $x = x_0$. We can then consider the second partials and use the second derivative test from Calculus 1: If $F_{xx} > 0$ then F_x is increasing, so it is a minimum along that line. This does not necessarily mean it is a minimum off the line.

However there is a fuller method: see Guidorizzi Vol 2 §16.3 [Gui02], and §3.6 of [HH15].

Theorem 5.32. *Let $F : \mathbb{R}^2 \to \mathbb{R}$ be in $\mathbb{C}^2$. We have:*

1. If F has a local max or min at $\mathbf{p}$ then the tangent plane there is horizontal. Equivalently, $F_x = F_y = 0$ at $\mathbf{p}$.

2. The second partials tell us the following:

(i) If $F_{xx} > 0$ at $\mathbf{p}$ and the Hessian $H(\mathbf{p}) > 0$ then $\mathbf{p}$ is a local minimum.

(i) If $F_{xx} < 0$ at $\mathbf{p}$ and the Hessian $H(\mathbf{p}) > 0$ then $\mathbf{p}$ is a local maximum.

(iii) if $H(\mathbf{p}) < 0$ then $\mathbf{p}$ is a saddle point. Thus it is neither max nor min.

(iv) If $H(\mathbf{p}) = 0$ then we cannot say from this test and have to look more closely.

Exercise 5.18. Compare the above tests for the functions we have encountered, see Figs. 3, 2: $F(x, y) = x^2 + y^2$, $F(x, y) = x^2 - y^2$, $F(x, y) = xy$.

The best way to understand this theorem is via the Taylor polynomials of order two, in two variables. First we examine the case of a single variable, and then return to the above.

5.23. **The Taylor polynomial and Taylor series. Taylor series in one dimension.**

Let us recall that a *polynomial of degree n* is

$$p(x) = a_0 + a_1x + \cdots + a_kx^k + \cdots + a_nx^n.$$

Here k is a nonnegtive integer and $a_k \in \mathbb{R}$. So since $x^0 = 1$ for any $x \in \mathbb{R}$, this is equal to

$$p(x) = a_0x^0 + a_1x^1 + \cdots + a_nx^n = \sum_{k=0}^{n} a_kx^k.$$

(Here we use the definition $0! = 1$.) Thus a polynomial of degree 0 is a constant function $p_0(x) = a_0$, of degree one is an affine function, $p_1(x) = a_0 + a_1 x$, of degree two is quadratic and so on.

Polynomials are great to work with as it is easy to compute their derivatives, integrals and to draw their graphs, and to compute their values at a point you only need to add and multiply. Also, as we describe in this section, more complicated functions (for example, the *transcendental functions* $\sin, \cos, \exp, \log, \tan \ldots$ and the *rational functions* (ratios of polynomials, such as $1/(1-x)$) can often be approximated quite well by polynomials.

Given a map $f : \mathbb{R} \to \mathbb{R}$, the terminology "$k^{\text{th}}$-order approximation" to f at a point $x \in \mathbb{R}$ comes from approximating the function with a certain polynomial of degree k. The polynomial of degree k which best fits the map near that point is termed "the best k^{th}-order approximation at $x \in \mathbb{R}$". This is is called the Taylor polynomial of order k, and is the polynomial which has all the same derivatives at that point, up to order k, as f. The Taylor series is a kind of "polynomial of infinite order" which can, in the nicest cases (for example, $f = \sin, \cos, \exp$) reproduce it exactly as a convergent infinite series.

Thus the best 0^{th}-order approximation of f at $p \in \mathbb{R}$ is the constant map with constant equal to the value of f at that point: the map $p_0(x) = f(p)$ for all x. The $\varepsilon - \delta$ definition of continuity guarantees that this approximates the function fairly well if it is continuous at that point. If f is differentiable, we can do much better: adding on the linear map given by the derivative $f'(p)$ gives the best *first-order approximation.* This is the affine map

$$p_1(x) = f(p) + f'(p)(x - p)$$

whose graph is the tangent line to the graph of f at that point.

For a function $f : \mathbb{R} \to \mathbb{R}$ in $\mathcal{C}^\infty$, meaning it has derivatives of all orders, we define a sequence of polynomials $p_n(x)$, each of degree n, which approximate f better and better as $n \to \infty$. For this we choose a point about which we make the approximation, and call this the *Taylor polynomial about* x_0. Here for simplicity we work with $x_0 = 0$, and remark that the Taylor polynomials in this case are also called *Maclaurin polynomials.*

We write p_n for the n^{th} Taylor polynomial (about 0). We also say this is the *best n^{th}-order approximation* to f.

