NYU-SH Honors Linear Algebra I – Lectures Summary

1 First class

Main reference: Axler §1.A or Treil §1.1 (the book titles are in the Syllabus)

Supplementary reading: Lay §4.1

- Usually "linearity" refers to operations involving the addition of objects of the same type and multiplication of these objects by numbers.
- *Linear Algebra* studies the mathematical structure of objects, sets and functions, as far as such structure is determined (or affected) by these operations.
- Vectors $\vec{x}$ on the plane are given by a pair of numbers $\vec{x} = (x_1, x_2) \in \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$.
- Vectors $\vec{x}$ on the 3-dimensional space are given by a triple $\vec{x} \in \mathbb{R}^3$.
- We can consider vectors on n-dimensional space as n-tuples $\vec{x} \in \mathbb{R}^n$.
- Adding two vectors $\vec{x}$ and $\vec{y}$ from $\mathbb{R}^n$, we get another vector $\vec{w} = \vec{x} + \vec{y} \in \mathbb{R}^n$.
- Multiplying a vector $\vec{x} \in \mathbb{R}^n$ by a number $\alpha \in \mathbb{R}$, we get a vector $\vec{w} = \alpha\vec{x} \in \mathbb{R}^n$.
- *Numbers* do not need to be real. We will consider both cases when the set $\mathbb{F}$ of numbers is given by $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. When $\mathbb{F} = \mathbb{C}$, we need the space to be $\mathbb{C}^n$ instead of $\mathbb{R}^n$, otherwise the previous property breaks down.
- A *complex number* $z \in \mathbb{C}$ is a number of the form $z = x + iy$ where $x, y \in \mathbb{R}$. In $\mathbb{C}$ we have usual algebraic properties of multiplication and addition, plus the property that $i^2 = -1$, so $(1 + 2i)(3 + 4i) = 3 + 4i + 6i + 8i^2 = -5 + 10i$.
- Why $\mathbb{C}$? Cutting a long story short:
  - Want to count: $\mathbb{N}$. Can add and multiply.
  - Want to subtract: $\mathbb{N} \rightsquigarrow \mathbb{Z}$
  - Want to divide: $\mathbb{Z} \rightsquigarrow \mathbb{Q}$
  - Want intermediate value theorem: $\mathbb{Q} \rightsquigarrow \mathbb{R}$
  - Want polynomials to have roots: $\mathbb{R} \rightsquigarrow \mathbb{C}$

  Even if one is ultimately interested in studying real quantities, using complex numbers may be more suitable because polynomials always have roots.

2 Vector spaces

Main reference: Treil §1.1 & Axler §1.A

Supplementary reading: Axler §1.B, Lay §4.1 and Hefferon §2.I.1 & §2.Fields

- A *field* $\mathbb{F}$ is a set with addition and multiplication operations satisfying: commutativity, associativity, additive identity 0, multiplicative identity 1, additive inverse $-\alpha$, multiplicative inverse $\frac{1}{\alpha}$, distributive property.
- Elements of $\mathbb{F}$ are called *numbers* or *scalars*. We will consider $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.
- A *vector space over the field* $\mathbb{F}$ is a set V together with the operations of addition and scalar multiplication (that is, for every $\vec{u}, \vec{v} \in V$ and $\alpha \in \mathbb{F}$, one has $\vec{u} + \vec{v} \in V$ and $\alpha\vec{u} \in V$) satisfying: commutativity, associativity, additive identity $\mathbf{0}$, additive inverse $-\vec{v}$, multiplicative identity, multiplicative associativity, distributive property for vector sum, distributive property for scalar sum.
- The additive identity $\mathbf{0}$ is unique, the additive inverse $-\vec{v}$ is unique for each $\vec{v}$.

  Proof. Expand $\mathbf{0} + \mathbf{0}'$ and $\vec{w} + \vec{v} + \vec{w}'$ using the above properties.
- Elements of a vector space are called *vectors* or *points*.

  A vector space over $\mathbb{R}$ is called a *real vector space*

  A vector space over $\mathbb{C}$ is called a *complex vector space*
- Examples of vector spaces: $\mathbb{F}^n$, the set $\mathcal{P}(\mathbb{F})$ of polynomials with real (or complex) coefficients, the set $\mathcal{P}_n(\mathbb{F})$ of polynomials of degree at most n.
- Another vector space is the set $\mathbb{F}^{m\times n}$ of $m \times n$ matrices $A = (a_{jk})_{j,k}$ written as

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}_{m\times n}$$

- The transpose of a matrix is defined by $A^T = (a_{kj})_{j,k} \in \mathbb{F}^{n\times m}$.

Notation. Treil denotes elements of $\mathbb{F}^n$ as column vectors, that is, matrices in $\mathbb{F}^{n\times 1}$:

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1} \qquad \text{or} \qquad \vec{x} = [x_1, x_2, \ldots, x_n]^T.$$

We will write $\vec{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{F}^n$, knowing that it denotes a column vector.

3 Linear combinations and bases

Main reference: Treil §1.2

Terminology. In these lecture notes, "proof" means just the main idea of the proof. The complete proof is the one written on the whiteboard or in the textbook.

- A *linear combination* of vectors $\vec{v}_1, \dots, \vec{v}_n$ is a sum of multiples of these vectors, resulting in some $\vec{u} = \alpha_1 \vec{v}_1 + \cdots + \alpha_n \vec{v}_n$ for some $\alpha_1, \dots, \alpha_n \in \mathbb{F}$.

- A family of vectors $\vec{v}_1, \dots, \vec{v}_n$ is *a basis of* V if every vector $\vec{u} \in V$ has a **unique** representation as a linear combination of $\vec{v}_1, \dots, \vec{v}_n$.

- Examples without proof: $(1,0), (0,1)$ is a basis of $\mathbb{R}^2$; $(1,1), (0,1)$ is a basis of $\mathbb{R}^2$; $(1,1), (2,2)$ is not a basis of $\mathbb{R}^2$; $(1,0), (0,1), (2,2)$ is not a basis of $\mathbb{R}^2$; $\vec{e}_1, \dots, \vec{e}_n$ is *the canonical basis* of $\mathbb{F}^n$; $1, t, t^2, t^3$ is a basis of $\mathcal{P}_3(\mathbb{F})$.

- Being a basis means that, for each $\vec{u} \in V$, the equation $\alpha_1 \vec{v}_1 + \cdots + \alpha_n \vec{v}_n = \vec{u}$ has a unique solution $(\alpha_1, \dots, \alpha_n)$. These numbers $\alpha_1, \dots, \alpha_n$ are called *the coordinates of* $\vec{u}$ *in the basis* $\vec{v}_1, \dots, \vec{v}_n$.

- A family of vectors $\vec{v}_1, \dots, \vec{v}_p$ is a *spanning family*, or *generating system*, or *complete system*, if every vector of V can be written as a linear combination of $\vec{v}_1, \dots, \vec{v}_p$.

  Examples without proof: $(1,0), (0,1)$ or $(1,1), (0,1)$ or $(1,1), (2,2)$ as well as $(1,0), (0,1), (2,2)$ are all spanning families of $\mathbb{R}^2$.

- The *trivial* linear combination of $\vec{v}_1, \dots, \vec{v}_n$ is the linear combination $0\vec{v}_1 + \cdots + 0\vec{v}_n$.

- A family of vectors is called *linearly independent* if the only linear combination equal to $\mathbf{0} \in V$ is the trivial linear combination. A family of vectors which is not linearly independent is called *linearly dependent*. $\emptyset$ is linearly independent.

- A family of vectors is a basis iff it is both spanning and linearly independent.

  Proof. For the more difficult direction, show that two linear combinations giving the same result must be the same by showing that the difference is trivial.

- ○ A family of vectors $\vec{v}_1, \dots, \vec{v}_n$ is linearly dependent iff there exists $k \in \{1, \dots, n\}$ and $\alpha_1, \dots, \alpha_{k-1} \in \mathbb{F}$ such that $\vec{v}_k = \mathbf{0} + \alpha_1 \vec{v}_1 + \cdots + \alpha_{k-1} \vec{v}_{k-1}$.

  Proof. Divide by the last non-zero coefficient in a non-trivial linear combination.

- Every finite spanning family contains a basis.

  Proof. Remove redundant vectors one by one until you get a basis.

4 Linear transformations and matrix-vector multiplication

Main reference: Treil §1.3 & §1.4

- A *linear map*, or *linear transformation*, is a function from a vector space V to a vector space W which satisfies the properties of additivity and homogeneity.
- Examples without proof: rotations on $\mathbb{R}^2$, reflections on $\mathbb{R}^2$, transposition of matrices, $T(x_1, x_2, x_3) = (x_1 - x_3, 4ix_2)$ from $\mathbb{C}^3$ to $\mathbb{C}^2$.
- Linear functions on $\mathbb{F}^1$: multiplication by a number. What about $\mathbb{F}^n$?
- For a linear map $T : \mathbb{F}^n \to \mathbb{F}^m$, define the vectors $\vec{a}_1 = T\vec{e}_1, \ldots, \vec{a}_n = T\vec{e}_n \in \mathbb{F}^m$. Then $\vec{a}_1, \ldots, \vec{a}_n$ determines T. Indeed, given $\vec{x} \in \mathbb{F}^n$, by linearity we have

  $$T\vec{x} = x_1\vec{a}_1 + \cdots + x_n\vec{a}_n = \sum_{k=1}^{n} x_k\vec{a}_k.$$

  Hence, the matrix

  $$A = \left[\, \vec{a}_1, \ldots, \vec{a}_n \,\right]_{m\times n}$$

  contains all the information about T. We denote this matrix A by $[T]$.
- *Multiplication of matrix by column.* Given $A \in \mathbb{F}^{m\times n}$ and $\vec{x} \in \mathbb{F}^n$, we define the product $\vec{y} = A\vec{x} \in \mathbb{F}^m$ by

  $$y_j = a_{j,1}x_1 + \cdots + a_{j,n}x_n = \sum_{k=1}^{n} a_{j,k}x_k.$$

  Writing $A = \left[\vec{a}_1, \ldots, \vec{a}_n\right]_{m\times n}$, this gives the same result as

  $$\vec{y} = x_1\vec{a}_1 + \cdots + x_n\vec{a}_n = \sum_{k=1}^{n} x_k\vec{a}_k.$$

  So with this definition we have $T\vec{x} = A\vec{x}$.
- To describe a linear transformation $T : \mathbb{F}^n \to \mathbb{F}^m$ we can consider any basis, it does not need to be $\vec{e}_1, \ldots, \vec{e}_n$. More generally, a linear transformation $T : V \to W$ is completely determined by the values that it takes on any given spanning family.
- Let $\mathcal{L}(V, W)$ denote the sets of all linear transformations defined on V and taking values on W. Then $\mathcal{L}(V, W)$ is itself a vector space!

