Linear Algebra

Bendit Chan

March 25, 2022

Disclaimer

This is a set of handouts based on the course M40003 – Linear Algebra & Groups in Imperial College London. The content is heavily based on the official course notes, as well as Dexter Chua's notes[1]and Evan Chen's "*An Infinitely Large Napkin*"[2], so all credits go to them.

[1] https://dec41.user.srcf.net/notes/IB_M/linear_algebra.pdf
[2] https://venhance.github.io/napkin/Napkin.pdf

Contents

Note. Following Ravi Vakil's style, we use $\star$ to denote topics worth knowing on a second (but not first) reading.

1 System of Linear Equations

In this section, we will be focusing on how to solve linear equations.

Warning: In retrospect...

Throughout studying linear algebra it would often seem like it is a study of *matrices* or computations for solutions of linear equations. But this is missing the main part of linear algebra, which is actually the study of *linear maps*. Everything should be stemmed from the concept of linear maps, and things would be much more meaningful than just an array of numbers. Hence:

> **! Keypoint**
> Treating this as a prerequisite towards linear algebra is fine, but putting this section as the first doesn't necessarily imply that it is important.

1.1 Definitions and Notations

By a system of **linear equations** we mean a family of equations in the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

Here, there are m equations with n unknowns, and we might represent this in matrix form:

Definition 1.1
Given a system of m linear equations in n unknowns, the **matrix form** is

$$A\mathbf{x} = B,$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ and $B = (b_1, b_2, \ldots, b_n)^T$ are column matrices, and

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

is an $m \times n$ matrix. The **augmented matrix** is then defined as

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right)$$

which is another way to represent the system.

Remark. A reasonable question is what class of sets are the coefficients and the unknowns in. For now, we might safely assume it to be $\mathbb{R}$, but any field F would work.

Matrix algebra is skipped here and is assumed to be familiar by the reader.

1.2 Row operations

To solve a system of linear equations, there are three operations we can do:

- multiply an equation by a non-zero factor;
- add a multiple of one equation to another;
- swap two equations.

These three operations are captured by the so-called **row operations** in an augmented matrix.

Definition 1.2 (Elementary row operations)

We define an **elementary row operation** to be one of the following operations performed on an augmented matrix:

- multiply a row by a non-zero factor;
- add a multiple of a row to another;
- swap two rows.

As augmented matrices represent the original system and each row corresponds to one equation, it is natural that these three operations preserve the solutions of a linear system. We also note here that each row operation has an inverse row operation. Thus it makes sense to define:

Definition 1.3

Two systems of linear equations are **equivalent** if either

- they both have no solution (so they are **inconsistent**);
- or the augmented matrix of the second system can be obtained via row operations from the first.

Equivalently, two systems are equivalent if they have the same set of solutions.

The process in solving system of equations using row operations comes from the following motivation:

Motivation

Intuitively, our goal is to perform row operations so that the matrix is transformed into the form

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right) \xrightarrow{\text{row operations}} \left(\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & c_1 \\ 0 & 1 & \cdots & 0 & c_2 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & c_m \end{array}\right)$$

so that we can just "read off" the solution as

$$\mathbf{x} = (c_1, c_2, \ldots, c_m)^T,$$

since the two matrices must have the same set of solutions as they are equivalent.

Unfortunately, this is clearly impossible if $m \neq n$, but turns out this is also not always possible (more details later) even if $m = n$. In any case, row operations can still be performed so that the goal is somehow reached.

We will first look at an example where the goal can be reached:

Example 1.4

Consider the system of equations

$$\begin{aligned} 3x - 2y + z &= -6 \\ 4x + 6y - 3z &= 5 \\ -4x + 4y &= 12 \end{aligned}$$

represented by the augmented matrix

$$\left(\begin{array}{ccc|c} 3 & -2 & 1 & -6 \\ 4 & 6 & -3 & 5 \\ -4 & 4 & 0 & 12 \end{array}\right)$$

By performing row operations,

$$\begin{aligned} \left(\begin{array}{ccc|c} 3 & -2 & 1 & -6 \\ 4 & 6 & -3 & 5 \\ -4 & 4 & 0 & 12 \end{array}\right) &\xrightarrow{R_3 \mapsto -\frac{1}{4}R_3} \left(\begin{array}{ccc|c} 3 & -2 & 1 & -6 \\ 4 & 6 & -3 & 5 \\ 1 & -1 & 0 & -3 \end{array}\right) \\ &\xrightarrow[R_1 \mapsto R_1 - 3R_3]{R_2 \mapsto R_2 - 4R_3} \left(\begin{array}{ccc|c} 0 & 1 & 1 & 3 \\ 0 & 10 & -3 & 17 \\ 1 & -1 & 0 & -3 \end{array}\right) \\ &\xrightarrow{R_2 \mapsto R_2 - 10R_1} \left(\begin{array}{ccc|c} 0 & 1 & 1 & 3 \\ 0 & 0 & -13 & -13 \\ 1 & -1 & 0 & -3 \end{array}\right) \\ &\xrightarrow{R_2 \mapsto -\frac{1}{13}R_2} \left(\begin{array}{ccc|c} 0 & 1 & 1 & 3 \\ 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & -3 \end{array}\right) \\ &\xrightarrow{R_1 \mapsto R_1 - R_2} \left(\begin{array}{ccc|c} 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & -3 \end{array}\right) \\ &\xrightarrow{R_3 \mapsto R_3 + R_1} \left(\begin{array}{ccc|c} 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & -1 \end{array}\right) \\ &\xrightarrow[R_2 \mapsto R_3]{\substack{R_3 \mapsto R_1 \\ R_1 \mapsto R_2}} \left(\begin{array}{ccc|c} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{array}\right) \end{aligned}$$

So the solution is $(x, y, z) = (-1, 2, 1)$.

However, in some cases it is impossible for the augmented matrix to be transformed into such a form:

Example 1.5

If a system is reduced to

$$\left(\begin{array}{ccc|c} 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 5 \end{array}\right)$$

then x has no constraints and its solution set corresponds to $(x, y, z) = (\lambda, 2, 5)$ where $\lambda \in \mathbb{R}$ is arbitrary.

The key idea here is that even though the matrix is not in the form of the goal mentioned before, it is still easy to figure out the solution in such a form. This gives the following definition.

Definition 1.6
We say a matrix is in **echelon form** if it satisfies the following:

- all of the zero rows are at the bottom;
- the first non-zero entry in each row is 1;
- the first non-zero entry in row i is strictly to the left of the first non-zero entry in row $i+1$.

We say a matrix is in **row reduced echelon form** if it is in echelon form and

- if the first non-zero entry in row i appears in column j, then every other elements in column j is zero.

For example,

$$\left(\begin{array}{ccc|c} 1 & 1 & 2 & 2 \\ 0 & 1 & 7 & 12 \\ 0 & 0 & 1 & -10 \\ 0 & 0 & 0 & 0 \end{array}\right) \qquad \left(\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

echelon form row reduced echelon form

Caution: Notice that in row reduced echelon form, the column of B has to satisfy the new condition too.

Clearly row reduced echelon form is the "best possible" goal that we can reach via row operations. But it is often simple enough to figure out the solution from an echelon form, so that less row operations have to be performed.

1.3 More matrices

We cover some more definitions on matrices here.

Definition 1.7
We say a matrix is **square** if its number of rows is equal to its number of columns.
A square matrix $A = (a_{ij})_{n\times n}$ is said to be

- **upper triangular** if $a_{ij} = 0$ whenever $i > j$ (all zeros below the diagonal).
- **lower triangular** if $a_{ij} = 0$ whenever $i < j$ (all zeros above the diagonal).
- **diagonal** if $a_{ij} = 0$ whenever $i \neq j$.

For example,

$$\begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 7 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

upper triangular lower triangular diagonal

Definition 1.8
The $n \times n$ **identity matrix**, denoted by I_n, is a matrix with all diagonal entries as 1 and all other entries as 0. It is the multiplicative identity for all $n \times n$ matrices A, i.e.

$$I_n A = A I_n = A.$$

If for square matrix A, there exists a matrix B such that $AB = BA = I$, then we say A is **invertible** and B is an **inverse** of A. If A is not invertible, we might sometimes call it a **singular** matrix too.

Example 1.9

Consider $A = \begin{pmatrix} 2 & 0 \\ 1 & -1 \end{pmatrix}$. Then $B = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & -1 \end{pmatrix}$ is an inverse of A. Indeed, one could check that

$$\begin{pmatrix} 2 & 0 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & -1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & -1 \end{pmatrix}\begin{pmatrix} 2 & 0 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

An important theorem is as follows.

Theorem 1.10 (Inverse is unique)

If there exists square matrices A, B, C such that $AB = CA = I$, then $B = C$.

Proof. We have

$$B = IB = (CA)B = C(AB) = CI = C$$

since matrix multiplication is associative. □

Hence, if A is invertible, we can talk about *the* inverse of A, denoted by A^{-1}.

Example 1.11

Suppose A, B are both invertible square matrices. Then AB is also invertible with inverse $B^{-1}A^{-1}$ since

$$AB \cdot B^{-1}A^{-1} = AIA^{-1} = I.$$

It can be shown that if $AB = I$ then we also have $BA = I$. A standard argument uses determinants: we have $\det(A)\det(B) = 1$ and so B is also invertible. Hence

$$I = BB^{-1} = B(AB)B^{-1} = (BA)BB^{-1} = BA.$$

Definition 1.12 (Transpose)

If $A = (a_{ij})_{m\times n}$, then the **transpose** of A is $A^T = (a_{ji})_{n\times m}$.

We have shown that $(AB)^{-1} = B^{-1}A^{-1}$ for square matrices A and B. Similarly, one could show the following:

Example 1.13

Suppose A, B are arbitrary matrices. Then $(AB)^T = B^TA^T$ since

- they have the same order $p \times n$;
- the ij^{th} entry of AB is $\sum a_{ik}b_{kj}$, which is the ji^{th} entry of $(AB)^T$;
- the ji^{th} entry of B^TA^T is $\sum (b^T)_{jk}(a^T)_{ki} = \sum a_{ik}b_{kj}$.

Transposing and taking inverse are also commutative:

Theorem 1.14

If A is an invertible square matrix, then $(A^T)^{-1} = (A^{-1})^T$.

Proof. From the definition of inverse,

$$I = AA^{-1} \Longrightarrow I = I^T = (AA^{-1})^T = (A^{-1})^TA^T$$

and similarly $A^T(A^{-1})^T = I$. Hence $(A^{-1})^T$ is the inverse of A^T, as needed. □

1.4 Elementary matrices

After an interlude of more matrices, we can go back to row operations. Indeed, row operations can also be described using matrix multiplications via the use of elementary matrices.