In the nicest cases, p_n actually converges to f as $n \to \infty$. This is true, for example, for $f(x) = \sin(x), \cos(x), e^x$.

The *Taylor series* is the infinite series which can be thought of as an *infinite polynomial.* For example, the n^{th} Taylor polynomial for e^x is

$$1 + x + x^2 + x^3/3! + \cdots + x^n/n!$$

and the Taylor series is

$$e^x = 1 + x + x^2 + x^3/3! + \cdots + x^n/n! + \cdots = \sum_{k=0}^{\infty} \frac{x^k}{k!} \tag{23}$$

For $f(x) = \sin x$ we have

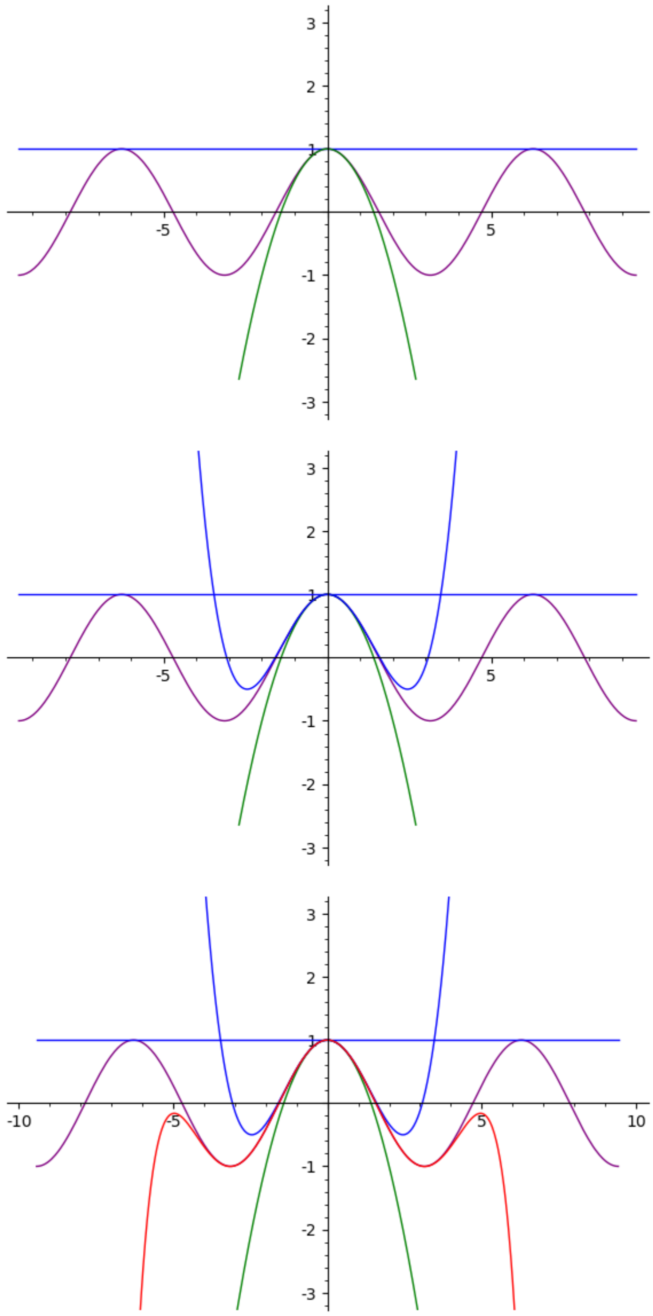


FIGURE 10. $\cos(x)$ and its Taylor polynomials p_n for $n = 0, 2, 4, 10$.

$$\sin(x) = x - x^3/3! + x^5/5! - \dots$$

with only the odd powers and

$$\cos(x) = 1 - x^2/2! + x^4/4! - \dots$$

with only the even powers, and both with alternating signs.