  Proof. Exercise.

5 Composition and matrix multiplication

Main reference: Treil §1.5

- Suppose $A \in \mathbb{F}^{m\times n}$ and $B \in \mathbb{F}^{n\times r}$, and let $\vec{b}_1, \dots, \vec{b}_r \in \mathbb{F}^n$ be the columns of B. Then the product $AB \in \mathbb{F}^{m\times r}$ is the matrix whose columns are $A\vec{b}_1, \dots, A\vec{b}_r$.
- Writing $C = AB$, we have

$$c_{j,k} = (j\text{-th row of } A)(k\text{-th column of } B) = \sum_{l=1}^{n} a_{j,l} b_{l,k}.$$

- It is defined when the rows of A have the same length as the columns of B.
- For $T_1 \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ and $T_2 \in \mathcal{L}(\mathbb{F}^r, \mathbb{F}^n)$, then $[T_1 \circ T_2] = [T_1][T_2]$.

  Proof. The k-th column equals $T(\vec{e}_k) = T_1(T_2(\vec{e}_k)) = T_1(B\vec{e}_k) = T_1(\vec{b}_k) = A\vec{b}_k$
- Example: reflection against the line $x_1 = 3x_2$ on $\mathbb{R}^2$. Then $T = R_\gamma T_0 R_{-\gamma}$ is a composition of rotations and a reflection against the line $x_2 = 0$. After some work, we get $T(x_1, x_2) = (0.8x_1 + 0.6x_2, 0.6x_1 - 0.8x_2)$.
- Properties: associativity, distributivity, commutativity with scalars.
- No commutative property: in general $AB \neq BA$.

  Remark. If we pick square matrices "at random," chances are they don't commute.
- $(AB)^T = B^T A^T$ if one of the products is defined.
- Identity operator: $I_V \in \mathcal{L}(V) = \mathcal{L}(V, V)$ defined by $I_V \vec{v} = \vec{v}$.

  Identity matrix: $I = I_n \in \mathbb{F}^{n\times n}$ with 1 on diagonal and 0 elsewhere.

6 Invertible matrices and isomorphisms

Main reference: Treil §1.6

- We say that $T \in \mathcal{L}(V, W)$ is *left invertible* if there exists $S \in \mathcal{L}(W, V)$ such that $ST = I_V$. In this case S is called *a left inverse* of T.

  We say that $T \in \mathcal{L}(V, W)$ is *right invertible* if there exists $R \in \mathcal{L}(W, V)$ such that $TR = I_W$. In this case R is called *a right inverse* of T.

  Remark. The left and right inverses need not be unique. Matrix $\binom{1}{1}$ has many left inverses and no right inverse, $[1\ 1]$ has many right inverses and no left inverse.

- We say that T is *invertible* if it is both left invertible and right invertible. In this case, the left and right inverses are unique and are the same, denoted T^{-1}.

  Proof. Expand STR.

- Examples: Identity $I^{-1} = I$, rotation $(R_\gamma)^{-1} = R_{-\gamma}$.

- $T \in \mathcal{L}(V, W)$ is invertible iff for each $\vec{y} \in W$ the equation $T\vec{x} = \vec{y}$ has a unique solution $\vec{x} \in V$. So T is invertible as a linear map if it is bijective as a function.

  Proof. In one direction, apply T^{-1} to the equation to see that $\vec{x} = T^{-1}\vec{y}$ is the only solution. Conversely, let $f(\vec{y})$ denote the unique solution, so that $f \circ T = I_V$ and $T \circ f = I_W$, and check that f is linear.

- A matrix is *(left, right) invertible* if the corresponding linear transformation is (left, right) invertible, and A^{-1} is called *the inverse* of A.

- If A and B are invertible and AB is defined, then $(AB)^{-1} = B^{-1}A^{-1}$.
  If A is invertible, then $(A^T)^{-1} = (A^{-1})^T$ and $(A^{-1})^{-1} = A$.

  Proof. Check that the product from the left and the right give the identity.

- An invertible linear transformation $T \in \mathcal{L}(V, W)$ is called an *isomorphism*. If T is an isomorphism, then so is T^{-1}. Two vector spaces V and W are called *isomorphic*, denoted by $V \cong W$, if there exists an isomorphism between them.

  Remark. This means that these spaces have exactly the same properties, as far as their linear structure is concerned.

- Let $T \in \mathcal{L}(V, W)$ be an isomorphism. Then $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ is a basis for V iff $T\vec{v}_1, T\vec{v}_2, \ldots, T\vec{v}_n$ is a basis for W.

  Proof. Check that the properties of being LI and spanning are preserved by T.

- Let $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ be a basis for V. Then $T \in \mathcal{L}(V, W)$ is invertible iff $T\vec{v}_1, T\vec{v}_2, \ldots, T\vec{v}_n$ is a basis for W.

  Proof. Define $R \in \mathcal{L}(W, V)$ by $R\vec{w}_k = \vec{v}_k$. Check that $RT = I_V$ and $TR = I_W$.

- Corollary: A matrix is invertible iff its columns form a basis.

7 Row reduction and echelon forms

Main reference: Treil §2.1 & §2.2. Supplementary reading: Hefferon §1.I.1 & §1.I.2

- A system of linear equations, or *linear system* can be seen as:
  – A collection of m linear equations with n unknown variables.
  – A *matrix-vector equation* $Ax = b$.
  – A *vector equation* $x_1\vec{a}_1 + \cdots + x_n\vec{a}_n = b$.

  Here $A \in \mathbb{F}^{m\times n}$ is the *coefficient matrix* and $b \in \mathbb{F}^{m\times 1}$ is the *right-hand side.*

- Linear system is encoded by the *augmented matrix* $[A|b]$.

- There are three types of *row operations*:
  – Row exchange: interchange two rows
  – Scaling: multiply a row by a non-zero scalar
  – Row replacement: add a multiple of a row to another row

  These operations do not change the set of solutions, because they can be reversed.

- Row reduction:

  1. find the left most non-zero column;
  2. make sure its topmost entry is non-zero (apply row exchange if needed), this entry is then called a *pivot*; maybe apply scaling so that the pivot equals 1;
  3. apply row replacement to zero out all entries below the pivot;
  4. now leave this row alone, and apply the procedure to the remaining submatrix.

  Example:
  $$\left(\begin{array}{rrr|r} 0 & -4 & -8 & 4 \\ 1 & 2 & 3 & 1 \\ 2 & 1 & 2 & 1 \end{array}\right).$$

- Echelon form (triangular is a particular case):

  1. Non-zero rows are above zero rows, their first non-zero element is called *pivot*
  2. Position of each row's pivot is to the right of previous rows' pivots

  Reduced echelon form:

  3. The value of pivot entries is 1, entries above the pivots are also zero (below pivots are already zero by the two previous items)

  Examples:
  $$\left(\begin{array}{cccccc|c} \mathbf{1} & 0 & 8 & \mathbf{0} & \mathbf{0} & 0 & 9 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & 3 & 6 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & 1 & 3 \end{array}\right) \text{ and } \left(\begin{array}{cc|c} \mathbf{1} & 3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array}\right).$$

- Row reduction yields an echelon form. To get a reduced echelon form we apply the backward phase, from right to left. General solution may have *free variables.*

8 Echelon form and bases

Main reference: Treil §2.3

The notions of row operation, echelon form and pivot help us not only solve a given linear system, but this process actually reveals fundamental properties of bases, linearly independent families, spanning families, and invertible matrices.

Notation. Henceforth we write u instead of $\vec{u}$, but we still write $\vec{v}_j$ to avoid confusion.

- $Ax = b$ is inconsistent iff the echelon form of $[A|b]$ has a pivot in the last column.

  The echelon form of A, denoted A_e, has a pivot in every column if and only if, for every $b \in \mathbb{F}^m$, the equation $Ax = b$ is either inconsistent or has a unique solution.

  A_e has pivots in every row iff $Ax = b$ has solutions for every b.

  A_e has pivots in every row and column iff there is a unique solution for every b.

  Each row and column of an echelon form have at most one pivot.

  Proof. Immediate. Equivalent to not having free variables. Direct implication follows immediately from the first observation; conversely, if A_e does not have a pivot in every row, the last row is zero, taking $b_\mathrm{e} = (0, \ldots, 0, 1) \in \mathbb{F}^m$ makes $[A_\mathrm{e}|b_\mathrm{e}]$ inconsistent, and reversing the row operations give $[A|b]$ inconsistent. Follows immediately from previous two observations. Follows from definition of echelon.

- For a family $\vec{v}_1, \ldots, \vec{v}_m \in \mathbb{F}^n$, writing $A = [\vec{v}_1, \ldots, \vec{v}_m]_{n\times m}$:
  – The family is LI iff A_e has a pivot in every column.
  – The family is spanning iff A_e has a pivot in every row.
  – The family is a basis iff A_e has a pivot in every row and every column.

  Proof. The definitions of LI and spanning match the previous observations.

- A family with more than n vectors in $\mathbb{F}^n$ cannot be LI.

  Proof. Denote the family $\vec{v}_1, \ldots, \vec{v}_m$ with $m > n$ (if it is infinite, reduce it). There are at most n pivots in $[\vec{v}_1, \ldots, \vec{v}_m]_\mathrm{e}$, so there cannot be one at each column.

- Any two bases of V have the same number of elements.