Definition 1.15 (Elementary matrices)

An **elementary matrix** is a matrix that can be obtained from an identity matrix by means of *one* elementary row operation. They are denoted by

- $E_r(\alpha)$: multiplying the row r by α;
- $E_{rs}(\alpha)$: adding a multiple of row s by a factor of α to row r.
- E_{rs}: swapping row r and s.

At the end, we will be able to use elementary matrices to compute the inverse of a matrix. The main two tools we will use are as follows:

Proposition 1.16

Let A be an $m \times n$ matrix and E be an elementary $m \times m$ matrix. Then EA is the matrix resulted in applying the row operation corresponding to E applied on A.

Proof. The cases of $E_r(\alpha)$ and E_{rs} are trivial. For $E_{rs}(\alpha)$, we can write

$$E_{rs}(\alpha)A = (I_m + \alpha Z_{rs})A = A + \alpha Z_{rs}A$$

where Z_{rs} is the matrix with all entries zero except the term in row r, column s being one. Then it is clear that $Z_{rs}A$ is the matrix with all entries zero except row r being the s-th row of A, so $A + \alpha Z_{rs}A$ is exactly the row operation of adding a multiple of row s by a factor of α to A. □

Proposition 1.17

Every elementary matrix is invertible and the inverse is also an elementary matrix.

Proof. From the corresponding row operations, one can check that

$$\begin{aligned} E_r(\alpha)E_r(\alpha^{-1}) = E_r(\alpha^{-1})E_r(\alpha) &= I \\ E_{rs}(\alpha)E_{rs}(-\alpha) = E_{rs}(-\alpha)E_{rs}(\alpha) &= I \\ E_{rs}E_{rs} &= I \end{aligned}$$

and hence they are invertible. □

Combining both tools, we have a result on how to compute inverses:

Theorem 1.18

Suppose a square matrix A can be reduced to an identity matrix via row operations. Then A is invertible and the inverse of A is found by applying the same row operations to I.

Proof. Let $E_1, E_2, \ldots, E_r$ be the elementary matrices corresponding to the row operations applied on A so that it is reduced to I. By Proposition 1.16,

$$E_r \cdots E_2 E_1 A = I.$$

Hence $A = E_1^{-1}E_2^{-1}\cdots E_r^{-1}$ as elementary matrices are invertible, and $A^{-1} = E_r \cdots E_2 E_1$. Yet this can be viewed as applying the row operations to the matrix I, which gives our desired result. □

Example 1.19

Consider the matrix

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 2 & 0 \\ 3 & 0 & 4 \end{pmatrix}$$

We now know that to find its inverse, we can reduce it into I and apply the same row operations to I. We can do both things at once by constructing the augmented matrix $A|I$:

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 & 1 & 0 \\ 3 & 0 & 4 & 0 & 0 & 1 \end{array}\right)$$

Then by applying row operations,

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 & 1 & 0 \\ 3 & 0 & 4 & 0 & 0 & 1 \end{array}\right) \xrightarrow[R_3 \mapsto R_3 - 3R_1]{R_2 \mapsto R_2 - R_1} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & -1 & -1 & 1 & 0 \\ 0 & 0 & 1 & -3 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{R_2 \mapsto R_2 + R_3} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & -4 & 1 & 1 \\ 0 & 0 & 1 & -3 & 0 & 1 \end{array}\right)$$

$$\xrightarrow[R_2 \mapsto \frac{1}{2} R_2]{R_1 \mapsto R_1 - R_3} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 4 & 0 & -1 \\ 0 & 1 & 0 & -2 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 & -3 & 0 & 1 \end{array}\right)$$

and hence the inverse of A is

$$A^{-1} = \begin{pmatrix} 4 & 0 & -1 \\ -2 & \frac{1}{2} & \frac{1}{2} \\ -3 & 0 & 1 \end{pmatrix}$$

1.5 Fields

As promised before, entries in matrices or coefficients in systems of linear equations are not limited to being in $\mathbb{R}$; in fact, all of our discussion above would have worked if we replace $\mathbb{R}$ by an arbitrary **field**.

Definition 1.20 (Field)

A **field** is a set F with two binary operations, addition $+$ and multiplication $\cdot$, such that

- $(F, +)$ is an abelian group with identity 0;
- $(F^\times, \cdot)$ (the set of non-zero elements) is an abelian group with identity 1;
- $a \cdot (b + c) = a \cdot b + a \cdot c$ for all $a, b, c \in F$ (distributive).

Example 1.21

Here are some classic examples (and non-examples) of fields:

- $\mathbb{R}, \mathbb{C}, \mathbb{Q}$ are all fields.
- If p is a prime, then $\mathbb{F}_p = \{0, 1, \ldots, p-1\}$ is a field with addition and multiplication modulo p.
- $\mathbb{F}_6$ as defined above is **not** a field, since 3 does not have an multiplicative inverse.

2 Vector Spaces

We now enter the regime of the most important concept in linear algebra: **vector spaces**. This set up the stage of all of our proceeding discussion.

2.1 Definitions and Examples

We already know intuitively that $\mathbb{R}^n$ is a "vector space":

Motivation

In $\mathbb{R}^n$, elements are represented by vectors $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ where $v_i \in \mathbb{R}$ for $1 \leq i \leq n$.

There are two crucial facts here:

- you can add and subtract two vectors, and there is an additive identity (zero vector) in $\mathbb{R}^n$;
- you can scale a vector **by a real number** and it will still be in $\mathbb{R}^n$.

This defines what's called a vector space, but $\mathbb{R}^n$ is just the canonical example: the elements are just lists of numbers.

Here comes the first definition which allows us to jump from viewing things as a list of numbers to abstract objects in a set:

Definition 2.1 (Vector space)

Let F be a field. A **vector space** over F is a non-empty set V together with the following maps:

$$\begin{aligned} \oplus : V \times V &\to V && \text{(addition)} \\ (\mathbf{v}_1, \mathbf{v}_2) &\mapsto \mathbf{v}_1 \oplus \mathbf{v}_2 \\ \odot : F \times V &\to V && \text{(scalar multiplication)} \\ (f, \mathbf{v}) &\mapsto f \odot \mathbf{v} \end{aligned}$$

satisfying the following axioms: for all $f, g \in F$ and $\mathbf{v}, \mathbf{w} \in V$,

- $(V, \oplus)$ is an abelian group with identity 0_V, the **zero vector**;
- scalar multiplication is distributive: $f \odot (\mathbf{v} \oplus \mathbf{w}) = (f \odot \mathbf{v}) \oplus (f \odot \mathbf{w})$ and $(f + g) \odot \mathbf{v} = (f \odot \mathbf{v}) \oplus (g \odot \mathbf{v})$;
- scalar multiplication is associative: $f \odot (g \odot \mathbf{v}) = (fg) \odot \mathbf{v}$;
- $1 \odot \mathbf{v} = \mathbf{v}$ where 1 is the identity element in F.

Then elements in V are called **vectors**, elements in F are called **scalars**, and we will sometimes refer to V as an F-vector space.

Caution: Notice that $\oplus$ and $\odot$ in V is different from the operations in F. However, we will still drop the notation from now on and simply use $+$ and $\cdot$.

Put aside all the abstract definitions, informally, this should be remembered as:

! Keypoint

An F-vector space is a structure where you can add two elements and scale by elements of F, subject to some constraints which make it well-behaved.

As said before, $\mathbb{R}^n$ is an $\mathbb{R}$-vector space. But many other vector spaces exist, and of course they can be more exotic and quite different from "a list of numbers":

Example 2.2 (Some $\mathbb{R}$-vector spaces)

Let's look at some of the easier examples of $\mathbb{R}$-vector spaces first:

- The set $M_{m\times n}(\mathbb{R})$ of all $m \times n$ matrices is a vector space: we can add two matrices and scale a matrix by a real number.
- The set of all polynomials of degree at most two, namely
$$\{ax^2 + bx + c : a, b, c \in \mathbb{R}\}$$
is a vector space: we can add two quadratics and multiply by constants.
- In fact, the set of *all* polynomials with real coefficients, denoted by $\mathbb{R}[x]$, is also an $\mathbb{R}$-vector space.

Note that in some of the examples above, we cannot multiply two elements in the vector space (for example, the product of two quadratics is not a quadratic). But that doesn't matter – in a vector space we do not need multiplication of two vectors.

Example 2.3 (Some stranger examples)

Some of the vector spaces can be more exotic:

- The set
$$\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$$
is a $\mathbb{Q}$-vector space. Notice how it is **not** an $\mathbb{R}$-vector space.
- The set of all **real valued functions** $\mathbb{R}^X$ on X, namely
$$\mathbb{R}^X := \{f : X \to \mathbb{R}\}$$
is an $\mathbb{R}$-vector space: addition is given by $(f + g)(x) := f(x) + g(x)$ and scalar multiplication is given by $(\lambda f)(x) := \lambda f(x)$. Notice how the operations on the left side differs in meaning from the right side.
- Similarly, let $[a, b]$ be a closed interval, then the set
$$C([a, b], \mathbb{R}) := \{f \in \mathbb{R}^{[a,b]} : f \text{ is continuous}\}$$
is an $\mathbb{R}$-vector space too by the operations above, since continuity is preserved.

Caution: The first example illustrates how the **base field is very important**. Indeed, a change of base field could result in the set turning into a non-vector space.

2.2 Subspaces

In mathematics, whenever we define "something", we would also like to define a "sub-something". In the case of vector spaces, this is a **subspace**.

Definition 2.4 (Subspace)

A subset W of a vector space V is a **subspace** of V, denoted by $W \leq V$, if

- W is non-empty;
- for $\mathbf{v}, \mathbf{w} \in W$, $\mathbf{v} + \mathbf{w}$ is also in W (closed under addition);
- for $\mathbf{v} \in W$ and $f \in \mathbf{R}$, $f\mathbf{v}$ is also in W (closed under scalar multiplication).

Remark. Note that V and the zero subspace are always subspaces of V. Any other subspace of V would hence be called a **proper subspace** of V.

Proposition 2.5

Every subspace of a F-vector space V must contain the zero vector.

Proof. We first show that $0\mathbf{v} = 0_V$ in any vector space V. Indeed, we have $0\mathbf{v} + 0\mathbf{v} = (0+0)\mathbf{v} = 0\mathbf{v}$, so $0\mathbf{v}$ is the additive identity in V, which is precisely the zero vector 0_V.