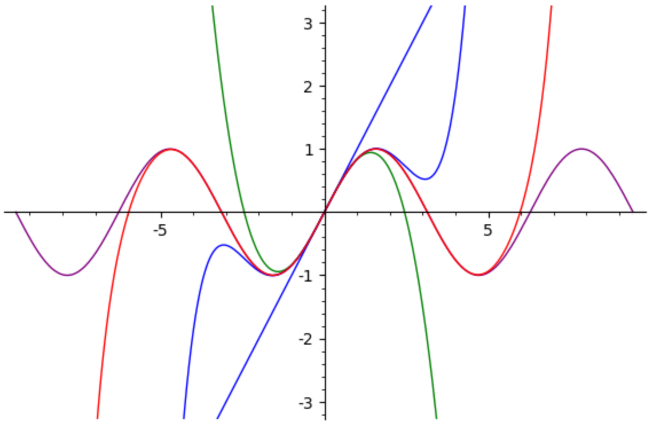


FIGURE 11. $\sin(x)$ and its Taylor polynomials p_n for $n = 1, 3, 5, 13$.

That the graphs of the polynomials p_n do approach the function can be seen in Figs. 10, 11. The 0^{th} -order approximation is a horizontal line with that value, the 1^{st}-order approximation is the tangent line to the graph at that point. The 2^{nd}-order approximation is the parabola which best fits the curve, and so on. In Fig. 10, 11 we show the functions $f = \cos$ and $\sin$ together with some of the Taylor polynomials p_n. Note how close the fit becomes as n increases!

Exercise 5.19. Check that for e^x the derivative of p_{n+1} is p_n, and that for $\sin(x)$ the derivative of p_{n+1} is p_n for $\cos(x)$. This agrees with $(e^x)' = e^x$ and $(\sin)' = \cos, (\cos)' = -\sin$!!

The definition of the Taylor series for a differentiable function (about 0) is

$$\sum_{k=0}^{\infty} a_k x^k$$

where $a_k = f^{(k)}(0)/k!$ (here $f^{(k)}(0)$ is the k^{th} derivative of f at 0)). (Note that we write $f^{(0)}$ for f itself).

Thus the Taylor polynomial of degree n is

$$p_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k.$$

Exercise 5.20. (1) Check that this general formula does give the above Taylor series for e^x, $\sin(x)$ and $\cos(x)$.
(2) Show that the polynomial p_n has the same derivatives as f at $x = 0$, of order $0, 1, \ldots, n$. That is, $p_n(0) = f(0), p_n'(0) = f'(0), p_n''(0) = f''(0)$, and so on.

To define the *Taylor polynomials about* x_0 we simply replace x by $(x - x_0)$ and the derivatives at 0 by those at x_0. Thus the Taylor polynomial of degree n about x_0 is

$$p_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k.$$

The Taylor polynomial in higher dimensions.

We can now understand the role of the Hessian matrix and Hessian determinant in understanding maxes and mins much better, by understanding the Taylor polynomial of a function of 2 variables.

Consider $F : \mathbb{R}^2 \to \mathbb{R}$. Then the best 0^{th} -order approximation about $\mathbf{0} = (0,0)$ is the constant function with the value $f(\mathbf{0})$. The the 1^{st}-order approximation is the tangent *plane* to the graph at $\mathbf{0}$. The best 2^{nd}-order approximation may for example be a paraboloid but could instead be a parabolic hyperboloid, Fig. 2. This depends on the partial derivatives of order 2 at the point.

The Taylor polynomials p_n will be functions of two variables (x, y), thus $p_n : \mathbb{R}^2 \to \mathbb{R}$. Just as for one dimension, a polynomial of degree n is a linear combination of *basic terms* of degree k for $k = 0$ (a constant) up to $k = n$. Each basic term of degree k will be of the form $x^i y^j$ such that $(i + j) = k$. Thus for example the degree of x^2y^3 is $2 + 3 = 5$, and the basic polynomials of degree 1 are $p(x, y) = x, p(x, y) = y$ and of order 2 are $p(x, y) = x^2, p(x, y) = y^2$ and $p(x, y) = xy$.

Taking a linear combination of terms of degree $\leq n$ gives a polynomial of degree n, for example

$$p(x, y) = 1 + x + 3y + x^2 + y^2 + 5xy$$

has degree 2.

The only polynomials in two variables of degree 0 are the constant functions, $p(x, y) = c$. Those of degree one have the form

$$p(x, y) = c + ax + by.$$

Thus the graph is a plane. For the case of the Taylor polynomial, this will be the tangent plane at that point.