  Proof. Can assume one of them, $\mathcal{A} = \vec{v}_1, \ldots, \vec{v}_n$ is finite. It is enough to show that the other one $\mathcal{B}$, cannot have more than n elements. Let $T \in \mathcal{L}(V, \mathbb{F}^n)$ be defined by $T\vec{v}_j = \vec{e}_j$. Then T is an isomorphism, hence $(T\vec{u})_{\vec{u}\in\mathcal{B}}$ is linearly independent. The claim then follows from the previous proposition.

- Every basis of $\mathbb{F}^n$ has n elements.

- A spanning family in $\mathbb{F}^n$ must have at least n elements.

  Proof. If it is infinite, it has a lot more. If it is finite, it contains a basis.

9 Echelon form and invertibility

Main reference: Treil §2.3 & §2.4

Notation. A " ∘ " indicates a point that it is not quite following the textbook.

- A matrix A is invertible iff A_{e} has a pivot in every row and every column.

  Proof. Both are equivalent to $Ax = b$ having unique solution for every $b \in \mathbb{F}^m$.
  Proof 2. Both are equivalent to $\vec{a}_1, \dots, \vec{a}_n$ being a basis.

- Only square matrices can be invertible.

  Proof. Let n be the number of pivots. Then A_{e} must have n rows and n columns.

- A square matrix is left invertible iff it is right invertible.

  Proof. If A is right invertible, $Ax = b$ has solution for every b, thus A_{e} has a pivot at every row, hence A_{e} has a pivot at every column and therefore A is invertible. If A is left invertible, $\mathbf{0}$ is the only solution to $Ax = \mathbf{0}$, thus A_{e} has a pivot at every column, hence A_{e} has a pivot at every row and therefore A is invertible.

- For square matrices, it is enough that $AB = I$ **or** $BA = I$ to have $B = A^{-1}$.

  Proof. It is a corollary of the previous proposition.

∘ A family $\vec{v}_1, \dots, \vec{v}_n \in \mathbb{F}^n$ is LI iff it is spanning.

  Proof. LI and spanning are equivalent to the matrix $[\vec{v}_1, \dots, \vec{v}_n]_{n \times n}$ having a row at every column or every row, which are in turn equivalent to each other.

∘ For a family with n vectors, it is enough to check LI **or** spanning to have a basis.

  Proof. It is a corollary of the previous proposition.

- Row operations on an $m \times n$ matrix A are equivalent to multiplying A from the left by an an *elementary matrix* E. Elementary matrices are invertible.

- To find the inverse of a square matrix A we can apply row reduction to $[A|I]$.

  If A_{e} has fewer than n pivots, we know that A is not invertible, and we can stop.

  If it has n pivots, the pivots are on the diagonal, and applying the backward phase of row reduction we get the reduced echelon form which is $[I|A^{-1}]$.

  Proof. Row reduction and backward phase consist in applying $B = E_k \cdots E_2 E_1$ to $[A|I]$, giving $B[A|I] = [BA|BI] = [I|B]$, and since $BA = I$ we have $B = A^{-1}$.

- Any invertible matrix can be represented as a product of elementary matrices.

10 Subspaces and dimension

Main reference: Treil §1.8 & §2.5

- A subset $W \subseteq V$ is called a *subspace* of V if W is itself a vector space, with the same operations as inherited from V.
- A subset $W \subseteq V$ is a subspace of V iff it satisfies:
  1. $\mathbf{0} \in W$.
  2. W is closed under addition, i.e., for every $u, v \in W$, we have $u + v \in W$.
  3. W is closed under scalar multiplication: $\alpha u \in W$ for every $u \in W$ and $\alpha \in \mathbb{F}$.

  Proof. All the properties are satisfied because W inherits the operations from V.

- ○ Examples: Trivial subspaces: $\{\mathbf{0}\}$ and V. The set of all linear combinations of a family $\mathcal{A} = \vec{u}_1, \dots, \vec{u}_k$, denoted $\mathsf{span}(\vec{u}_1, \dots, \vec{u}_k)$. The set of all solutions to $Ax = 0$. The *range* of $T \in \mathcal{L}(V, W)$, denoted $\mathsf{range}\, T = \{Tv : v \in V\} \subseteq W$. The *null space* or *kernel* of T, is given by $\ker T = \{v \in V : Tv = \mathbf{0}\} \subseteq V$.

  Useful properties: $\mathsf{span}(\mathsf{span}\, \mathcal{A}) = \mathsf{span}\, \mathcal{A}$, $\ker(TR) \supseteq \ker R$, $\mathsf{range}(TR) \subseteq \mathsf{range}\, T$.

- The *dimension* $\dim V$ of a vector space V is the number of vectors in a basis (note that $\dim\{\mathbf{0}\} = 0$ because $\emptyset$ is a basis). We say that V is *finite-dimensional* if it has a finite basis, otherwise it is *infinite-dimensional.*

  Examples: $\mathbb{F}^n$ and $\mathcal{P}_n(\mathbb{R})$ are finite-dimensional, $\mathcal{P}(\mathbb{R})$ and the space of all continuous functions defined on $[0, 1]$ are infinite-dimensional.

- ○ Suppose $n = \dim V < \infty$. A family $\mathcal{A}$ with n vectors is LI iff it is spanning. If it has fewer vectors, it cannot be spanning. If it has more vectors, it cannot be LI.

  For a family with n vectors, it is enough to check LI **or** spanning to have a basis.

  Proof. Take an isomorphism $T \in \mathcal{L}(V, \mathbb{F}^n)$ and use the result for $\mathbb{F}^n$.

- ○ Suppose $\dim V < \infty$. If $\mathcal{A} \subseteq \mathcal{C} \subseteq V$ and $\mathcal{A}$ is linearly independent, then there exists a finite basis $\mathcal{B}$ for $\mathsf{span}\, \mathcal{C}$ such that $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{C}$.

  Proof. Exercise.

- ○ Suppose $\dim V < \infty$. If $\mathcal{A}$ is a LI family, there is a basis $\mathcal{B}$ that contains $\mathcal{A}$.
  If $\mathcal{C}$ is a spanning family, there is a basis $\mathcal{B}$ contained in $\mathcal{C}$.

  Proof. Take $\mathcal{C} = V$. Take $\mathcal{A} = \emptyset$.

- Suppose $\dim V < \infty$. If W is a subspace of V, then $\dim W \leqslant \dim V$. Moreover, $\dim W = \dim V$ only if $W = V$.

  Proof. Take a basis for W, extend to a basis of V, if same number then $W = V$.

11 Fundamental subspaces and rank theorems

Main reference: Treil §2.6 & §2.7. Supplementary reading: Hefferon §2.III.3

Logic. Often in our sentences, we are implicitly saying that a certain statement is true *for all* V, *for all* v, etc. In order to negate such sentences, one needs to show that the claim is false *for some* v, etc. It is also implicit that V is a vector space. When we say that U and W are subspaces, it is implicit that they are subsets of *the same* space V. When we say "if vectors x and y ... then ...," usually it means in the same space.

- If $Ax = b$ has a solution v, then the set of solutions is given by $\{v + u : Au = \mathbf{0}\}$.

  Proof. If x in this set, $Ax = Av + Au = b + \mathbf{0} = b$, so x is a solution. Conversely, if $Ax = b$, take $u = x - v$, so $Au = Ax - Av = b - b = \mathbf{0}$, and x is in this set.

Suppose we are given a parametrized family $\mathcal{A}$ of solutions as one fixed vector plus the span of a few other vectors. How can we tell whether $\mathcal{A}$ contains **all** solutions to $Ax = b$?

- We associate to a given matrix $A \in \mathbb{F}^{m\times n}$ four *fundamental subspaces*:
  - *Null space* or *kernel*: $\operatorname{ker} A = \{v \in \mathbb{F}^n : Av = \mathbf{0}\} \subseteq \mathbb{F}^n$.
  - *Column space* or *range*: $\operatorname{range} A = \operatorname{span}(\vec{a}_1, \ldots, \vec{a}_n) = \{Ax : x \in \mathbb{F}^n\} \subseteq \mathbb{F}^m$.
  - *Row space*, given by $\operatorname{range}(A^T) \subseteq \mathbb{F}^n$.
  - *Left null space*, given by $\operatorname{ker}(A^T) \subseteq \mathbb{F}^m$.
- How to find bases the range, row space and kernel?

  First, use row reduction to find an echelon form A_{e}.
  We say that column k is a *pivot column* if it contains a pivot of A_{e}.

  1. The pivot columns of *the original matrix* A form a basis for $\operatorname{range} A$.
  2. The non-zero rows of A_{e} form a basis for $\operatorname{range} A^T$.
  3. Expressing solutions of $A_{\mathrm{re}}\, x = \mathbf{0}$ in vector form gives a basis for $\operatorname{ker} A$, each vector in the basis corresponding to one free variable.
- We define *the rank of* A as $\operatorname{rank} A = \dim \operatorname{range} A$.
- **Rank Theorem:** For $A \in \mathbb{F}^{m\times n}$, $\operatorname{rank} A = \operatorname{rank} A^T$.

  Proof. From previous procedures, both correspond to the number of pivots in A_{e}.
- **Rank-Nullity Theorem:** For $A \in \mathbb{F}^{m\times n}$, $\operatorname{rank} A + \dim \operatorname{ker} A = n$.
  If $\dim V < \infty$ and $T \in \mathcal{L}(V, W)$, then $\dim \operatorname{range} T + \dim \operatorname{ker} T = \dim V$.

  Proof. From previous procedures, $\operatorname{rank} A$ equals the number of pivots in A_{e} and $\dim \operatorname{ker} A$ equals the number of columns without pivots. These add up to n. For a linear map $T \in \mathcal{L}(V, W)$, consider isomorphisms to subspaces of $\mathbb{F}^n$.

12 Finding bases and completing bases

Main reference: Treil §2.7

- How to find bases the range, row space and kernel?
  1. The pivot columns of A (those where A_e has a pivot) form a basis for $\mathsf{range}\, A$.
  2. The non-zero rows of A_e form a basis for $\mathsf{range}\, A^T$.
  3. Expressing the solutions of $A_\mathrm{re}\, x = \mathbf{0}$ in vector form gives a basis for $\ker A$, each vector in the basis corresponding to one free variable.