Now let U be a subspace of V. Since U is non-empty, we can pick $\mathbf{v} \in U$. Then $0_V = 0\mathbf{v} \in U$ since U is closed under scalar multiplication. □

An important fact below gives an equivalent definition of a subspace (which we will not prove):

! Keypoint

A subspace of a F-vector space V is itself a F-vector space, inheriting the operations from V.

If we have two subspaces U and W, there are several things we can do with them. For example, we can take the intersection $U \cap W$. We will show that this will be a subspace:

Theorem 2.6

Let U, W be subspaces of V. Then $U \cap W$ is a subspace of V.

Proof. It suffices to check:

- $U \cap W$ is non-empty since $0_V \in U$ and W;
- if $\mathbf{v}_1, \mathbf{v}_2 \in U \cap W$, then $\mathbf{v}_1 + \mathbf{v}_2 \in U$ and W as addition is closed in U and W. Hence $\mathbf{v}_1 + \mathbf{v}_2 \in U \cap W$;
- similarly scalar multiplication is closed.

Thus by definition, $U \cap W$ is a subspace of V. □

However, taking the union will in general not produce a vector space. For instance, consider

$$U = \{(x, 0) : x \in \mathbb{R}\} \quad \text{and} \quad W = \{(0, y) : y \in \mathbb{R}\}$$

where $V = \mathbb{R}^2$. Then $(1,0)$ and $(0,1)$ are in $U \cup W$, yet $(1,0) + (0,1) = (1,1) \notin U \cup W$.

2.3 Spans, Linear independence, and Bases

Recall that in $\mathbb{R}^n$, the "standard bases" made of the form $\mathbf{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ with 1 in the i-th component represent all vectors in a linear combination. This section generalises the idea for general vector spaces.

Motivation

As a first motivation, consider

$$V = \{ax^2 + bx + c : a, b, c \in \mathbb{R}\}$$

as above. Then every element can be represented as a sum of copies of $\{1, x, x^2\}$. It is not the only one: $\{2, x, x^2\}, \{x+4, x-2, x^2+x\}$ work just as fine, though not as natural. But $S = \{3 + x^2, x + 1, 5 + 2x + x^2\}$ should **not** be considered a basis for two reasons:

- it is impossible to write x^2 as a sum of elements of S;
- the representation is not unique: we have $0 = (3 + x^2) + 2(x + 1) - (5 + 2x + x^2)$ so we can just add whatever multiple of this expression to the linear combination.

This motivates the definition of **spanning** and **linearly independent**, as we will see later.

The most important result in this section is to prove that for any vector space, **any two basis must contain the same number of elements**. This means we can define the **dimension** of a vector space as the number of elements in a basis. We will also prove that the dimension is well-behaved. For example, a subspace of a vector space must have a smaller dimension than the larger space.

Remark. For experts: this is **not** true for modules over a ring R. Indeed, not all modules have a basis, and even for those that have a basis, the behaviour of the "dimension" is vastly different.

Definition 2.7 (Span)

Given an F-vector space V, a **linear combination** of some vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ is a sum of the form $\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n$, where $\alpha_1, \dots, \alpha_n \in F$.

The **span** of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ is then the set of linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$, i.e.

$$\operatorname{Span}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = \{\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n \in V : \alpha_1, \dots, \alpha_n \in F\}.$$

Note in particular that, by convention, we take the empty sum to be 0_V, so $\operatorname{Span}\emptyset = \{0_V\}$. In addition, for an infinite set S, we still only take finite sums, i.e.

$$\operatorname{Span}(S) = \left\{ \sum_{\mathbf{v}_i \in S'} \alpha_i \mathbf{v}_i : S' \overset{\text{finite}}{\subset} S, \alpha_i \in F \right\},$$

as infinite sums and the notion of convergence do not make sense in a general vector space.

Lemma 2.8

Let V be an F-vector space, and $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \in V$. Then $\operatorname{Span}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ is a subspace of V.

Proof. Clearly $S := \operatorname{Span}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) \subset V$, so it suffices to check:

- S is non-empty since $\mathbf{v}_1 \in S$;
- if $\mathbf{u}, \mathbf{w} \in S$ then $\mathbf{u} = \sum \alpha_i \mathbf{v}_i$ and $\mathbf{w} = \sum \beta_i \mathbf{v}_i$ for some α_i and β_i. Hence

  $$\mathbf{u} + \mathbf{w} = \sum_{i=1}^{n} (\alpha_i + \beta_i)\mathbf{v}_i \in S$$

  as F is closed under addition;
- similarly, if $\mathbf{u} \in S$ and $\lambda \in F$, then $\mathbf{u} = \sum \alpha_i \mathbf{v}_i$ and so $\lambda\mathbf{u} = \sum_{i=1}^{n} (\lambda\alpha_i)\mathbf{v}_i \in S$. □

Example 2.9

Let $V = \mathbb{R}^3$ and $S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \right\}$. Then $\operatorname{Span}(S) = \left\{ \begin{pmatrix} a \\ b \\ b \end{pmatrix} : a, b \in \mathbb{R} \right\}$.

Note that any subset of S with two elements has the same span as S.

Definition 2.10 (Spanning set)

Let V be an F-vector space, and suppose $S \subset V$ satisfies $\operatorname{Span}(S) = V$. Then we say S **spans** V or equivalently S is a **spanning set** for V.

Again, spanning sets are not unique for if $\operatorname{Span}(S) = V$ and $\mathbf{v} \in \operatorname{Span}(S)$, then $\operatorname{Span}(S \cup \{\mathbf{v}\}) = \operatorname{Span}(S) = V$, in the sense that adding $\mathbf{v}$ brings no contribution since the rest of the elements are "enough".

To avoid redundancy, we introduce a new concept:

Definition 2.11 (Linear independence)

Let V be an F-vector space. We say $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in V$ are **linearly independent** if whenever

$$\sum_{i=1}^{n} \alpha_i \mathbf{v}_i = 0_V$$

we must have $\alpha_i = 0$ for all i. A set is **linearly dependent** if it is not linearly independent.

Example 2.12

Using the above example, $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\}$ is a linearly independent subset of $\mathbb{R}^3$, but $S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \right\}$ is linearly dependent since

$$1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + (-1) \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} = 0_V.$$

In addition, S does not span V since $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \notin \operatorname{Span}(S)$.

Before moving on to bases, we have a property for linearly independent sets which are useful later on.

Lemma 2.13

Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be linearly independent in an F-vector space V, and let $\mathbf{v}_{n+1}$ be a vector such that $\mathbf{v}_{n+1} \notin \operatorname{Span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$. Then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1}$ is linearly independent.

Proof. Suppose $\alpha_1 \mathbf{v}_1 + \cdots + \alpha_{n+1} \mathbf{v}_{n+1} = 0_V$ for some $\alpha_i \in F$. If $\alpha_{n+1} \neq 0$, then

$$\mathbf{v}_{n+1} = -\frac{1}{\alpha_{n+1}} \sum_{i=1}^{n} \alpha_i \mathbf{v}_i \in \operatorname{Span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n),$$

which is a contradiction.

Hence $\alpha_{n+1} = 0$ and so $\alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n = 0_V$. But $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent, so $\alpha_1 = \cdots = \alpha_n = 0$. □

After studying spans and linear independence, we can finally state what a basis is.

Definition 2.14 (Basis)

Let V be an F-vector space. A **basis** of V is a linearly independent spanning set of V.

If V has a finite basis, then we say V is a **finite dimensional** vector space.

Recall from the motivation that $\{1, x, x^2\}$ is indeed a basis for $V = \{ax^2 + bx + c : a, b, c \in \mathbb{R}\}$, and one can now verify it easily. Here are more examples:

Example 2.15

- Regard $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$ as a $\mathbb{Q}$-vector space. Then $\{1, \sqrt{2}\}$ is a basis.
- If $V = \mathbb{R}[x]$, the set of all real polynomials, then there is an infinite basis $\{1, x, x^2, \ldots\}$. The condition that we only use finitely many terms guarantees that the polynomials must have finite degree (which is good). Note in particular that V is **not** finite dimensional.

Besides this definition, we also have an equivalent definition of a basis:

Proposition 2.16

Let V be an F-vector space, and $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \subseteq V$. Then S is a basis of V if and only if every vector in V can be uniquely expressed as a linear combination of elements in S.

Proof. ($\Rightarrow$) Suppose S is a basis of V. Take $\mathbf{v} \in V$. Since V is spanned by S there exists $\alpha_1, \ldots, \alpha_n \in F$ such that $\mathbf{v} = \sum \alpha_i \mathbf{v}_i$. Now suppose there exists $\beta_1, \ldots, \beta_n \in F$ such that $\mathbf{v} = \sum \beta_i \mathbf{v}_i$. Then

$$\mathbf{v} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i = \sum_{i=1}^{n} \beta_i \mathbf{v}_i \Longrightarrow \sum_{i=1}^{n} (\alpha_i - \beta_i)\mathbf{v}_i = 0,$$

which implies $\alpha_i = \beta_i$ as S is linearly independent. Thus the choice of α_i's is unique.

($\Leftarrow$) Suppose conversely that for every $\mathbf{v} \in V$ there are unique α_i such that $\mathbf{v} = \sum \alpha_i \mathbf{v}_i$. Then it suffices to check:

- S is spanning: let $\mathbf{v} \in V$, then $\mathbf{v} = \sum \alpha_i \mathbf{v}_i \in \operatorname{Span}(S)$.
- S is linearly independent: since $0\mathbf{v}_1 + \cdots + 0\mathbf{v}_n = 0_V$, if $\sum \beta_i \mathbf{v}_i = 0$ then by uniqueness we must have $\beta_i = 0$.

Hence S is a basis for V. □

In the above proof, we can see that spanning ensures the **existence** of a linear combination, and linearly independence ensures the **uniqueness** of the linear combination. So the result is in fact quite natural.

Proposition 2.17

Let V be a non-trivial F-vector space and suppose V has a finite spanning set S. Then S contains a basis.

Proof. Consider $T \subseteq S$ such that T is linearly independent and for any linearly independent subset T' of S, we must have $|T'| \leq |T|$. This is possible since there exists $\mathbf{v} \in S$, and $\{\mathbf{v}\}$ is linearly independent.

We claim that T is spanning. Indeed, suppose not then there is a $\mathbf{v} \in S \setminus \operatorname{Span}(T)$. But by Lemma 2.13 $T \cup \{\mathbf{v}\}$ is linearly independent, contradiction since $T \cup \{\mathbf{v}\} \subseteq S$ is a larger linearly independent subset. □

As a final side-note: having a basis $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ of V is really cool because it means that to specify $\mathbf{v} \in V$, we only have to specify $\alpha_1, \ldots, \alpha_n \in F$, and let $\mathbf{v} = \sum \alpha_i \mathbf{e}_i$. We can even think of $\mathbf{v}$ as $(\alpha_1, \ldots, \alpha_n)$.