Those with *only* terms of degrre two have a special name:

Definition 5.17. A *quadratic form* on $\mathbb{R}^2$ is a degree-two polynomial with no constant or linear terms, hence of the form

$$Q(x, y) = ax^2 + by^2 + cxy.$$

That is, it is a linear combination of all the possible terms of degree 2.

Consider for example $p(x) = x^2 + y^2$. Its graph is a paraboloid, while the graph of $q(x) = xy$ is a hyperbolic paraboloid. See Figs. 3, 2.

Both of these polynomials have degree 2. Both have a horizontal tangent plane at $\mathbf{0}$. The first has a minimum there while the second is a saddle point hence neither max nor min. For $q(x, y) = xy$, on the line $x = y$ we have $q(x, y) = xy = x^2$, an upward parabola, so a minimum along the line $x = y$. When $x = -y$ we have $q(x, y) = -x^2$, so a maximum. Thus $(0, 0)$ can be neither max nor min. This is the essence of a saddle point. Furthermore the level curves are hyperbolas; see Fig. 3.

In fact the terms of order 2 can be understood with the help of the Hessian, a key point being that this is a symmetric matrix.

Proposition 5.33. *Given a quadratic form Q, there is a symmetric (2×2) matrix A such that for $\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix}$, then*

$$Q(\mathbf{v}) = \mathbf{v}^t A \mathbf{v}.$$

That is,

$$Q(\mathbf{v}) = \begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix}.$$

Proof. In fact, for $A = \begin{bmatrix} a & c/2 \\ c/2 & b \end{bmatrix}$ we have

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & c/2 \\ c/2 & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax^2 + by^2 + cxy \end{bmatrix} = Q(x, y).$$

□

Exercise 5.21. Check that when $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ then $Q(x, y) = 2xy$. What do we get for $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$? For $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$?

To graph a quadratic form we have the following:

Theorem 5.34. *A quadratic form*

$$Q(x, y) = ax^2 + by^2 + cxy = \begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix}$$

has either
(i) a local min or max at $\mathbf{0}$, *if* $detA > 0$;
(ii) a saddle point at $\mathbf{0}$, *if* $detA < 0$.
If $detA = 0$, *we cannot tell from this test.*

Proof. (Sketch) From Linear Algebra, a symmetric matrix A can be diagonalized: there exists an orthogonal matrix U such that $U^{-1}AU = D$ where D is diagonal. Now an orthogonal matrix is a rotation, a reflection, or a product of these. That does not change whether $\mathbf{0}$ is a saddle point, max or min. Also, $\det D = \det U^{-1}AU = \det A$. This proves that $\det A$ is the product of its eigenvalues, since the eigenvalues of A and D are the same. The graph of the quadratic form defined by D has the two types described, completing the proof. See the above examples. □

We use this to study the Taylor polynomial of order 2 of $F : \mathbb{R}^2 \to \mathbb{R}$. Write A for the Hessian matrix. Then

$$A = D^2 F = \begin{bmatrix} F_{xx} & F_{yx} \\ F_{xy} & F_{yy} \end{bmatrix} \tag{24}$$

as explained above in (24); note that A is symmetric!

Then the Taylor polynomials about $\mathbf{p} = (x_0, y_0)$ are, for degree 0:

$$p_0(x, y) = F(\mathbf{p})$$

For degree 1, writing $h = (x - x_0), k = (y - y_0)$ then

$$p_1(x,y) = F(\mathbf{p}) + F_x(\mathbf{p})(x-x_0) + F_y(\mathbf{p})(y-y_0) = F(\mathbf{p}) + F_x(\mathbf{p})h + F_y(\mathbf{p})k$$

(this is the tangent plane);
For degree 2 we have

$$p_2(x,y) = F(\mathbf{p}) + F_x(\mathbf{p})h + F_y(\mathbf{p})k + \frac{1}{2}\begin{bmatrix} h & k \end{bmatrix} \begin{bmatrix} F_{xx} & F_{yx} \\ F_{xy} & F_{yy} \end{bmatrix} \begin{bmatrix} h \\ k \end{bmatrix}$$

$$= F(\mathbf{p}) + F_x(\mathbf{p})h + F_y(\mathbf{p})k + \frac{1}{2}(F_{xx}h^2 + 2F_{xy}hk + F_{yy}k^2)$$

Thus

$$p_2(x,y) = p_1(x,y) + \begin{bmatrix} h & k \end{bmatrix} \frac{D^2F(\mathbf{p})}{2} \begin{bmatrix} h \\ k \end{bmatrix}$$

which reminds us of the formula for dimension 1. Note that the last term is a quadratic form since the Hessian matrix D^2F is symmetric.