  Proof. We need a few preliminary lemmas.

  Exercise: $\ker A$ determines which columns are spanned by which other columns.

  Exercise: If S is invertible, then $\ker(ST) = \ker T$ and $\mathsf{range}(RS) = \mathsf{range}\, R$.

  1. Pivot columns of A_re are LI and span the other columns. Since $A_\mathrm{re} = EA$ with E invertible, after applying E^{-1} the corresponding columns are still LI and still span the other columns, hence they are a basis for the column space.
  2. First, check that non-zero rows of A_e are linearly independent, so they form a basis for $\mathsf{range}\, A_\mathrm{e}^T$. Second, note that $A_\mathrm{e}^T = A^T E^T$ with E is invertible, and by the second exercise $\mathsf{range}\, A_\mathrm{e}^T = \mathsf{range}\, A^T$.
  3. These vectors span the null space $\ker A$ by construction. Since the k-th coordinate of the general solution always equals the free variable x_k, the only linear combination that produces $\mathbf{0}$ is the trivial one, so they are also LI.

- For two subspaces U and W of V, the *sum of U and W* is the subspace

$$U + W = \{u + w : \ u \in U, \ w \in W\} \subseteq V.$$

- $\mathsf{span}(\vec{x}_1, \ldots, \vec{x}_j) + \mathsf{span}(\vec{y}_1, \ldots, \vec{y}_r) = \mathsf{span}(\vec{x}_1, \ldots, \vec{x}_j, \vec{y}_1, \ldots, \vec{y}_r)$.

  Proof. Exercise.

- How can we complete a LI family in $\mathbb{F}^n$ to get a basis? Write them as rows, find the pivot columns, and add canonical rows $\vec{e}_k$ corresponding to the free variables.

  Proof. Let A be the matrix $[\vec{v}_1, \ldots, \vec{v}_j]^T$ and $A_\mathrm{e} = [\vec{u}_1, \ldots, \vec{u}_j]^T$ be its echelon form. Let B be the square matrix $[\vec{u}_1, \ldots, \vec{u}_j, \vec{e}_{k_1}, \ldots, \vec{e}_{k_r}]^T$. With only row exchanges we get B_e with n pivots, so $\mathsf{rank}\, B = n$ and thus $\mathsf{range}\, B^T = \mathbb{F}^n$. On the other hand,

$$\begin{aligned} \mathsf{range}\, B^T &= \mathsf{span}(\vec{u}_1, \ldots, \vec{u}_j, \vec{e}_{k_1}, \ldots, \vec{e}_{k_r}) \\ &= \mathsf{range}\, {A_\mathrm{e}}^T + \mathsf{span}(\vec{e}_{k_1}, \ldots, \vec{e}_{k_r}) \\ &= \mathsf{range}\, A^T + \mathsf{span}(\vec{e}_{k_1}, \ldots, \vec{e}_{k_r}) \\ &= \mathsf{span}(\vec{v}_1, \ldots, \vec{v}_j, \vec{e}_{k_1}, \ldots, \vec{e}_{k_r}). \end{aligned}$$

  Since this family is spanning and contains $j + r = n$ vectors, it is a basis.

13 Coordinate and change of basis

Main reference: Treil §2.8

- Let $\mathcal{A} = \vec{a}_1, \dots, \vec{a}_n$ be a basis of a vector space V. For a vector $v \in V$ such that $v = x_1\vec{a}_1 + \cdots + x_n\vec{a}_n$, the *coordinate vector* of v in the basis $\mathcal{A}$ is defined as
$$[v]_{\mathcal{A}} = (x_1, \dots, x_n) \in \mathbb{F}^n$$
and the numbers $x_1, \dots, x_n$ are *the coordinates of v relative to the basis $\mathcal{A}$.*
- The map $v \mapsto [v]_{\mathcal{A}}$ is an isomorphism between V and $\mathbb{F}^n$.
- For a linear map $T \in \mathcal{L}(V, W)$ and bases $\mathcal{A} = \vec{a}_1, \dots, \vec{a}_n$ of V and $\mathcal{B} = \vec{b}_1, \dots, \vec{b}_m$ of W, the matrix of T with input basis $\mathcal{A}$ and output basis $\mathcal{B}$, denoted $[T]_{\mathcal{BA}} \in \mathbb{F}^{m\times n}$ is the matrix whose k-th column is $[T\vec{a}_k]_{\mathcal{B}}$. With this definition,
$$[Tv]_{\mathcal{B}} = [T]_{\mathcal{BA}}\,[v]_{\mathcal{A}}$$
for every $v \in V$, and $[T]_{\mathcal{BA}}$ is the only matrix with this property.
- ◦ A basis is a basis regardless of how vectors are ordered.
  But, for the purpose of writing $[v]_{\mathcal{B}}$ and $[T]_{\mathcal{BA}}$, the order does matter.
- If $S \in \mathcal{L}(U, V)$ and $\mathcal{C}$ is a basis of U, then
$$[TS]_{\mathcal{BC}} = [T]_{\mathcal{BA}}[S]_{\mathcal{AC}}.$$
Proof. $[(TS)u]_{\mathcal{B}} = [T(Su)]_{\mathcal{B}} = [T]_{\mathcal{BA}}\,[Su]_{\mathcal{A}} = [T]_{\mathcal{BA}}\,[S]_{\mathcal{AC}}\,[u]_{\mathcal{C}}$.
- The change of coordinate matrix from a basis $\mathcal{A} = \vec{a}_1, \dots, \vec{a}_n$ of V to another basis $\mathcal{B} = \vec{b}_1, \dots, \vec{b}_n$ of V is the matrix of I_V with input basis $\mathcal{A}$ and output basis $\mathcal{B}$:
$$[v]_{\mathcal{B}} = [I_V]_{\mathcal{BA}}[v]_{\mathcal{A}}.$$
Moreover, the change of basis from $\mathcal{B}$ to $\mathcal{A}$ is the matrix $[I_V]_{\mathcal{AB}} = ([I_V]_{\mathcal{BA}})^{-1}$.
- If $\mathcal{S} = \vec{e}_1, \dots, \vec{e}_n$ denote the canonical basis of $\mathbb{F}^n$, and let $\mathcal{A} = \vec{a}_1, \dots, \vec{a}_n$ denote another basis. Then $[I_V]_{\mathcal{SA}} = A = [\,\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n]_{n\times n}$ and $[I_V]_{\mathcal{AS}} = A^{-1}$.

  Examples: $\mathcal{A} = (1, 2), (2, 1)$. $\mathcal{A} = 1, 1 + t$ and $\mathcal{B} = 1 + 2t, 1 - 2t$.
- The change of basis for the matrix of a linear map $T \in \mathcal{L}(V, W)$, with $\mathcal{A}$, $\mathcal{A}'$ bases of V and $\mathcal{B}$, $\mathcal{B}'$ bases of W, is given by:
$$[T]_{\mathcal{B}'\mathcal{A}'} = [I_W]_{\mathcal{B}'\mathcal{B}}[T]_{\mathcal{BA}}[I_V]_{\mathcal{AA}'}.$$
In case $T \in \mathcal{L}(V)$, we have
$$[T]_{\mathcal{B}} = [I_V]_{\mathcal{BA}}[T]_{\mathcal{A}}[I_V]_{\mathcal{AB}}.$$
- Two matrices A and $B \in \mathbb{F}^{n\times n}$ are *similar* is there exists an invertible matrix $Q \in \mathbb{F}^{n\times n}$ such that $A = Q^{-1}BQ$. This splits $\mathbb{F}^{n\times n}$ into *classes*.

14 Determinant: axiomatic definition

Main reference: Treil §§3.1–3.3, **with row instead of column!**

- We want to define the *determinant* of a **square** matrix as a quantity, function of its **rows** $\vec{a}_j$, which in some sense measures the "volume" induced by vectors $\vec{a}_j$, and which is meaningful for Linear Algebra. This function should satisfy:

  (0) – Invariance under row replacement
  (1) – Linearity in each row
  (3) – Normalization

- Assuming (1), Property (0) is equivalent to the following:

  (2) – Antisymmetry under row exchange

  Proof. For (0) $\Rightarrow$ (2), add $\vec{a}_j$ to $\vec{a}_k$, then $-\vec{a}_k$ to $\vec{a}_j$, then $\vec{a}_j$ to $\vec{a}_k$, and use (1). For (2) $\Rightarrow$ (0), suppose C is obtained by taking $\vec{c}_j = \vec{a}_j + \alpha\vec{a}_k$. Using (1), $\det C = \det A + \alpha \det B$, where rows j and k of B are identical. Using (2), $\det B = 0$.

- We say that $\det : \mathbb{F}^n \to \mathbb{F}$ is *a determinant* if it satisfies Properties (1)-(2)-(3).

  For now, let us assume existence of such a function. We will see that, using only these properties, we can compute $\det A$. So we can call it *the determinant.*

- How do row operations affect $\det$? From Properties (0)-(1)-(2),

  – Row replacement: does not change $\det$.
  – Scaling: multiply $\det$ by α.
  – Row exchange: multiply $\det$ by -1.

- A matrix $B \in \mathbb{F}^{n\times n}$ is *upper triangular* if all entries below the main diagonal are zero. If B is upper triangular, we have $\det B = b_{1,1}b_{1,2}\cdots b_{n,n}$.

  Proof. If B has zero on the diagonal, then B_{e} has a zero row and $\det B = 0$ by (1). If not, then row replacements make B diagonal, and scaling makes it identity.

- Row reduction consists of row operations which yield an upper triangular matrix. So we can indeed compute $\det A$ assuming only (1)-(2)-(3)!

- $\det A = 0$ iff A is not invertible.

  Proof. Row operations do not change whether or not a matrix's determinant is zero. If A is invertible, row operations yield the identity. If A is not invertible, row operations yield a zero row.

- $\det A = 0$ iff one of the rows is a linear combination of the others.

  Proof. Equivalent to A is not being invertible.

- By linearity in each row, $\det(\alpha A) = \alpha^n \det A$.

- We still haven't proved existence of the determinant.