2.4 Dimension

Ideally, we would want to define the **dimension** as the number of vectors in a basis. However, we must first show that this is well-defined: it is certainly plausible that a vector space has a basis of size 7 as well as one of size 3.

The key step is as follows:

Lemma 2.18 (Steinitz Exchange Lemma)

Let V be an F-vector space. Take $S \subseteq V$ and suppose $\mathbf{u} \in \operatorname{Span}(S)$ but $\mathbf{u} \notin \operatorname{Span}(S \setminus \{\mathbf{v}\})$ for some $\mathbf{v} \in S$. Then $\operatorname{Span}(S) = \operatorname{Span}((S \setminus \{\mathbf{v}\}) \cup \{\mathbf{u}\})$.

Proof. Since $\mathbf{u} \in \operatorname{Span}(S)$, there exists $\alpha_1, \ldots, \alpha_n \in F$ and $\mathbf{v}_1, \ldots, \mathbf{v}_n \in S$ such that $\mathbf{u} = \sum \alpha_i \mathbf{v}_i$. Now $\mathbf{u} \notin \operatorname{Span}(S \setminus \{\mathbf{v}\})$ for some $\mathbf{v}$ implies that $\mathbf{v} = \mathbf{v}_i$ for some i, so WLOG assume $\mathbf{v} = \mathbf{v}_n$ and $\alpha_n \neq 0$. Then

$$\mathbf{v} = \mathbf{v}_n = \frac{1}{\alpha_n}(\mathbf{u} - \alpha_1 \mathbf{v}_1 - \cdots - \alpha_{n-1}\mathbf{v}_{n-1}).$$

Now if $\mathbf{w} \in \text{Span}((S \setminus \{\mathbf{v}\}) \cup \{\mathbf{u}\})$ then there exists $\beta_0, \beta_1, \ldots, \beta_m \in F$ and $\mathbf{u}_1, \ldots, \mathbf{u}_m \in S \setminus \{\mathbf{v}\}$ such that

$$\mathbf{w} = \beta_0 \mathbf{u} + \sum_{i=1}^{m} \beta_i \mathbf{u}_i = \beta_0 \sum_{i=1}^{n} \alpha_i \mathbf{v}_i + \sum_{i=1}^{m} \beta_i \mathbf{u}_i \in \text{Span}(S),$$

so $\text{Span}((S \setminus \{\mathbf{v}\}) \cup \{\mathbf{u}\}) \subseteq \text{Span}(S)$.

For the other direction, if $\mathbf{w} \in \text{Span}(S)$, $\mathbf{w}$ can be written as a linear combination of elements in S. Then we can replace $\mathbf{v}_n$ by $\frac{1}{\alpha_n}(\mathbf{u} - \alpha_1 \mathbf{v}_1 - \cdots - \alpha_{n-1}\mathbf{v}_{n-1})$, so it is a linear combination of elements in $(S \setminus \{\mathbf{v}\}) \cup \{\mathbf{u}\}$, so $\text{Span}(S) \subseteq \text{Span}((S \setminus \{\mathbf{v}\}) \cup \{\mathbf{u}\})$, which completes the proof. □

Remark. It is called the *exchange* lemma since it essentially states that we can exchange $\mathbf{v}$ for $\mathbf{u}$ and still preserve the span under certain conditions.

We now proceed to the main theorem of this section. It has multiple corollaries, including the fact that dimension is well-defined. Unfortunately, the proof is going to be slightly technical and notationally daunting, but the idea should be simple.

Theorem 2.19
Let V be a finite dimensional F-vector space. Let S, T be finite subsets of V. If S is linearly independent and T spans V then $|S| \leq |T|$.

Before the proof, this means

! Keypoint
Linear independent sets are at most as big as spanning sets.

Proof. Let $S = \{\mathbf{s}_1, \ldots, \mathbf{s}_m\}$ be linearly independent and $T = \{\mathbf{t}_1, \ldots, \mathbf{t}_n\}$ be spanning. Moreover let $T_0 = T$.

Since $\text{Span}(T_0) = V$ there is a minimal i such that $\mathbf{s}_1 \in \text{Span}(\{\mathbf{t}_1, \ldots, \mathbf{t}_i\})$. In particular, $\mathbf{s}_1 \notin \text{Span}(\{\mathbf{t}_1, \ldots, \mathbf{t}_{i-1}\})$. Thus by Steinitz Exchange Lemma (SEL),

$$\text{Span}(\mathbf{s}_1, \mathbf{t}_1, \ldots, \mathbf{t}_{i-1}) = \text{Span}(\mathbf{t}_1, \ldots, \mathbf{t}_i).$$

Now let $T_1 = \{\mathbf{s}_1, \mathbf{t}_1, \ldots, \mathbf{t}_{i-1}, \mathbf{t}_{i+1}, \ldots, \mathbf{t}_n\} = T_0 \setminus \{\mathbf{t}_i\} \cup \{\mathbf{s}_1\}$. Then $\text{Span}(T_1) = \text{Span}(T_0) = V$. We continue this process inductively: suppose for some j with $1 \leq j \leq m$ we have $T_j = \{\mathbf{s}_1, \ldots, \mathbf{s}_j, \mathbf{t}_{i_1}, \ldots, \mathbf{t}_{i_{n-j}}\}$, with $\text{Span}(T_j) = V$ and $\mathbf{t}_{i_k} \in T$. Then $\mathbf{s}_{j+1} \in \text{Span}(T_j)$ so there is a minimal k such that $\mathbf{s}_{j+1} \in \text{Span}(\mathbf{s}_1, \ldots, \mathbf{s}_j, \mathbf{t}_{i_1}, \ldots, \mathbf{t}_{i_k})$.

We let $T_{j+1} = T_j \setminus \{\mathbf{t}_{i_k}\} \cup \{\mathbf{s}_{j+1}\}$ and by SEL, $\text{Span}(T_{j+1}) = \text{Span}(T_j) = V$. By relabelling the elements of T_{j+1} we have a set of the form

$$T_{j+1} = \{\mathbf{s}_1, \ldots, \mathbf{s}_{j+1}, \mathbf{t}_{i_1}, \ldots, \mathbf{t}_{i_{n-(j+1)}}\},$$

so the process can be continued.

Now after j steps we have replaced j members of T with j members of S. We cannot run out of members of T before we run out of members of S, or else $T_k = \{\mathbf{s}_1, \ldots, \mathbf{s}_k\}$ for some k but $\text{Span}(T_k) = V$, so $\mathbf{s}_{k+1} \in \text{Span}(T_k)$, contradiction to S being linearly independent. Hence $m \leq n$. □

Behind the notational mess, the main idea in the proof is to replace elements of T one-by-one by elements of S, and to preserve the spanning property of T, where we used the Steinitz Exchange Lemma.

We finally have the following result we have promised a long time ago:

Corollary 2.20 (Dimension theorem for vector spaces)
Let V be a finite dimensional vector space, and S, T be bases of V. Then S and T are both finite and $|S| = |T|$.

Proof. Since V is finite dimensional, it has a finite basis B. Now as S, T are linearly independent, by Theorem 2.19 $|S| \leq |B|$ and $|T| \leq |B|$, so both sets are finite.

Now S is spanning and T is linearly independent, so $|T| \leq |S|$. Similarly $|S| \leq |T|$, so $|S| = |T|$. □

Remark. In fact, the theorem is true for an infinite basis as well if we interpret "the size of a basis" as "cardinality".

The dimension theorem, true to its name, allows us to define:

Definition 2.21
Let V be a finite dimensional vector space. The **dimension** of V, written $\dim V$, is the size of *any* basis of V.

For example,

$$V = \{ax^2 + bx + c : a, b, c \in \mathbb{R}\}$$

has dimension three, because $\{1, x, x^2\}$ is a basis. That's not the only basis: we could as well have written

$$\{a(x^2 - 4x) + b(x + 2) + c : a, b, c \in \mathbb{R}\}$$

and gotten the exact same vector space. But the beauty of the theorem is that no matter how we try to contrive the basis, we always will get **exactly three elements**. That's why it makes sense to say V has dimension three.

Besides the dimension theorem, we have multiple other corollaries:

Corollary 2.22
Suppose that $\dim V = n$. Then

(i) any spanning set of size n is a basis;

(ii) any linearly independent set of size n is a basis;

(iii) S is spanning if and only if it contains a basis;

(iv) S is linearly independent if and only if it is contained in a basis;

(v) any subset of V with size $> n$ is linearly dependent.

Proof. Most of the proofs are simple routine:

(i). Let T be spanning with size n. If T were linearly dependent, then there is some $\mathbf{t}_0, \ldots, \mathbf{t}_m \in T$ and $\alpha_1, \ldots, \alpha_m \in F$ such that $\mathbf{t}_0 = \sum \alpha_i \mathbf{t}_i$. Hence $\operatorname{Span}(T \setminus \{\mathbf{t}_0\}) = \operatorname{Span}(T) = V$, but $T \setminus \{\mathbf{t}_0\}$ has size $n - 1 < n$, contradicting Theorem 2.19.

(ii). Let T be linearly independent with size n, but $\operatorname{Span}(T) \neq V$. Then there exists $\mathbf{v} \notin \operatorname{Span}(T)$. By Lemma 2.13, $T \cup \{\mathbf{v}\}$ is linearly independent, contradiction.

(iii). ($\Rightarrow$) is proved in Proposition 2.17, and ($\Leftarrow$) is trivial.

(iv). ($\Rightarrow$) is by the proof of Theorem 2.19: pick a basis B of V, then there is some $B' \subseteq B$ with $|B'| = |T|$ such that $(B \setminus B') \cup T$ spans V, which is a basis by (i). ($\Leftarrow$) is trivial.

(v). Contrapositive of "linear independent $\Rightarrow$ size $\leq n$". □

2.5 More subspaces

We have shown that $U \cap W$ is a subspace of V while $U \cup W$ in general isn't. Turns out, we need the sum:

Definition 2.23
Let U, W be subspaces of V. Then the **sum** of U and V is

$$U + W = \{\mathbf{u} + \mathbf{w} : \mathbf{u} \in U, \mathbf{w} \in W\},$$

which is a subspace of V.

It is simple to check that it is indeed a subspace, so we omit it here. Note that $U, W \subseteq U + W$ as for any $\mathbf{u} \in U$, $\mathbf{u} = \mathbf{u} + 0 \in U + W$ and similarly for W.