Looking for maximum and minimum points, first we see if the tangent plane is horizontal. Then the first-order term is 0 so the Taylor polynomial is simply

$$p_2(x,y) = F(\mathbf{p}) + \begin{bmatrix} h & k \end{bmatrix} \frac{D^2F(\mathbf{p})}{2} \begin{bmatrix} h \\ k \end{bmatrix} = F(\mathbf{p}) + Q(x,y)$$

where Q is a quadratic form.

As shown in Theorem 5.34, the sign of the determinant then tells us whether the surface is a max or min, as for a paraboloid or *elliptic paraboloid* (like a paraboloid but with an ellipse as crossection) or a saddle point, as for $F(x,y) = xy$, discussed above.

For $F : \mathbb{R}^n \to \mathbb{R}$ with $n \geq 2$ a similar formula can be given. The higher terms of the Taylor series also have a nice expression when the derivatives D^kF are viewed as k-linear functions. For a clear treatment see [Mar74]. See also §3.6 of [HH15].

Remark 5.12. The Hessian (matrix) gives a *local form* for critical points, see the Morse Lemma in §1.7 of [GP74], and §I.2 of [Mil16], with a proof in proof Lemma 2.2. See also §7.6 of of [Mar74]. This is related to what we have seen about the Taylor series.

5.24. **Lagrange Multipliers.**

Theorem 5.35. *Given an open set $\mathcal{U} \subseteq \mathbb{R}^n$ and two $\mathcal{C}^1$ functions $F, G : \mathcal{U} \to \mathbb{R}$, let B be a level set for G, so $B = \{\mathbf{x} \in \mathcal{U} : G(\mathbf{x}) = c\}$. Assume that for some point $\mathbf{p} \in \mathcal{U}$, $\nabla G_\mathbf{p} \neq \mathbf{0}$. Then if F has a local maximum at $\mathbf{p} \in B$, there exists some $\lambda \in \mathbb{R}$ such that*

$$\nabla F_\mathbf{p} = \lambda \nabla G_\mathbf{p}.$$

Proof. Since $\nabla G_\mathbf{p} \neq \mathbf{0}$, the linear transformation (the matrix with those entries) $DG_\mathbf{p}$ is surjective, which allows us to use the Implicit Function Theorem, Theorem 5.27. So the level set B has a parametrization. Calling one of the coordinates t, we have a curve $\gamma(t)$ in B which passes through $\mathbf{p}$ at time 0, and such that given any chosen vector $\mathbf{v}$ tangent to B, $\gamma'(0) = \mathbf{v} \neq \mathbf{0}$.

Then $G(\gamma(t)) = c$ so for all t, $D(G \circ \gamma)(t) = 0$. By the Chain Rule this is

$$D(G \circ \gamma)(t) = DG_{\gamma(t)} D\gamma(t) = \nabla G_{\gamma(t)} \cdot \gamma'(t).$$

In particular fo $t = 0$, $\nabla G_\mathbf{p} \cdot \gamma'(0) = 0$.

On the other hand, F has a local maximum at $\mathbf{p}$, so in particular, $F \circ \gamma(t)$ has a local maximum at $t = 0$.

Therefore,

$$D(F \circ \gamma)(0) = DF_{\mathbf{p}} D\gamma(0) = \nabla F_{\mathbf{p}} \cdot \gamma'(0) = 0.$$

Since $\mathbf{v} = \gamma'(0) \neq \mathbf{0}$ was any tangent vector to B, both $\nabla F_{\mathbf{p}}$ and $\nabla G_{\mathbf{p}}$ are orthogonal to any such $\mathbf{v}$. Thus they must be multiples (think for example of a level curve, or a level surface).

□

Remark 5.13. Note that the derivatives are 0 for two completely different reasons: that $G \circ \gamma$ is constant, and that $F \circ \gamma$ has a maximum.