15 Determinant: factorization and permutation formula

Main reference: Treil §3.3 with row instead of column, and §3.4

- $\mathsf{det}(AB) = (\mathsf{det}\, A)(\mathsf{det}\, B)$ and $\mathsf{det}(A^T) = \mathsf{det}\, A$.

  Proof. Lemma: If E is an elementary matrix, then $\mathsf{det}(EB) = (\mathsf{det}\, E)(\mathsf{det}\, B)$. Indeed, performing row operations is equivalent to multiplying from the left by elementary matrices, whose determinant coincides with the factor affecting the determinant of B. To prove the above identities, we can assume A is invertible (otherwise AB and A^T are not invertible, and we get $0 = 0$), so $A = E_N \cdots E_2 E_1$. By the lemma, $\mathsf{det}(AB) = (\mathsf{det}\, E_N) \cdots (\mathsf{det}\, E_2)(\mathsf{det}\, E_1)(\mathsf{det}\, B) = (\mathsf{det}\, A)(\mathsf{det}\, B)$. Moreover, $A^T = E_1^T E_2^T \cdots E_N^T$, so it is enough to prove the second identity for elementary matrices, i.e., $\mathsf{det}(E^T) = \mathsf{det}\, E$, which can be checked case by case.

- The determinant of $A = (a_{j,k})_{j,k} \in \mathbb{F}^{n \times n}$ exists and is given by

  $$\mathsf{det}\, A = \sum_{\sigma} a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n}\, \mathsf{sgn}(\sigma).$$

  The above sum is over all *permutations* σ of $\{1, 2, \ldots, n\}$. Finally, $\mathsf{sgn}\, \sigma$ is defined as ± 1 according to the parity of how many *disorders* are present in σ, i.e.

  $$\mathsf{sgn}(\sigma) = (-1)^{\#\{(j,k):\ 1 \leqslant j < k \leqslant n,\ \sigma(j) > \sigma(k)\}}.$$

  Derivation. First, if A has exactly one 1 in each column, one 1 in each row, and 0 elsewhere, then A is a *permutation of the identity*, i.e., $A = [\vec{e}_{\sigma(1)}, \ldots, \vec{e}_{\sigma(n)}]$ for some permutation σ. In this case, the product $a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n}$ equals 1 for this permutation σ and 0 for all others, and the above formula states that $\mathsf{det}\, A = \mathsf{sgn}\, \sigma$. This is consistent with properties of det, as can be seen by applying neighbor column permutations to I_n while using Property (2'), and using Property (3) for I_n itself. Now consider the general case, $A \in \mathbb{F}^{n \times n}$. Write $A = [\vec{a}_1, \ldots, \vec{a}_n]$, so $\vec{a}_k = [a_{1,k}, \ldots, a_{n,k}]^T = \sum_j a_{j,k} \vec{e}_j$. Using Property (1') of det for $\vec{a}_1$,

  $$\mathsf{det}\, A = \sum_{j_1} a_{j_1,1}\, \mathsf{det}[\vec{e}_{j_1}, \vec{a}_2, \ldots, \vec{a}_n].$$

  Repeating the same argument for $\vec{a}_2, \ldots, \vec{a}_n$,

  $$\mathsf{det}\, A = \sum_{j_1} \sum_{j_2} \cdots \sum_{j_n} a_{j_1,1} a_{j_2,2} \cdots a_{j_n,n}\, \mathsf{det}[\vec{e}_{j_1}, \vec{e}_{j_2}, \ldots, \vec{e}_{j_n}].$$

  The above sum has n^n terms, but most of them are zero for having repeated columns. The nonzero terms are exactly when the j_k's are all different, i.e., when for some permutation σ, $j_k = \sigma(k)$ for all k. For this term, $\mathsf{det}(\vec{e}_{\sigma(1)}, \ldots, \vec{e}_{\sigma(k)}) = \mathsf{sgn}(\sigma)$. So, a function satisfying (1)-(2)-(3) must agree with the above formula.

  Proof. (1) the above summand has exactly one term from each column. (2) column exchange results from an odd number of neighbor column permutations. (3) when $A = I$, only the neutral permutation gives a non-zero summand.

16 Determinant: volume and cofactor expansion

Main reference: Treil §3.5

- ∘ Given $T \in \mathcal{L}(\mathbb{R}^n)$, for $\Omega \subset \mathbb{R}^n$ open and bounded, $\mathsf{vol}(T(\Omega)) = |\det T| \times \mathsf{vol}(\Omega)$.

  Proof. Seen in HLA-2, using Isometries and Singular Value Decomposition.

- *Cofactor expansion.* For $A = (a_{jk})_{j,k} \in \mathbb{F}^{n\times n}$ and for $j, k \in \{1, \dots, n\}$, let $A_{j,k} \in \mathbb{F}^{(n-1)\times(n-1)}$ be the submatrix obtained by erasing row j and column k from A. We can expand the determinant of A with respect to any given row j:

$$\det A = \sum_{k=1}^{n} (-1)^{j+k} a_{j,k} \det A_{j,k}.$$

  We can also expand the determinant of A with respect to any given column j:

$$\det A = \sum_{k=1}^{n} (-1)^{j+k} a_{k,j} \det A_{k,j}.$$

  Remark. This method has theoretical importance, and can be helpful when computing examples of size 2 and 3, or for a matrix with many zeros.

  Explanation. The expansion for column j can be seen as splitting the permutation formula according to the value of $\sigma(j)$. This gives invertible functions from $\{1, \dots, n\}\backslash\{j\}$ to $\{1, \dots, n\}\backslash\{k\}$, which in turn can be identified with permutations of $\{1, \dots, n-1\}$ with the sgn changed accordingly.

- The coefficients $c_{j,k} = (-1)^{j+k} \det A_{j,k}$ are called *cofactors* of A.

- ∘ Writing $C = [\vec{c}_1, \dots, \vec{c}_n]$ and $A = [\vec{a}_1, \dots, \vec{a}_n]$, we have $\vec{c}_j \cdot \vec{a}_j = \sum_k c_{k,j} a_{k,j} = \det A$. In general, for any vector $\vec{y}$ we have $\vec{c}_j \cdot \vec{y} = \det[\vec{a}_1, \dots, \vec{a}_{j-1}, \vec{y}, \vec{a}_{j+1}, \dots, \vec{a}_n]$.

- Let $A \in \mathbb{F}^{n\times n}$ be invertible and have cofactor matrix C. Then

$$A^{-1} = \frac{1}{\det A} C^T.$$

  Remark. Same as before.

  Proof. Let $D = C^T A$. Then $d_{j,k} = \vec{c}_j \cdot \vec{a}_k = \det[\vec{a}_1, \dots, \vec{a}_{j-1}, \vec{a}_k, \vec{a}_{j+1}, \dots, \vec{a}_n]$. When $k = j$, this gives $\det A$. When $k \neq j$, this is the determinant of a matrix with two repeated columns, which is zero. So $C^T A = (\det A) I$.

- Cramer's rule: If A is invertible, then the solution to $Ax = b$ is given by

$$x_k = \frac{\det B_k}{\det A},$$

  where B_k is the matrix obtained when we replace the k-th column of A by b.

  Proof. Writing $x = A^{-1}b = \frac{1}{\det A} C^T b = \frac{1}{\det A}\vec{v}$, we have $v_k = \vec{c}_k \cdot b = \det B_k$.

17 Eigenvalues and eigenvectors

Main reference: Treil §4.1. Review on polynomials: Axler §4

- Let $A \in \mathbb{F}^{n\times n}$. A scalar $\lambda \in \mathbb{F}$ is an *eigenvalue* if there exists $v \in \mathbb{F}^n \setminus \{\mathbf{0}\}$ such that $Av = \lambda v$, and v is called an *eigenvector* of A associated with eigenvalue λ.
- The subspace $\mathsf{ker}(A - \lambda I)$ is called the *eigenspace* associated to λ.
  The set of all the eigenvalues of A is called the *spectrum* of A.
- $p_A(\lambda) = \mathsf{det}(A - \lambda I)$ has degree n and is called the *characteristic polynomial* of A.

  A number $\lambda \in \mathbb{F}$ is an eigenvalue of A iff $p_A(\lambda) = 0$.

  Remark. Not very practical unless n is small or A has many zeros.
- Similar matrices have the same characteristic polynomial.

  Proof. Exercise.
- The *algebraic multiplicity* of an eigenvalue is its multiplicity as a root of $p_A \in \mathcal{P}(\mathbb{C})$.
- The n eigenvalues of an upper triangular matrix $A \in \mathbb{C}^{n\times n}$, listed with algebraic multiplicity, are exactly the diagonal entries $a_{1,1}, a_{2,2}, \ldots, a_{n,n}$.

  Proof. Exercise.

Below we consider $\mathbb{F} = \mathbb{C}$. In some situations it may be useful to treat real numbers, vectors and matrices as particular cases of complex numbers, vectors and matrices.

- *Fundamental Theorem of Algebra.* Non-constant complex polynomials have roots.

  We call z_0 a *root* of p if $p(z_0) = 0$. The *multiplicity* of a root z_0 is the highest power of $(z - z_0)$ that divides $p(z)$. Any complex polynomial of degree n can written as $p(z) = c(z - z_1)\cdots(z - z_n)$, where $z_1, \ldots, z_n$ are its roots, counting multiplicity.
- Every $A \in \mathbb{C}^{n\times n}$ has n eigenvalues $\lambda_1, \ldots, \lambda_n$, counted with algebraic multiplicity.

  Proof. Follows from the factorization of complex polynomials of degree n.
- In this case, $\mathsf{det}\, A = \lambda_1 \cdots \lambda_n$ and $\mathsf{trace}\, A = \lambda_1 + \cdots + \lambda_n$.