Example 2.24
Take the vector space $\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$. We can consider it as a sum of its x-axis and y-axis:

$$X = \{(x, 0) : x \in \mathbb{R}\} \quad \text{and} \quad Y = \{(0, y) : y \in \mathbb{R}\}.$$

Then $\mathbb{R}^2 = X + Y$.

Remark. Extra side-note: for a vector space V, if $V = U + W$ and every element in V can be *uniquely* written as $\mathbf{u} + \mathbf{w}$ for some $\mathbf{u} \in U, \mathbf{w} \in W$, then the sum is also called a **direct sum**, and is denoted $V = U \oplus W$.

For instance, $\mathbb{R}^2 = X \oplus Y$ above and if $V = \{ax^2 + bx + c : a, b, c \in \mathbb{R}\}$ then we can write $V = x^2\mathbb{R} \oplus x\mathbb{R} \oplus \mathbb{R}$. This gives us a "top-down" way to break down vector spaces.

Proposition 2.25
Let V be an F-vector space and U, W be subspaces of V. Suppose we have $U = \text{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_s)$ and $W = \text{Span}(\mathbf{w}_1, \ldots, \mathbf{w}_r)$, then

$$U + W = \text{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_s, \mathbf{w}_1, \ldots, \mathbf{w}_r).$$

Proof. ($\subseteq$) Let $\mathbf{v} \in U + W$. Then $\mathbf{v} = \mathbf{u} + \mathbf{w}$ for some $\mathbf{u} \in U$ and $\mathbf{w} \in W$. Therefore there exists α_i and β_i such that

$$\mathbf{u} = \sum_{i=1}^{s} \alpha_i \mathbf{u}_i \quad \text{and} \quad \mathbf{w} = \sum_{i=1}^{r} \beta_i \mathbf{w}_i,$$

so $\mathbf{v} = \sum \alpha_i \mathbf{u}_i + \sum \beta_i \mathbf{w}_i \in \text{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_s, \mathbf{w}_1, \ldots, \mathbf{w}_r)$. The case of ($\supseteq$) is similar. □

The following theorem combines all the definition, and showcases the idea of "choosing the correct basis" well. Also note that Corollary 2.22(iv) is particularly useful here for extending a basis, which is a trick you should internalise.

Theorem 2.26
Let V be an F-vector space and u, W be subspaces of V. Then

$$\dim(U + W) = \dim U + \dim W - \dim(U \cap W).$$

Proof. Let $B_{U \cap W} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ be a basis of $U \cap W$. Regarding U as a vector space, $U \cap W \subseteq U$ is a subspace. Since $B_{U \cap W}$ is linearly independent it is contained in a basis

$$B_U = \{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{u}_{m+1}, \ldots, \mathbf{u}_r\},$$

and similarly we have a basis

$$B_W = \{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{w}_{m+1}, \ldots, \mathbf{w}_s\}$$

for W. We want to show that $\dim(U + W) = r + s - m$, so it suffices to prove that $B_U \cup B_W$ is a basis for $U + W$.

- $B_U \cup B_W$ is spanning: immediate by Proposition 2.25.
- $B_U \cup B_W$ is linearly independent: suppose we have a linear combination

$$\sum_{i=1}^{m} \lambda_i \mathbf{v}_i + \sum_{j=m+1}^{r} \mu_j \mathbf{u}_j + \sum_{k=m+1}^{s} \nu_k \mathbf{w}_k = 0_V$$

where $\lambda_i, \mu_j, \nu_k \in F$. Then

$$\sum \lambda_i \mathbf{v}_i + \sum \mu_j \mathbf{u}_j = -\sum \nu_k \mathbf{w}_k.$$

Since the left hand side is something in U and the right hand side is something in W, they both lie in $U \cap W$. Hence we can write the the left hand side as $\sum \beta_i \mathbf{v}_i$ for some $\beta_i \in F$ as it is in $\text{Span}(B_{U\cap W})$. Then

$$\sum \beta_i \mathbf{v}_i + \sum \nu_k \mathbf{w}_k = 0_V$$

and since $B_W = \{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{w}_{m+1}, \ldots, \mathbf{w}_s\}$ is a basis for W, we have $\beta_i = \nu_i = 0$. Therefore

$$\sum \lambda_i \mathbf{v}_i + \sum \mu_j \mathbf{u}_j = 0_V$$

and since $B_U = \{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{u}_{m+1}, \ldots, \mathbf{u}_r\}$ is a basis for U, we have $\lambda_i = \mu_i = 0$.

Thus $B_U \cup B_W$ is linearly independent.

This completes the proof as $|B_U \cup B_W| = r + s - m$. □

We end this section by providing a concrete example in computing the basis for $U \cap W, U, W$ and $U + W$.

Example 2.27

Let $V = \mathbb{R}^3, U = \{(x, y, z) \in V : x + y + z = 0\}$ and $W = \{(x, y, z) \in V : -x + 2y + z = 0\}$.

- For $U \cap W$, a vector $\mathbf{v} = (x, y, z)$ is in $U \cap W$ if and only if $x + y + z = 0$ and $-x + 2y + z = 0$, which reduces to $y = 2x$ and $z = -3x$. Hence $U \cap W = \{(x, 2x, -3x) : x \in \mathbb{R}\}$ and thus $\{(1, 2, -3)\}$ is a basis.
- A general vector in U is of the form $(a, b, -a-b)$ for $a, b \in \mathbb{R}$, so it is easy to see that $\{(1, 0, -1), (0, 1, -1)\}$ is a spanning set. As the two vectors are linearly independent, this is a basis for U.

  But now let's try to *exchange* one of the vectors in the basis by $(1, 2, -3)$, so our life will be easier in the case of $U + W$. Notice that $(1, 2, -3) \notin \text{Span}(1, 0, -1)$, so by SEL $\{(1, 0, -1), (1, 2, -3)\}$ is a basis too.
- Similarly, one can see that $\{(1, 0, 1), (1, 2, -3)\}$ is a basis for W.
- Finally, for $U + W$, by Proposition 2.25, $\{(1, 0, -1), (1, 0, 1), (1, 2, -3)\}$ is spanning. By Theorem 2.26 we already know $\dim(U + W) = 2 + 2 - 1 = 3$, so by Corollary 2.22(i) this is a basis.

Motivation

Notice that in the above example, if we simply used a non-intersecting basis for U and W, the spanning set for $U + W$ would have size 4, so we need extra work to remove one of the vectors and show that it is linearly independent or spanning. This explains why we exchanged the vector in the basis of U.

Remark. As a final side-note: the direct sum that we defined before is actually the notion of **internal direct sums**. The seemingly opposite notion, **external direct sum** is defined as

$$U \oplus W = \{(\mathbf{u}, \mathbf{w}) : \mathbf{u} \in U, \mathbf{w} \in W\}$$

where U and W are F-vector spaces. This creates a **new** vector space (addition and scalar multiplication are defined componentwise) instead of decomposing the given vector space, thus the name external. But one can show that they are actually isomorphic, so they deserve the same name and notation.

2.6 Rank of a matrix

The final subsection is a glimpse at the whole picture of linear transformations that we will soon cover, and possibly the most important concept in matrix algebra. We will define the (row and column) **rank** of a matrix, which represents a measure of "information content" of a matrix.

Motivation

Consider the system of equations

$$\begin{aligned} x + 2y &= 0 \\ 2x + 4y &= 0 \end{aligned}$$

From first sight we know that there are infinitely many solutions: the second equation is just a multiple of the first, so any point of the form $(-2y, y)$ works. There are multiple ways to think of this:

- The matrix $A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ has two linearly dependent rows, so although there are two equations, only one is useful, or in other words $\dim \text{Span}(\text{row vectors}) = 1$.
- If we regard A as a map $A : \mathbb{R}^2 \to \mathbb{R}^2$ which sends $\mathbf{x}$ to $A\mathbf{x}$, then the **image** of the map

$$\text{im } A = \left\{ \begin{pmatrix} x_1 + 2x_2 \\ 2x_1 + 4x_2 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} x \\ 2x \end{pmatrix} : x \in \mathbb{R} \right\}$$

  is a subspace of $\mathbb{R}^2$ with dimension one: it is just a line $y = 2x$ in the plane.

These two perspectives both agree to say that the matrix A has **(row) rank one**. We will later see how this relates to the system having infinitely many solutions.

Enough talking, here are the definitions:

Definition 2.28

Let A be an $m \times n$ matrix with entries in a field F. Define

- the **row space** of A, denoted by $\text{RSp}(A)$ as the span of the rows of A. This is a subspace of F^n;
- the **row rank** of A to be $\dim(\text{RSp}(A))$.

Similarly we can define the column space $\text{CSp}(A)$ of A which is a subspace of F^m and the column rank.

Example 2.29

Let $F = \mathbb{R}$ and $A = \begin{pmatrix} 3 & 1 & 2 \\ 0 & -1 & 1 \end{pmatrix}$. Then

$$\text{RSp}(A) = \text{Span}((3,1,2),(0,-1,1)) \quad \text{and} \quad \text{CSp}(A) = \text{Span}\left(\begin{pmatrix} 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right).$$

Now since $(3,1,2)$ and $(0,-1,1)$ are linearly independent, $\dim(\text{RSp}(A)) = 2$, while the set $\left\{ \begin{pmatrix} 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}$ is linearly dependent since $\begin{pmatrix} 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. So

$$\text{CSp}(A) = \text{Span}\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right)$$

and consequently $\dim(\text{CSp}(A)) = 2$.

To find the row rank of a matrix, we can adopt the following steps:

- Reduce A to row echelon form using row operations:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \xrightarrow{\text{row operations}} A_{\text{ech}} = \begin{pmatrix} 1 & * & * & * & \cdots \\ 0 & 0 & 1 & * & \cdots \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots \end{pmatrix}$$

- The non-zero rows of A_{ech} form a basis for $\text{RSp}(A)$ so the row rank is the number of non-zero rows in A_{ech}.

The reason why this works is based on the following lemma:

Lemma 2.30
Define A_{ech} as above, then (i) $\text{RSp}(A) = \text{RSp}(A_{\text{ech}})$ and (ii) the rows of A_{ech} are linearly independent.