Note that it is possible for λ to be 0, and also possible for $\nabla F_{\mathbf{p}}$ to be $\mathbf{0}$. However for the proof $\nabla G_{\mathbf{p}}$ must be nonzero to be able to apply the Implicit Function Theorem.

6. Double and triple integrals (TO DO!)

6.1. **Review of Riemann integration.**

6.2. **Double integrals and Fubini's Theorem.**

6.3. **Equality of mixed partials.** As a nice application of Fubini's Theorem, we now prove Proposition stated above. We like the following "Fubini's Theorem argument", partly because it leads in to Green's Theorem of Calculus 3.

Lemma 6.1. *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 $R \subseteq \Omega$ we have

$$\int\int_R \varphi \,\mathrm{d}x\,\mathrm{d}y = \int\int_R \widetilde{\varphi}\,\mathrm{d}x\,\mathrm{d}y,$$

then we can conclude that $\varphi = \widetilde{\varphi}$ on Ω. (Because, if they differ at a point, then one is larger than the other on a small rectangle about that point, and the integrals there are different, a contradiction).

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 $R = [a, b] \times [c, d]$ we have the above equality of integrals, and the result will then follow.

Fubini's Theorem tells us that

$$\int\int_R \varphi(x, y)\,\mathrm{d}x\,\mathrm{d}y = \int_c^d \left(\int_a^b \varphi(x, y)\,\mathrm{d}x\right)\mathrm{d}y = \int_c^d \left(\int_a^b \frac{\partial}{\partial x}\left(\frac{\partial}{\partial y} f(x, y)\right)\mathrm{d}x\right)\mathrm{d}y$$

Now for any differentiable function $G(x, y)$ we have by the Fundamental Theorem of Calculus that for any fixed y, $\int_a^b \frac{\partial}{\partial x} G(x, y)\mathrm{d}x = G(a, y) - G(b, y)$, so

$$\int_a^b \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)(x,y)\mathrm{d}x = \frac{\partial f}{\partial y}(b,y) - \frac{\partial f}{\partial y}(a,y)$$

so the iterated integral equals

$$\int_c^d \frac{\partial}{\partial y} f(b,y)\mathrm{d}y - \int_c^d \frac{\partial}{\partial y} f(a,y)\mathrm{d}y =$$
$$\Big(f(b,d) - f(b,c)\Big) - \Big(f(a,d) - f(a,c)\Big).$$

Again, by Fubini's Theorem:

$$\int\int_R \widetilde{\varphi}(x,y)\ \mathrm{d}x\ \mathrm{d}y = \int_a^b \int_c^d \widetilde{\varphi}(x,y)\ \mathrm{d}y\ \mathrm{d}x = \int_a^b \left(\int_c^d \frac{\partial}{\partial y}\left(\frac{\partial \varphi}{\partial x} f(x,y)\right)\mathrm{d}y\right)\ \mathrm{d}x$$

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. □

6.4. **Change of variables and the determinant of the derivative matrix.**

References

[Axl97] Sheldon Jay Axler. *Linear Algebra Done Right*, volume 2. Springer, 1997.

[DC16] Manfredo P Do Carmo. *Differential geometry of curves and surfaces: revised and updated second edition*. Courier Dover Publications, 2016.

[GP74] Victor Guillemin and Alan Pollack. *Differential topology*. Prentice-Hall, 1974.

[Gui02] Hamilton Luiz Guidorizzi. Um Curso de Cálculo, Vols I- III, ltc, 5a, 2002.

[Hal74] P. R. Halmos. *Naive Set Theory*. Undergraduate Texts in Mathematics. Springer, 1974.

[HH15] John H Hubbard and Barbara Burke Hubbard. *Vector Calculus, Linear Algebra, and Differential Forms: a unified approach*. Matrix Editions, 2015.

[Mar74] Jerrold E. Marsden. *Elementary Classical Analysis*. W. H. Freeman, 1974.

[Mil16] John Milnor. Morse theory.(am-51), volume 51. In *Morse Theory.(AM-51), Volume 51*. Princeton university press, 2016.

[MW97] John Milnor and David W Weaver. *Topology from the differentiable viewpoint*, volume 21. Princeton university press, 1997.

[Str12] Gilbert Strang. *Linear algebra and its applications*. 2012.

[War71] Frank W Warner. *Foundations of differentiable manifolds and Lie groups*, volume 94 of *Graduate Texts in Mathematics*. Springer Verlag, 1971.

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