  Proof. Let us analyze the coefficients of $p_A(z) = b_n z^n + b_{n-1} z^{n-1} + \cdots + b_1 z + b_0$. We first expand the product $p_A(z) = c(z - \lambda_1)\cdots(z - \lambda_n)$, which gives $b_n = c$, $b_{n-1} = -c(\lambda_1 + \cdots + \lambda_n)$, and $b_0 = c\,(-\lambda_1)\cdots(-\lambda_n)$.
  Expanding the permutation formula for $\mathsf{det}(A - zI)$, only the diagonal permutation has terms involving z^n or z^{n-1}. Other permutations miss at least two positions in the diagonal. So b_{n-1} and b_n come from $(a_{1,1} - z)\cdots(a_{n,n} - z)$, giving $b_n = (-1)^n$ and $b_{n-1} = (-1)^{n-1}(a_{1,1} + a_{2,2} + \cdots + a_{n,n})$. Moreover, $b_0 = p_A(0) = \mathsf{det}\, A$.

18 Diagonalization

Main reference: Treil §4.2, skipping 4.2.4. We do not always treat $\mathbb{R}$ as a subset of $\mathbb{C}$.

- ∘ For $T \in \mathcal{L}(\mathbb{F}^n)$, it would be very convenient to have a basis $\mathcal{B}$ for which $[T]_\mathcal{B}$ is a diagonal matrix. Denoting $[T]_\mathcal{S} = A$, this means $Q^{-1}AQ = D$, or $A = QDQ^{-1}$.

  Remark. In this case, $A^N = QD^NQ^{-1}$, $p(A) = Qp(D)Q^{-1}$, $e^{tA} = Qe^{tD}Q^{-1}$, etc.

- ∘ We say that $A \in \mathbb{F}^{n\times n}$ is *diagonalizable (over* $\mathbb{F}$*)* if $A = QDQ^{-1}$ for some $Q \in \mathbb{F}^{n\times n}$ invertible and D diagonal. In this case we say that Q *diagonalizes* A.

- ∘ Let $A \in \mathbb{F}^{n\times n}$, $B = [\vec{v}_1, \dots, \vec{v}_r] \in \mathbb{F}^{n\times r}$, and $D = \mathsf{diag}(\lambda_1, \dots, \lambda_r) \in \mathbb{F}^{r\times r}$. Then $AB = BD$ if and only if $A\vec{v}_j = \lambda_j\vec{v}_j$ for $j = 1, \dots, r$.

  Proof. Check what the columns of AB and BD are.

- • A matrix $A \in \mathbb{F}^{n\times n}$ is diagonalizable iff there is a basis of $\mathbb{F}^n$ made of eigenvectors.

  Proof. Write $Q = [\vec{v}_1, \dots, \vec{v}_n]_{n\times n}$.

- • If $\vec{v}_1, \dots, \vec{v}_r \in \mathbb{F}^n$ are eigenvectors of $A \in \mathbb{F}^{n\times n}$ corresponding to distinct eigenvalues, then $\vec{v}_1, \dots, \vec{v}_r$ are linearly independent.

  Proof. Apply $A - \lambda_r I$ to a null linear combination and use induction on r.

- • If $A \in \mathbb{F}^{n\times n}$ has n distinct eigenvalues $\lambda_1, \dots, \lambda_n \in \mathbb{F}$, then A is diagonalizable.

  Proof. Take n corresponding eigenvectors, they are LI, so they form a basis.

- • The *geometric multiplicity* of an eigenvalue λ is given by $\mathsf{dim\,ker}(A - \lambda I)$.

- • The geometric multiplicity of λ cannot exceed its algebraic multiplicity.

  Proof. Exercise.

- ∘ Let $\lambda_1, \dots, \lambda_r \in \mathbb{F}$ denote the distinct eigenvalues of $A \in \mathbb{F}^{n\times n}$. Then A is diagonalizable over $\mathbb{F}$ if and only if the sum of algebraic multiplicities $m_1, \dots, m_r$ equals n and they equal the geometric multiplicities $g_1, \dots, g_r$.

  Proof. ($\Rightarrow$) A LI family of eigenvectors has at most $\sum_j g_j \leqslant \sum_j m_j \leqslant n$ vectors, and a basis has n vectors. ($\Leftarrow$) For each $j = 1, \dots, r$, take $\vec{v}_{j,1}, \vec{v}_{j,2}, \dots, \vec{v}_{j,m_j}$ as a basis for $\mathsf{ker}(A - \lambda_j I)$. Let $\mathcal{B} = \vec{v}_{1,1}, \vec{v}_{1,2}, \dots, \vec{v}_{1,m_1}, \dots, \vec{v}_{r,1}, \vec{v}_{r,2}, \dots, \vec{v}_{r,m_r}$. Note that each vector in $\mathcal{B}$ is an eigenvector of A, since $A\vec{v}_{j,k} = \lambda_j\vec{v}_{j,k}$. Suppose $\sum_{j=1}^{r}\sum_{k=1}^{m_j} \alpha_{j,k}\vec{v}_{j,k} = \mathbf{0}$ for some collection of scalars $(\alpha_{j,k})_{(j,k)}$. Take $\vec{u}_j = \sum_{k=1}^{m_j} \alpha_{j,k}\vec{v}_{j,k}$. Then $\vec{u}_j$ is either $\mathbf{0}$ or a λ_j-eigenvector. Sine $\sum_j \vec{u}_j = \mathbf{0}$ and eigenvectors with distinct eigenvalues are LI, we have $\vec{u}_j = \mathbf{0}$ for every j. Since $\vec{v}_{j,1}, \vec{v}_{j,2}, \dots, \vec{v}_{j,m_j}$ are LI, we have $\alpha_{j,k} = 0$ for $k = 1, \dots, m_j$. Therefore, $\mathcal{B}$ is LI.

- ∘ A square matrix with real entries is diagonalizable over $\mathbb{R}$ if and only if it is diagonalizable over $\mathbb{C}$ and all the eigenvalues are real.

  Proof. The geometric multiplicity of a real eigenvalue is the same over $\mathbb{R}$ or $\mathbb{C}$.

19 Orthogonality and projection

Main reference: Lay §§6.1–6.3. Most proofs of this topic will be skipped.

Notation. Henceforth we write u_j instead of $\vec{u}_j$ and no longer refer to coordinates.

- For vectors $u, v \in \mathbb{R}^n$ we define the *dot product* by $u \cdot v = u^T v$.
- For $u, v, w \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, we have :
  $u \cdot v = v \cdot u$, $(u + v) \cdot w = u \cdot w + v \cdot w$, $(\alpha u) \cdot v = \alpha(u \cdot v)$, and $u \cdot u > 0$ if $u \neq \mathbf{0}$.
- The *length* (or *norm*) of v is given by $\|v\| = \sqrt{v \cdot v} \geqslant 0$.

  We call v a *unit vector* if $\|v\| = 1$.
- We say that u is *orthogonal* to v, denoted $u \perp v$, if $u \cdot v = 0$.
- Two vectors u and v are orthogonal if and only if $\|u + v\|^2 = \|u\|^2 + \|v\|^2$.
- We say that u is orthogonal to $\mathcal{B}$ if $u \perp w$ for all $w \in \mathcal{B}$. The *orthogonal complement* of $\mathcal{B}$ and is denoted by $\mathcal{B}^\perp = \{u \in \mathbb{R}^n : u \text{ is orthogonal to } \mathcal{B}\}$.
- Let $\mathcal{B}$ be a family of vectors and $W = \mathsf{span}(\mathcal{B})$. Then $W^\perp = \mathcal{B}^\perp$.
- A family $\mathcal{B}$ of vectors is called an *orthogonal family* if $u \perp v$ for all $u \neq v$ in $\mathcal{B}$.
  If moreover all the vectors in $\mathcal{B}$ are unit vectors, we call $\mathcal{B}$ an *orthonormal family.*
- The columns of $Q \in \mathbb{R}^{n\times k}$ are orthonormal iff $Q^T Q = I_{k\times k}$. In this case, $(Qu) \cdot (Qv) = u \cdot v$ for all $u, v \in \mathbb{R}^k$. In particular, $\|Qv\| = \|v\|$ for all $v \in \mathbb{R}^k$.
- Any orthogonal family $\{u_1, \ldots, u_r\}$ of nonzero vectors is linearly independent. Moreover, if $y \in \mathsf{span}(u_1, \ldots, u_r)$, then $y = \sum_j \frac{y \cdot u_j}{u_j \cdot u_j} u_j$.

  Proof. Write $y = \sum_j \alpha_j u_j$ and compute $y \cdot u_k$ to determine α_k.
- *Orthogonal decomposition and best approximation.* Let $U \subseteq \mathbb{R}^n$ be a subspace and $u_1, \ldots, u_r$ an orthogonal basis for U. For each $y \in \mathbb{R}^n$ there are unique $\hat{y} \in U$ and $w \in U^\perp$ such that $y = \hat{y} + w$. The vector $\hat{y}$ is called the *projection of y onto U*, it is given by $\hat{y} = P_U\, y = \sum_j \frac{y \cdot u_j}{u_j \cdot u_j} u_j$ and has the property that $\|z - y\| > \|\hat{y} - y\|$ for any other $z \in U$. If the basis is orthonormal, then $\hat{y} = QQ^T y$, where $Q = [u_1, ..., u_r]$.

  Proof. Define $\hat{y}$ and w by the formulas. Check that $\hat{y} \in U$ and $w \in U^\perp$. Note that $\|y - z\|^2 = \|y - \hat{y}\|^2 + \|\hat{y} - z\|^2$, which used twice also gives uniqueness. Last, the j-th entry of $Q^T y \in \mathbb{R}^{r\times 1}$ equals $y \cdot u_j$, hence $Q(Q^T y) = \sum_j (y \cdot u_j) u_j$.
- *Cauchy-Schwarz Inequality.* For every $u, v \in \mathbb{R}^n$, we have $|u \cdot v| \leqslant \|u\| \cdot \|v\|$.

  Proof. If $v \neq \mathbf{0}$, write $u = \frac{u \cdot v}{v \cdot v} v + w$, so $\|u\|^2 = (\frac{u \cdot v}{v \cdot v}\|v\|)^2 + \|w\|^2 \geqslant (\frac{u \cdot v}{\|v\|})^2$.
- *Triangle Inequality.* For every $u, v \in \mathbb{R}^n$, we have $\|u + v\| \leqslant \|u\| + \|v\|$.

  Proof. Expand $\|u + v\|^2$ and use Cauchy-Schwarz Inequality.