Proof. To show (i), notice that to obtain A_{ech} from A the possible row operations are

$$\begin{cases} R_i \mapsto \lambda R_i & \lambda \in F,\ i \neq j \\ R_i \mapsto R_i + \lambda R_j & \lambda \in F \setminus \{0\} \\ R_i \leftrightarrow R_j & i \neq j \end{cases}$$

Let A' be obtained from A by one row operation. Then clearly every row of A' lies in $\text{RSp}(A)$, as the row operations that we perform are linear. Hence $\text{RSp}(A') \subseteq \text{RSp}(A)$. Moreover, every row operation is invertible:

$$\begin{cases} R_i \mapsto \lambda R_i & \text{has inverse} & R_i \mapsto \frac{1}{\lambda} R_i \\ R_i \mapsto R_i + \lambda R_j & \text{has inverse} & R_i \mapsto R_i - \lambda R_j \\ R_i \leftrightarrow R_j & \text{has inverse} & R_i \leftrightarrow R_j \end{cases}$$

which are all row operations too. Hence $\text{RSp}(A) \subseteq \text{RSp}(A')$ and $\text{RSp}(A) = \text{RSp}(A')$. After a finite number of row operations, we would obtain A_{ech}, thus $\text{RSp}(A) = \text{RSp}(A_{\text{ech}})$ as desired.

For (ii), let $i_1, \ldots, i_k$ be the positions of the leading entries in each row:

$$A_{\text{ech}} = \begin{pmatrix} \overset{i_1}{1} & * & \overset{i_2}{*} & * & \overset{\cdots}{\cdots} \\ 0 & 0 & 1 & * & \cdots \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots \end{pmatrix}$$

and let $\mathbf{r}_1, \ldots, \mathbf{r}_k$ denonte the non-zero row vectors of A_{ech}. Now suppose $\alpha_1 \mathbf{r}_1 + \cdots + \alpha_k \mathbf{r}_k = 0$ for scalars α_i.

Then the i_1-th entry of $\sum \alpha_i \mathbf{r}_i$ is $\alpha_1 \cdot 1 = \alpha_1$ since the entry in all other rows are 0. Hence $\alpha_1 = 0$ and $\alpha_1 \mathbf{r}_1 + \cdots + \alpha_k \mathbf{r}_k = \alpha_2 \mathbf{r}_2 + \cdots + \alpha_k \mathbf{r}_k$. Similarly, the i_2-th entry is α_2 so $\alpha_2 = 0$. By induction we can show that $\alpha_i = 0$ for all i, so the row vectors are linearly independent. □

In particular, (i) in the above lemma says that

! Keypoint
Row operations have no effect on the row space.

Remark. Note that the above argument still works if the leading entries in A_{ech} are not 1s.

A neat application of the above procedure is that we can now compute dimension of a span more easily:

Example 2.31

Consider

$$W = \text{Span}((-1,1,0,1),(2,3,1,0),(0,1,2,3)) \subseteq \mathbb{R}^4.$$

We will try to find the dimension of W. Viewing the vectors as rows of a matrix, i.e. let $A = \begin{pmatrix} -1 & 1 & 0 & 1 \\ 2 & 3 & 1 & 0 \\ 0 & 1 & 2 & 3 \end{pmatrix}$, then W is the row space of this matrix and the dimension is the row rank. By row operations,

$$A \xrightarrow{R_2 \mapsto R_2+2R_1} \begin{pmatrix} -1 & 1 & 0 & 1 \\ 0 & 5 & 1 & 2 \\ 0 & 1 & 2 & 3 \end{pmatrix} \xrightarrow{R_3 \mapsto 5R_3} \begin{pmatrix} -1 & 1 & 0 & 1 \\ 0 & 5 & 1 & 2 \\ 0 & 5 & 10 & 15 \end{pmatrix} \xrightarrow{R_3 \mapsto R_3-R_2} \begin{pmatrix} -1 & 1 & 0 & 1 \\ 0 & 5 & 1 & 2 \\ 0 & 0 & 9 & 13 \end{pmatrix} = A_{\text{ech}}$$

which has three non-zero rows, so $\dim W = \dim(\text{RSp}(A)) = 3$.

To find the column rank (as well as a basis of the column space) of A, we simply have to find the row rank of A^T, since the columns of A are the rows of A^T.

The interesting result is as below:

Theorem 2.32 (Row rank = Column rank)

For any matrix A, the row rank of A is equal to the column rank of A.

Proof. Let $A = (a_{ij}) \in M_{m\times n}(F)$, and let the rows of A be $\mathbf{r}_1, \ldots, \mathbf{r}_m$, so $\mathbf{r}_i = (a_{i1}, \ldots, a_{in})$. Similarly let the columns of A be $\mathbf{c}_1, \ldots, \mathbf{c}_m$. Assume A has row rank k.

Then $\text{RSp}(A)$ has a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$. Every row would then be a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_k$, so

$$\mathbf{r}_i = \alpha_{i1}\mathbf{v}_1 + \alpha_{i2}\mathbf{v}_2 \cdots + \alpha_{ik}\mathbf{v}_k$$

for some α_{ij}. Now let $\mathbf{v}_i = (b_{i1}, b_{i2}, \ldots, b_{in})$. Then by looking at the j-th coordinate in the above linear combination,

$$a_{ij} = \alpha_{i1}b_{1j} + \alpha_{i2}b_{2j} + \cdots + \alpha_{ik}b_{kj}.$$

Now

$$\mathbf{c}_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} = \begin{pmatrix} \alpha_{11}b_{1j} + \alpha_{12}b_{2j} + \cdots + \alpha_{1k}b_{kj} \\ \alpha_{21}b_{1j} + \alpha_{22}b_{2j} + \cdots + \alpha_{2k}b_{kj} \\ \vdots \\ \alpha_{m1}b_{1j} + \alpha_{m2}b_{2j} + \cdots + \alpha_{mk}b_{kj} \end{pmatrix} = b_{1j}\begin{pmatrix} \alpha_{11} \\ \alpha_{21} \\ \vdots \\ \alpha_{m1} \end{pmatrix} + b_{2j}\begin{pmatrix} \alpha_{12} \\ \alpha_{22} \\ \vdots \\ \alpha_{m2} \end{pmatrix} + \cdots + b_{kj}\begin{pmatrix} \alpha_{1k} \\ \alpha_{2k} \\ \vdots \\ \alpha_{mk} \end{pmatrix},$$

i.e. $\mathbf{c}_j$ is a linear combination of the vectors

$$\begin{pmatrix} \alpha_{11} \\ \alpha_{21} \\ \vdots \\ \alpha_{m1} \end{pmatrix}, \begin{pmatrix} \alpha_{12} \\ \alpha_{22} \\ \vdots \\ \alpha_{m2} \end{pmatrix}, \ldots, \begin{pmatrix} \alpha_{1k} \\ \alpha_{2k} \\ \vdots \\ \alpha_{mk} \end{pmatrix}.$$

Hence $\text{CSp}(A)$ is spanned by these vectors, thus $\dim(\text{CSp}(A)) \leq k = \dim(\text{RSp}(A))$. By the same argument, $\dim(\text{CSp}(A^T)) \leq \dim(\text{RSp}(A^T))$, but the column rank of A^T is the row rank of A and vice versa. Hence $\dim(\text{RSp}(A)) = \dim(\text{CSp}(A))$, as needed. □

As we now know that row and column rank are the same, it makes sense to define:

Definition 2.33 (Rank)

Let A be a matrix. The **rank** of A, written $\operatorname{rank}(A)$, is the row rank of A.

This completely explains our first perspective as mentioned in the motivation. For the second perspective, we will elaborate in a more detailed manner once we introduced linear transformations.

Finally, we have a proposition for the special case when A is a square matrix to end this section.

Proposition 2.34

Let A be an $n \times n$ matrix with entries in F. Then the following are equivalent:

(i) $\operatorname{rank}(A) = n$, otherwise known as "$A$ has *full rank*".

(ii) the rows of A form a basis for F^n.

(iii) the columns of A form a basis for F^n.

(iv) A is invertible.

Proof. ((i) $\Leftrightarrow$ (ii)): We simply note that

$$\operatorname{rank}(A) = n \iff \dim(\operatorname{RSp}(A)) = n \iff \operatorname{RSp}(A) = F^n$$

which is equivalent to the desired statement. ((i) $\Leftrightarrow$ (iii)) is the same but with columns.

((i) $\Leftrightarrow$ (iv)): Notice that $\operatorname{rank}(A) = n$ if and only if the row echelon form of A looks like

$$A_{\text{ech}} = \begin{pmatrix} 1 & & & \\ & 1 & & * \\ & & 1 & \\ & 0 & & \ddots \\ & & & & 1 \end{pmatrix}$$

Since all of the $*$ entries can be eliminated using row operations, A is reducible to I by row operations. Thus by Theorem 1.18 this is equivalent to A being invertible. □

3 Linear Transformations

Here comes possibly the most important concept in linear algebra: **linear transformations**.

3.1 Definitions

In mathematics, apart from studying objects, we would like to study functions between objects as well. In particular, we would like to study **structure-preserving** functions. This motivates the definition of linear transformations:

Definition 3.1 (Linear transformations)

Let V, W be F-vector spaces and $T : V \to W$ be a function from V to W. T is a **linear transformation** if

- T preserves addition: $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2)$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$;
- T preserves scalar multiplication: $T(\lambda \mathbf{v}) = \lambda T(\mathbf{v})$ for all $\mathbf{v} \in V, \lambda \in F$.

If this map is a bijection, it is an **isomorphism**. We then say V and W are **isomorphic** and write $V \cong W$.

The definition is natural as the structures determining whether a set is a vector space are exactly the two which are preserved. Note particularly that $T(0_V) = 0_W$. Also notice that the two conditions can be combined to the single requirement that

$$T(\lambda \mathbf{v}_1 + \mu \mathbf{v}_2) = \lambda T(\mathbf{v}_1) + \mu T(\mathbf{v}_2)$$

for all $\mathbf{v}_1, \mathbf{v}_2 \in V$ and $\lambda, \mu \in F$.

Example 3.2

Here are a myriad of examples:

- The identity map $\mathrm{id} : V \to V$ is clearly a linear transformation and isomorphism.
- The map $\mathbb{R}^3 \to \mathbb{R}$ via $(x, y, z) \mapsto 4x + 2y + z$ is a linear transformation.
- If A is an $n \times m$ matrix with entries in F, then the map $F^m \to F^n$ by $\mathbf{v} \mapsto A\mathbf{v}$ is a linear transformation.
- The map $\mathbb{R}[x] \to \mathbb{R}[x]$ via differention, i.e. $f(x) \mapsto f'(x)$ is a linear transformation.
- Let V be the set of real polynomials of degree at most 2, i.e. $V = \{ax^2 + bx + c : a, b, c \in \mathbb{R}\}$. Then
  - the map $\mathbb{R}^3 \to V$ by $(a, b, c) \mapsto ax^2 + bx + c$ is an isomorphism;
  - the map $V \to \mathbb{R}$ by $ax^2 + bx + c \mapsto 9a + 3b + c$ is a linear transformation, which can be described as "evaluation at 3".
- Consider the map of complex conjugation, i.e. $T : \mathbb{C} \to \mathbb{C}$ via $z \mapsto \overline{z}$. Then
  - if we view $\mathbb{C}$ as an $\mathbb{R}$-vector space, T is a linear transformation since $\lambda = \overline{\lambda}$ for all $\lambda \in \mathbb{R}$;
  - if we view $\mathbb{C}$ as a $\mathbb{C}$-vector space, then T is **not** a linear transformation.