20 Factorizations and least squares

Main reference: Lay §2.5, file PLU.pdf, Lay §§6.4–6.5

- The *PLU factorization* consists in row reduction with bookkeeping, combined with *partial pivoting* (choose the largest candidate for pivot). Start with $P = I, L = I, U = A$, so $PA = LU$. At each step, update the factors while keeping the factorization valid: $(QP)A = (QLQ)(QU)$ for row exchange and $PA = (LE^{-1})(EU)$ for row replacement. This way, P is always a permutation, L is always lower triangular, and U becomes upper triangular at the end. Example:

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} -2 & -2 & -1 \\ 1 & -1 & 6 \\ -4 & 1 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ -\frac{1}{4} & \frac{3}{10} & 1 \end{bmatrix} \begin{bmatrix} -4 & 1 & -2 \\ 0 & -\frac{5}{2} & 0 \\ 0 & 0 & \frac{11}{2} \end{bmatrix}.$$

- Let $v_1, \ldots, v_m$ be LI and $W_j = \mathsf{span}(v_1, \ldots, v_j)$. The *Gram-Schmidt* procedure gives orthogonal vectors $u_1, \ldots, u_k$ such that $\mathsf{span}(u_1, \ldots, u_j) = W_j$, as follows:

$$u_1 = v_1, \quad u_{j+1} = v_{j+1} - P_{W_j} v_{j+1}.$$

- To get an orthonormal family we can take $w_j = \frac{1}{\|u_j\|} u_j$.

- The *QR factorization* consists in writing $A \in \mathbb{R}^{n \times k}$ as $A = QR$ where $Q \in \mathbb{R}^{n \times k}$ has orthonormal columns and $R \in \mathbb{R}^{k \times k}$ is upper triangular. Q can be found by applying Gram-Schmidt to the columns of A, and $R = Q^T A$. Example:

$$\begin{bmatrix} 1 & -1 & 4 \\ 1 & 4 & -2 \\ 1 & 4 & 2 \\ 1 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 1/2 & -1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 \end{bmatrix} \begin{bmatrix} 2 & 3 & 2 \\ 0 & 5 & -2 \\ 0 & 0 & 4 \end{bmatrix}.$$

- Given $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, a *least-squares solution* to the equation $Ax = b$ is a vector $\hat{x} \in \mathbb{R}^n$ that minimizes $\|Ax - b\|$.

- Least-squares solutions exist and are given by *normal equations* $A^T A \hat{x} = A^T b$.

  Proof. Since the set of possible values of Ax is exactly the subspace $\mathsf{range}\, A$, the distance $\|Ax - b\|$ will be minimized when $A\hat{x}$ equals the orthogonal projection of b onto $\mathsf{range}\, A$. This is equivalent to $(A\hat{x} - b) \perp a_j$ for each column a_j.

- The minimizer $\hat{x}$ is unique when $A^T A$ is invertible. In this case, $R\hat{x} = Q^T b$.

  Example: with same A as above and $b = (20, 20, 20, 0)$ we have $\|A\hat{x} - b\| = 10$.

21 Real spectral theorem and sketching simple conics

Main reference: Lay §7.1, §7.2

- For symmetric $A \in \mathbb{R}^{n\times n}$, eigenvectors of different eigenvalues are orthogonal.

  Proof. Follows from $(Av_1)\cdot v_2 = v_1 \cdot (Av_2)$.

- We say that $A \in \mathbb{R}^{n\times n}$ is *orthogonally diagonalizable* is there is an orthogonal matrix $P \in \mathbb{R}^{n\times n}$ such that $A = PDP^T$.

- *Real Spectral Theorem.* $A \in \mathbb{R}^{n\times n}$ is orthogonally diagonalizable iff A is symmetric.

  Proof. We postpone the proof that symmetric matrices are always diagonalizable. Assuming this fact, by Gram-Schmidt we can find an orthogonal basis to each eigenspace, and the reunion of the bases of all eigenspaces is orthogonal by the previous proposition, so a symmetric matrix is orthogonally diagonalizable. The converse is immediate: $A^T = (P^T)^T D^T P^T = PDP^T = A$.

- *Spectral Decomposition.* Let $P = [u_1, \ldots, u_n]$ be an orthogonal matrix that diagonalizes A. Then A can be decomposed as a sum of rank-1 matrices:

  $$A = \sum_{j=1}^{n} \lambda_j [u_j u_j^T]_{n\times n}$$

  Remark. The matrix $u_j u_j^T$ projects vectors orthogonally onto $\mathsf{span}(u_j)$.

  Proof. This is the column-row expansion of the product $(PD)P^T$.

- ○ A *quadratic form* on $\mathbb{R}^n$ is a polynomial of n variables having only terms of degree two. It can represented in a unique way as $x \cdot Ax$ for symmetric $A \in \mathbb{R}^{n\times n}$.

- If we make an orthogonal *change of variables* $x = Py$, where y represents the coordinates of x with respect to the columns of P, the quadratic form becomes $y \cdot (P^TAP)y$. By the Spectral Theorem, it is possible to choose P so that (P^TAP) is diagonal, so the quadratic form has no cross-product terms.

- Example: Sketch the graph of $5x_1^2 - 4x_1x_2 + 5x_2^2 = 48$.

  Diagonalizing $[5, -2; -2, 5]$, we get $P = [u_1, u_2]$ with $u_1 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, $u_2 = (\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ and $D = \mathsf{diag}(3, 7)$, so this is an ellipse with $a = 4$ and $b = \sqrt{48/7}$.

- Example: Sketch the graph of $x_1^2 - 8x_1x_2 - 5x_2^2 = 16$.

  Diagonalizing $[1, -4; -4, -5]$, we get $P = [u_1, u_2]$ with $u_1 = (\frac{2}{\sqrt{5}}, \frac{-1}{\sqrt{5}})$, $u_2 = (\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}})$ and $D = \mathsf{diag}(3, -7)$, so this is a hyperbola with $a = \frac{4\sqrt{3}}{3}$ and $b = \frac{4\sqrt{7}}{7}$.

22 Spaces and subspaces revisited

Main reference: Axler §1.C, §2.A, §2.B, §2.C

The last lectures were all about matrices and the spaces $\mathbb{R}^n$, $\mathbb{C}^n$ or $\mathbb{F}^n$. We now switch back to abstract vector spaces V over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and consider subspaces U, W, etc.

- The sum $U_1 + \cdots + U_m$ is a *direct sum* if for every $x \in (U_1 + \cdots + U_m)$, there exist unique vectors $u_1 \in U_1, \ldots, u_m \in U_m$ such that $x = u_1 + \cdots + u_m$.

- In case $U_1 + \cdots + U_m$ is a direct sum, we also denote it by $U_1 \oplus \cdots \oplus U_m$ as a way to indicate this property.

- The sum $U_1 + \cdots + U_m$ is a direct sum iff the only m-tuple $u_1 \in U_1, \ldots, u_m \in U_m$ that gives $u_1 + \cdots + u_m = \mathbf{0}$ is the trivial combination $u_1 = \cdots = u_m = \mathbf{0}$.

  Proof. For the converse, take two representations of a given x and subtract.

- The sum $U + W$ is a direct sum if and only if $U \cap W = \{\mathbf{0}\}$.

  Proof. If sum is direct, for $v \in U \cap W$ we have $v + (-v) = \mathbf{0}$, implying that $v = \mathbf{0}$. If $U \cap W = \{\mathbf{0}\}$, solutions to $u + w = \mathbf{0}$, are trivial since $w = -u \in U \cap W$.

- If $\dim V < \infty$ and U is a subspace, there is a subspace W such that $V = U \oplus W$.

  Proof. Complete a basis and show uniqueness of $v = u + w$.

◦ For a direct sum $U \oplus W$, we have $\dim(U \oplus W) = \dim U + \dim W$.

  Proof. Join any two bases $u_1, \ldots, u_k$ for U and $w_1, \ldots, w_m$ for W. See what linear combinations give $\mathbf{0}$ by first considering $u + w = \mathbf{0}$. Infinite case is trivial.

◦ Suppose $\dim V < \infty$. If $\dim(U + W) = \dim U + \dim W$, then the sum is direct.

  Proof. Assume the general equality below holds for every vector space V and subspaces U and W. When $\dim V < \infty$ we can subtract and get $\dim(U \cap W) = \dim U + \dim W - \dim(U + W) = 0$, so $U \cap W = \{\mathbf{0}\}$ and hence $U + W = U \oplus W$.

◦ For V vector space, U, W subspaces, $\dim U + \dim W = \dim(U + W) + \dim(U \cap W)$.

  Proof. If $\dim U = \infty$ or $\dim W = \infty$, we have $\dim(U + W) = \infty$ and the equality holds. So we can assume that V is finite-dimensional (otherwise instead of V use $\tilde{V} = U + W$ which is finite-dimensional). Let $Z = U \cap W$. Take $\tilde{U}$ and $\tilde{W}$ such that $U = Z \oplus \tilde{U}$ and $W = Z \oplus \tilde{W}$. We will show that $(\tilde{U} \oplus Z) + \tilde{W}$ is a direct sum, so $\dim(U + W) = \dim(\tilde{U} \oplus Z) + \dim \tilde{W} = \dim \tilde{U} + \dim Z + \dim \tilde{W} = \dim U + \dim W - \dim Z$, proving the desired equality. Suppose $u + z + w = \mathbf{0}$ with $u \in \tilde{U}, z \in Z, w \in \tilde{W}$. Then $w = -z - u \in U$, so $w \in U \cap \tilde{W} \subseteq Z$. But $Z \cap \tilde{W} = \{\mathbf{0}\}$, hence $w = \mathbf{0}$, proving the claim.

23 Linear maps revisited

Main reference: Axler §3.B, §3.D

- A function $T : V \to W$ is called *injective* if $Tu = Tv$ implies $u = v$.
- Let $T \in \mathcal{L}(V, W)$. Then T is injective if and only if $\ker T = \{\mathbf{0}\}$.

  Proof. Use that $Tu = Tv$ if and only if $(u - v) \in \ker T$.
- A function $T : V \to W$ is called *surjective* if $\operatorname{range} T = W$.
- *Rank-Nullity Theorem.* For $T \in \mathcal{L}(V, W)$, $\dim \operatorname{range} T + \dim \ker T = \dim V$.