Caution: Despite the name "linear", many "linear" functions are not linear transformations. For instance, $T : \mathbb{R} \to \mathbb{R}$ defined by $x \mapsto x + 1$ is **not** a linear transformation since neither conditions are satisfied.

We also have an equivalent formulation for isomorphisms:

Proposition 3.3

A linear transformation is an isomorphism if and only if it has an linear inverse function.

Proof. ($\Leftarrow$) is simple since a function having an inverse is equivalent to it being bijective.

($\Rightarrow$) Suppose $T : V \to W$ is a bijective linear transformation, then it has an inverse $T^{-1} : W \to V$. We want to show that it is linear. Let $\mathbf{w}_1, \mathbf{w}_2 \in W$ and $\lambda, \mu \in F$. Then

$$T(T^{-1}(\lambda\mathbf{w}_1 + \mu\mathbf{w}_2)) = \lambda\mathbf{w}_1 + \mu\mathbf{w}_2 = \lambda T(T^{-1}(\mathbf{w}_1)) + \mu T(T^{-1}(\mathbf{w}_2)) = T(\lambda T^{-1}(\mathbf{w}_1) + \mu T^{-1}(\mathbf{w}_2)).$$

Since T is injective, $T^{-1}(\lambda\mathbf{w}_1 + \mu\mathbf{w}_2) = \lambda T^{-1}(\mathbf{w}_1) + \mu T^{-1}(\mathbf{w}_2)$. So T^{-1} is linear. □

3.2 What is a matrix?

Using linear transformations, one can finally explain matrices in a morally correct manner: it is a way of representing a linear transformation in terms of bases. At the end we will show that in fact, *all* linear transformations come from matrices.

To reach there, we need to establish some basic ideas of what is happening. We will first show:

! Keypoint

To define a linear transformation, it suffices to define its values on a basis.

In addition, a nice corollary we will show later is the following fact:

Motivation

At the end of Section 2.3, we mentioned how given a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$, we can just think of a vector $\mathbf{v} \in V$ as a list of scalars $(\alpha_1, \dots, \alpha_n)$ since $\mathbf{v} = \sum \alpha_i \mathbf{e}_i$ is a unique representation of $\mathbf{v}$. But notice that the list of scalars is actually just an element in F^n. So essentially this means:

$$V \cong F^{\dim V}$$

for any finite-dimensional vector space V.

So let's begin. The following example explains the keypoint:

Example 3.4

Consider a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ such that $T(1,0) = (1,-1,2)$ and $T(0,1) = (0,1,3)$. Then we can **extend linearly**, in the following sense:

Since $\{(1,0),(0,1)\}$ is a basis of $\mathbb{R}^2$, if we have $(a,b) \in \mathbb{R}^2$ we can write $(a,b) = a(1,0) + b(0,1)$. Then by the definition of linear transformations,

$$T(a,b) = T(a(1,0) + b(0,1)) = aT(1,0) + bT(0,1) = a(1,-1,2) + b(0,1,3) = (a, -a+b, 2a+3b).$$

So the whole map is actually determined.

Proposition 3.5

Let V and W be F-vector spaces and $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis of V. Suppose we fix some $\mathbf{w}_1, \dots, \mathbf{w}_n$. Then there is a unique linear transformation $T : V \to W$ such that $T(\mathbf{e}_i) = \mathbf{w}_i$ for all i.

Proof. Let $\mathbf{v} \in V$. Since $\mathbf{e}_1, \dots, \mathbf{e}_n$ is a basis, we can write $\mathbf{v} = \sum \lambda_i \mathbf{e}_i$ uniquely. Then we define $T : V \to W$ by

$$T(\mathbf{v}) = \lambda_1 \mathbf{w}_1 + \cdots + \lambda_n \mathbf{w}_n.$$

Uniqueness is immediate since $T(\sum \lambda_i \mathbf{e}_i) = \sum \lambda_i T(\mathbf{e}_i) = \sum \lambda_i \mathbf{w}_i$, so this is the only way to define T. It remains to check that T is indeed linear. Let $\mathbf{u} = \sum \lambda_i \mathbf{e}_i$ and $\mathbf{v} = \sum \mu_i \mathbf{e}_i$. Then we have

$$T(\alpha\mathbf{u} + \beta\mathbf{v}) = T\left(\sum(\alpha\lambda_i)\mathbf{e}_i + \sum(\beta\mu_i)\mathbf{e_i}\right) = \sum(\alpha\lambda_i + \beta\mu_i)\mathbf{w}_i = \alpha T(\mathbf{u}) + \beta T(\mathbf{v}),$$

which completes the proof. □

This implies our motivation:

Corollary 3.6 (n-dimensional vector spaces are isomorphic)
If V is a n-dimensional vector space, then $V \cong F^n$.

Proof. Let $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis of V. As in Proposition 3.5, choose $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n$ as the standard basis of F^n, i.e. $\mathbf{w}_i = (0, \dots, 0, 1, 0, \dots, 0)^T$ where the 1 is at the i-th position. Then there is a unique linear transformation $T : V \to F^n$ such that $T(\mathbf{e}_i) = \mathbf{w}_i$, and if $\mathbf{v} = \sum \lambda_i \mathbf{e}_i$ we have

$$T(\mathbf{v}) = \sum_{i=1}^{n} \lambda_i \mathbf{w}_i = (\lambda_1, \lambda_2, \dots, \lambda_n)^T.$$

This is clearly bijective, so $V \cong F^n$ as needed. □

Remark. Important: You could technically say that all finite-dimensional vector spaces are just F^n and that no other space is worth caring about. But this seems kind of rude, as spaces often are more than just tuples:

$$V = \{ax^2 + bx + c : a, b, c \in \mathbb{R}\}$$

is a space of polynomials, and so it has some "essence" to it that you'd lose if you compressed it into (a, b, c).

Moreover, a lot of spaces, like the set of vectors (x, y, z) with $x + y + z = 0$, do not have an obvious choice of basis. Thus to cast such a space into F^n would require you to make arbitrary decisions.

We also define a new notation for convenience:

Definition 3.7
Let V be an n-dimensional vector space with $B = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ as a basis and $\mathbf{v} \in V$ has $\mathbf{v} = \sum \lambda_i \mathbf{e}_i$. Then the **vector of v w.r.t** B is

$$[\mathbf{v}]_B = (\lambda_1, \lambda_2, \dots, \lambda_n)^T,$$

which is well-defined as, again, the linear combination is unique.

so that the isomorphism from V to F^n is just $T(\mathbf{v}) = [\mathbf{v}]_B$.

Motivation
Consider two vector spaces V and W, with bases $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ and $\{\mathbf{f}_1, \dots, \mathbf{f}_m\}$ respectively.

Recall in Proposition 3.5 that if we fix some $\mathbf{w}_i$, there will be a unique linear map sending $\mathbf{e}_i$ to $\mathbf{w}_i$. Being more concrete, we can write $\mathbf{w}_i$ in a basis of W as well, then there is a unique map $T : V \to W$ which satisfies

$$T(\mathbf{e}_1) = a_{11}\mathbf{f}_1 + a_{21}\mathbf{f}_2 + \cdots + a_{m1}\mathbf{f}_m$$

and similarly for $\mathbf{e}_2, \dots, \mathbf{e}_n$ for some a_{ij}. So telling you a_{ij} would tell you everything you need to know about T. We can then define "the matrix for T" to be

$$A = \begin{pmatrix} | & | & & | \\ T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n) \\ | & | & & | \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

So all of this means:

! Keypoint
To define a linear transformation, it suffices to give a matrix.

Writing the above more rigorously, we have

Corollary 3.8 (Linear transformations are matrices)

Let V, W be finite-dimensional vector spaces over F with bases $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ and $\{\mathbf{f}_1, \dots, \mathbf{f}_m\}$ respectively, and let $\mathcal{L}(V, W)$ denote the set of linear transformations from V to W. Then there is a bijection

$$M_{m\times n}(F) \longleftrightarrow \mathcal{L}(V, W)$$

sending $A = (a_{ij})$ to the unique linear map $T(\mathbf{e}_i) = \sum a_{ji}\mathbf{f}_j$.

Proof. The proof is actually just the argument in the motivation. Nonetheless, a more fancy way of saying the same thing is to represent the result in a diagram:

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ {\scriptstyle [-]_B}\uparrow\downarrow & & \uparrow\downarrow{\scriptstyle [-]_C} \\ F^n & \dashrightarrow & F^m \end{array}$$

Suppose we are given a linear transformation $T : V \to W$. Then we can define a map $F^n \to F^m$ by following the diagram around. But such a map must be a matrix transformation. If it is represented by a matrix A, then $A[\mathbf{v}]_B = [T(\mathbf{v})]_C$. We can then calculate A by figuring out its columns. We write

$$T(\mathbf{e}_i) = a_{1i}\mathbf{f}_1 + a_{2i}\mathbf{f}_2 + \cdots + a_{mi}\mathbf{f}_m, \tag{1}$$

and let $\mathbf{c}_i$ be the canonical basis of F^n. Then

$$[T(\mathbf{e}_i)]_C = A[\mathbf{e}_i]_B = A\mathbf{c}_i$$

which is exactly the i-th column of A. Such a construction is unique since there is only one T satisfying (1). □

Definition 3.9

Given a linear transformation $T : V \to W$, the matrix corresponding to it by the above construction is called the **matrix of T** w.r.t basis $B = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ and $C = \{\mathbf{f}_1, \dots, \mathbf{f}_m\}$. This is denoted by ${}_C[T]_B$.

By the diagram, we then have ${}_C[T]_B[\mathbf{v}]_B = [T(\mathbf{v})]_C$. If $V = W$ and $B = C$ we sometimes write the matrix simply as $[T]_B$, then $[T]_B[\mathbf{v}]_B = [T(\mathbf{v})]_B$.

Caution: Despite of the notation, when computing ${}_C[T]_B$, we put in elements of B and express the output as linear combinations of elements of C.