  Proof. Seen in Lecture 11.
- If $\dim W < \dim V < \infty$, then $T \in \mathcal{L}(V, W)$ cannot be injective.

  Proof. By Rank-Nullity Theorem, $\dim \ker T > 0$.
- If $\dim V < \dim W < \infty$, then $T \in \mathcal{L}(V, W)$ cannot be surjective.

  Proof. By Rank-Nullity Theorem, $\dim \operatorname{range} T < \dim W$.
- A linear map is invertible if and only if it is injective and surjective.

  Proof. Seen in Lecture 6. Need to check that the inverse is linear.
- ○ Finite-dimensional spaces are isomorphic iff they have the same dimension.

  Proof. Let V and W be finite-dimensional spaces and let $v_1, \ldots, v_n$ be a basis for V. If there exists an isomorphism $T \in \mathcal{L}(V, W)$, then $Tv_1, \ldots, Tv_n$ is a basis for W and hence $\dim W = n$. Conversely, suppose $\dim W = n$. Take $w_1, \ldots, w_n$ a basis for W and define $T \in \mathcal{L}(V, W)$ by $Tv_1 = w_1, \ldots, Tv_n = w_n$. Then T maps a basis to a basis, and hence it is an isomorphism.
- For finite-dimensional spaces V and W, $\dim \mathcal{L}(V, W) = (\dim V)(\dim W)$.

  Proof. We will show that the space $\mathcal{L}(V, W)$ is isomorphic to $\mathbb{F}^{m \times n}$. Fix a basis $\mathcal{A} = v_1, \ldots, v_n$ for V and $\mathcal{B} = w_1, \ldots, w_m$ for W. Define $R : \mathcal{L}(V, W) \to \mathbb{F}^{m \times n}$ by $R(T) = [T]_{\mathcal{B}\mathcal{A}}$. This R is linear and bijective, so it is an isomorphism.
- Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then the following are equivalent: (a) T is invertible; (b) T is injective; (c) T is surjective.

  Proof. By the Rank-Nullity Theorem, (c) is equivalent to $\ker T = \{\mathbf{0}\}$, which in turn is equivalent to (b). By above proposition, (a) is equivalent to "(b) and (c)" and this completes the proof.

24 Invariant spaces and eigenvectors

Main reference: Axler §5.A, §5.B

- A number $\lambda \in \mathbb{F}$ is called an *eigenvalue of* $T \in \mathcal{L}(V)$ if $Tv = \lambda v$ for some $v \neq \mathbf{0}$.
- A number $\lambda \in \mathbb{F}$ is an eigenvalue of T if and only if $T - \lambda I$ is not injective.
  Proof. $Tv = \lambda v$ is equivalent to $v \in \mathsf{ker}(T - \lambda I)$.
- A vector v is an *eigenvector of* T corresponding to $\lambda \in \mathbb{F}$ if $v \neq \mathbf{0}$ and $Tv = \lambda v$.
- Eigenvectors corresponding to distinct eigenvalues are linearly independent.
  Proof. Apply $T - \lambda_m I$ to a null linear combination, and use induction on m.
- If V is finite-dimensional then $T \in \mathcal{L}(V)$ has at most $\mathsf{dim}\, V$ distinct eigenvalues.
  Proof. A LI family has at most $\mathsf{dim}\, V$ vectors.
- A subspace U of V is said to be *invariant under* T if $Tu \in U$ for any $u \in U$.
  Examples: $\{\mathbf{0}\}$, V, $\mathsf{ker}\, T$, $\mathsf{range}\, T$, $\mathsf{range}\, T^2$.
- We define $T^0 = I$, $T^{m+1} = T^m T$.
  For $p \in \mathcal{P}(\mathbb{F})$ and $T \in \mathcal{L}(V)$, we define $p(T) = a_n T^n + \cdots + a_2 T^2 + a_1 T + a_0 I \in \mathcal{L}(V)$.
- Factoring polynomials: $(pq)(T) = p(T)q(T)$. In particular, $p(T)q(T) = q(T)p(T)$.
  Proof. Expanding and using the distributive property works for T as it does for z.
- Let $\mathcal{B} = v_1, \ldots, v_n$ be a basis for V and $T \in \mathcal{L}(V)$. These are equivalent:
  (a) $[T]_{\mathcal{B}}$ is upper-triangular;
  (b) $Tv_j \in \mathsf{span}(v_1, \ldots, v_j)$ for $j = 1, \ldots, n$;
  (c) $\mathsf{span}(v_1, \ldots, v_j)$ is invariant under T for $j = 1, \ldots, n$.
  Proof. (b⇒c) For $v = \alpha_1 v_1 + \cdots + \alpha_j v_j$, $Tv \in \mathsf{span}(v_1) + \cdots + \mathsf{span}(v_1, \ldots, v_j)$.
- If V is complex finite-dimensional, and $T \in \mathcal{L}(V)$, then T has an eigenvalue.
  Proof without determinant. Since $\mathsf{dim}\, \mathcal{L}(V) = n^2$, the family $I, T, T^2, \ldots, T^{n^2}$ is LD. Hence there is a linear combination $\alpha_0 I + \alpha_1 T + \alpha_2 T^2 + \cdots + \alpha_k T^k v = \mathbf{0}$ with $\alpha_k = 1$. Now the polynomial $\sum_{j=0}^{k} \alpha_j z^j$ can be factorized as $(z - \lambda_1) \cdots (z - \lambda_k)$, so $(T - \lambda_1 I) \cdots (T - \lambda_k I) = \mathbf{0}$, and thus one of the factors is not injective.
- If V is complex finite-dimensional, then $[T]_{\mathcal{B}}$ is upper-triangular for some basis $\mathcal{B}$.
  Proof. We prove by induction on n. Take λ as an eigenvalue. Subspace $U = \mathsf{range}(T - \lambda I) \neq V$ is invariant because $Tu = (T - \lambda I)u + \lambda u$. For the restriction $T_{|_U}$, by induction there is a basis $u_1, \ldots, u_k$ for U such that $Tu_j \in \mathsf{span}(u_1, \ldots, u_j)$ for $j = 1, \ldots, k$. Complete it to a basis $u_1, \ldots, u_k, v_{k+1}, \ldots, v_n$ for V. Now $Tv_j = \lambda v_j + u$ for $u \in U$, so $Tv_j \in \mathsf{span}(u_1, \ldots, u_k, v_j)$, hence $[T]_{\mathcal{B}}$ is upper triangular.
  Counter-example. $T(x, y) = (-y, x)$ on $\mathbb{R}^2$ cannot be made upper-triangular.

25 Decomposition into eigenspaces

Main reference: Axler §5.C

Assume the dimension of V is finite, denoted n.

- $T \in \mathcal{L}(V)$ is *diagonalizable* if there exists a basis $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}$ is diagonal.
- The *eigenspace* of T corresponding to $\lambda \in \mathbb{F}$ is defined as
$$E(\lambda, T) = \mathsf{ker}(T - \lambda I).$$
- Let $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$ denote distinct eigenvalues of T. Then
$$E(\lambda_1, T) + \cdots + E(\lambda_m, T) = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T).$$
Proof. Check that $u_1 + \cdots + u_m = \mathbf{0}$, $u_j \in E(\lambda_j, T)$ only has the trivial solution.
- Let $\lambda_1, \ldots, \lambda_m$ be all distinct eigenvalues of T. The following are equivalent:
  1. T is diagonalizable;
  2. V has a basis $u_1, \ldots, u_n$ consisting of eigenvectors of T;
  3. There are invariant one-dimensional $U_1, \ldots, U_n$ such that $V = U_1 \oplus \cdots \oplus U_n$;
  4. $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$;
  5. $\mathsf{dim}\, E(\lambda_1, T) + \cdots + \mathsf{dim}\, E(\lambda_m, T) = n$.

  Proof. ($1 \Leftrightarrow 2$) by definition of $[T]_{\mathcal{B}}$.
  ($2 \Rightarrow 3$) Take $U_j = \mathsf{span}(u_j)$. Then $U_1 + \cdots + U_n = V$, and the sum is direct.
  ($3 \Rightarrow 2$) Take $u_j \in U_j \setminus \{\mathbf{0}\}$ eigenvector. $\{u_1, \ldots, u_n\}$ spans V, so it is a basis.
  ($2 \Rightarrow 4$) If eigenvectors span V, we have $E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T) = V$.
  ($4 \Rightarrow 2$) Let $\mathcal{A}_j$ be a basis for $E(\lambda_j, T)$ and take $\mathcal{A} = \mathcal{A}_1, \ldots, \mathcal{A}_m$. Since $\mathsf{span}\, \mathcal{A} = V$, it contains a basis for V, and its elements are all eigenvectors.
  ($4 \Leftrightarrow 5$) Property of direct sum.
- If T has n distinct eigenvalues, then T is diagonalizable.

  Proof. There are n linearly independent eigenvectors, which thus form a basis.
- We define the *determinant of an operator* $T \in \mathcal{L}(V)$ by $\mathsf{det}\, T = \mathsf{det}[T]_{\mathcal{B}}$ for some basis $\mathcal{B}$. The *trace* is defined as $\mathsf{trace}\, T = \mathsf{trace}[T]_{\mathcal{B}}$. The definitions do not depend on the choice of basis because similar bases have the same trace and determinant.
- We define the *characteristic polynomial of an operator* $T \in \mathcal{L}(V)$ by $p_T(z) = \mathsf{det}(T - zI)$. A number $\lambda \in \mathbb{F}$ is an eigenvalue if and only if it is a root of p_T. In this case, we define its *algebraic multiplicity* as its multiplicity as a root of p_T, and its geometric multiplicity as $\mathsf{dim}\, \mathsf{ker}(T - \lambda I)$.
- If V is a complex vector space, then $T \in \mathcal{L}(V)$ has n eigenvalues counting algebraic multiplicity. Moreover, $\mathsf{det}\, T = \prod_{j=1}^{n} \lambda_j$ and $\mathsf{trace}\, T = \sum_{j=1}^{n} \lambda_j$.