Example 3.10

Here is a concrete example on how to find A. Consider $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 - x_2 \\ x_1 + 2x_2 \end{pmatrix}$.

- Take $E = \{\mathbf{e}_1, \mathbf{e}_2\}$, the standard basis of $\mathbb{R}^2$. Then as

  $$T(\mathbf{e}_1) = \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 2\mathbf{e}_1 + 1\mathbf{e}_2 \quad \text{and} \quad T(\mathbf{e}_2) = \begin{pmatrix} -1 \\ 2 \end{pmatrix} = -1\mathbf{e}_1 + 2\mathbf{e}_2,$$

  we conclude that $[T]_E = \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}$.

- Similarly, if we take the basis $B = \{\mathbf{e}_1 + \mathbf{e}_2, \mathbf{e}_2\}$, one can show $[T]_B = \begin{pmatrix} 1 & -1 \\ 2 & 3 \end{pmatrix}$ and ${}_B[T]_E = \begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix}$.

Example 3.11

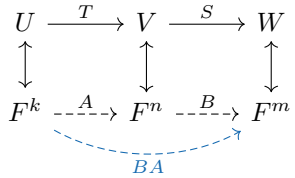
We again let $V = \{ax^2 + bx + c : a, b, c \in \mathbb{R}\}$ and $B = \{1, x, x^2\}$ be a basis for it.

Consider the "evaluation at 3" map, which is a linear transformation $T : V \to \mathbb{R}$. Choose $C = \{1\}$ as the basis of $\mathbb{R}$, then T is represented by

$$_C[T]_B = \begin{pmatrix} 1 & 3 & 9 \end{pmatrix}$$

with the columns corresponding to $T(1), T(x)$ and $T(x^2)$.

From here, we can actually work out what it means to mulitply two matrices. Suppose we have picked a basis for three spaces U, V, W. Then given map $T : U \to V$ and $S : V \to W$, we can consider their composition $S \circ T$, i.e.



where A represents T and B represents S. Then BA represents the composition $S \circ T$:

Proposition 3.12 (Matrix multiplication is composition of linear maps)

Let U, V, W be finite-dimensional vector spaces over F with bases $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$, $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ and $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$ respectively. If $T : U \to V$ and $S : V \to W$ are linear transformations, then $S \circ T$ is linear and

$$_{(\mathbf{w}_i)}[S \circ T]_{(\mathbf{u}_i)} = {}_{(\mathbf{w}_i)}[S]_{(\mathbf{v}_i)} \cdot {}_{(\mathbf{v}_i)}[T]_{(\mathbf{u}_i)}.$$

Proof. Verifying that $S \circ T$ is linear is straightforward. In the following we simply denote $[T]$ as the matrix of T. Writing $S \circ T(\mathbf{u}_i)$ as a linear combination of $\mathbf{w}_i$, we then have

$$S \circ T(\mathbf{u}_i) = S\left(\sum_k [T]_{ki}\mathbf{v}_k\right) = \sum_k [T]_{ki} S(\mathbf{v}_k) = \sum_k [T]_{ki} \sum_j [S]_{jk}\mathbf{w}_j = \sum_j \left(\sum_k [S]_{jk}[T]_{ki}\right)\mathbf{w}_j = \sum_j [S][T]_{ji}\mathbf{w}_j,$$

which precisely implies that $[S][T]$ is the matrix representing $S \circ T$. □

In particular, since function composition is associative, it follows that matrix multiplication is as well.

Motivation

This in fact has a deeper consequence: the correlation between matrices and linear transformations means that we can define concepts like the determinant or the trace of a matrix, both in terms of an **"intrinsic" map** $T : V \to W$ and in terms of the entries of the matrix. Since the map T itself doesn't refer to any basis, the abstract definition will imply that the **numerical definition doesn't depend on the choice of basis**.

To sum up this section and answer the title,

! Keypoint

A matrix is the laziest possible way to specify a linear transformation.

3.3 Image and Kernel

To study functions, it is natural to look at the set of zeroes, or the set of solutions of the function. At the same time, it is useful to consider the possible outputs of the function, so that we can predict how will the function behave. Over linear algebra (and in general any algebraic structure), this is the concept of **kernel** and **image**.

Definition 3.13 (Image and kernel)

Let $T : U \to V$ be a linear transformation. Then the **image** of T is

$$\operatorname{im} T = \{T(\mathbf{u}) : \mathbf{u} \in U\}.$$

The **kernel** of T is

$$\ker T = \{\mathbf{u} : T(\mathbf{u}) = 0_U\}.$$

It is easy to see that they are subspaces of V and U respectively.

Example 3.14

Let $A \in M_{m\times n}(F)$ and $T : F^n \to F^m$ be the linear map $\mathbf{v} \mapsto A\mathbf{v}$. Consider the system of linear equations

$$\sum_{j=1}^{m} A_{ij}x_j = b_i, \qquad 1 \le i \le n.$$

- The system has a solution iff $(b_1, \dots, b_n) \in \operatorname{im} T$.
- The kernel of T contains all solutions to $\sum_j A_{ij}x_j = 0$.

Image and kernel have a lot of properties; one of them is the following fact that appears often.

Proposition 3.15

Let $T : V \to W$ be a linear transformation and $\mathbf{v}_1, \mathbf{v}_2 \in V$. Then $T(\mathbf{v}_1) = T(\mathbf{v}_2)$ iff $\mathbf{v}_1 - \mathbf{v}_2 \in \ker T$.

Proof. This is just unfolding the definition:

$$T(\mathbf{v}_1) = T(\mathbf{v}_2) \iff T(\mathbf{v}_1 - \mathbf{v}_2) = 0 \iff \mathbf{v}_1 - \mathbf{v}_2 \in \ker T,$$

as T is a linear transformation. □

Although I advertise the result as appearing often, the more useful corollary is the following:

Corollary 3.16

A linear transformation T is injective if and only if $\ker T = 0$.

Remark. For experts again: this result is standard and can actually be generalised to any group.

For images, a nice result is as follows, which would explain a terminology later on:

Proposition 3.17

Let $T : V \to W$ be a linear transformation. Suppose that $B = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis for V. Then $\operatorname{im} T = \operatorname{Span}(T(\mathbf{v}_1), \dots, T(\mathbf{v}_n))$.

Proof. $(\supseteq)$ clearly holds. For $(\subseteq)$, let $\mathbf{w} \in \operatorname{im} T$, then $\mathbf{w} = T(\mathbf{v})$ for some $\mathbf{v}$. Write $\mathbf{v} = \sum \lambda_i \mathbf{v}_i$, then

$$\mathbf{w} = T\left(\sum \lambda_i \mathbf{v}_i\right) = \sum \lambda_i T(\mathbf{v}_i) \in \operatorname{Span}(T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)),$$

which completes the proof. □

Does this look familiar? Recall from the last section that

$$[T(\mathbf{v}_i)]_B = [T]_B[\mathbf{v}_i]_B = [T]_B\mathbf{e}_i = i\text{-th column of } [T]_B$$

where $\mathbf{e}_i$ is the canonical basis of F^n. So the above proposition can be restated to say

! Keypoint

$\operatorname{im} T$ is the column space of the matrix representing V after fixing a basis.

Hence, we sometimes refer to $\dim(\operatorname{im} T)$ as the **rank** of T, as the dimension is exactly the column rank of $[T]_B$.

We now proceed to a super important result – not because of its actual statement, but because of how it will give you the right picture in your head of how a linear transformation is supposed to look.

Theorem 3.18 (Rank-nullity theorem)

Let $T : V \to W$ be a linear transformation. Then

$$\dim(\operatorname{im} T) + \dim(\ker T) = \dim V.$$

Proof. Let $\{\mathbf{f}_1, \ldots, \mathbf{f}_k\}$ be a basis of $\operatorname{im} T$, and note there exists $\mathbf{e}_i \in V$ such that $T(\mathbf{e}_i) = \mathbf{f}_i$. Now let $\{\mathbf{e}_{k+1}, \ldots, \mathbf{e}_n\}$ be a basis for $\ker T$. We claim that $B = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a basis for V (notice that $\mathbf{e}_i \neq \mathbf{e}_j$ even if $i \leq k$ and $j > k$, for if not then $\mathbf{f}_i = T(\mathbf{e}_i) = T(\mathbf{e}_j) = 0$, which is impossible).

- B is spanning: Let $\mathbf{v} \in V$. Since $T(\mathbf{v}) \in \operatorname{im} T$ there exists $\lambda_i \in F$ such that

$$T(\mathbf{v}) = \sum_{i=1}^{k} \lambda_i \mathbf{f}_i = T\left(\sum_{i=1}^{k} \lambda_i \mathbf{e}_i\right),$$

  so $\mathbf{v} - \sum \lambda_i \mathbf{e}_i \in \ker T$. Thus there exists $\mu_j \in F$ such that

$$\mathbf{v} - \sum_{i=1}^{k} \lambda_i \mathbf{e}_i = \sum_{j=k+1}^{n} \mu_j \mathbf{e}_j \implies \mathbf{v} = \lambda_1 \mathbf{e}_1 + \ldots + \lambda_k \mathbf{e}_k + \mu_{k+1} \mathbf{e}_{k+1} + \ldots + \mu_n \mathbf{e}_n \in \operatorname{Span}(B).$$

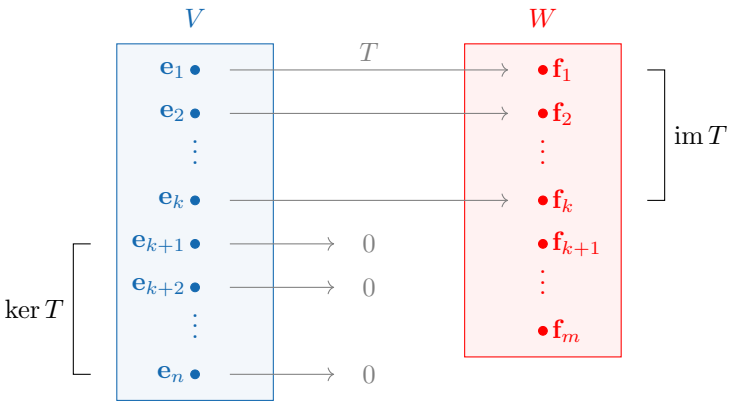
- B is linearly independent: Suppose $\sum \lambda_i \mathbf{e}_i + \sum \mu_j \mathbf{e}_j = 0_V$ for some $\lambda_i, \mu_j \in F$. Then by applying T,

$$0_W = T\left(\sum_{i=1}^{k} \lambda_i \mathbf{e}_i + \sum_{j=k+1}^{n} \mu_j \mathbf{e}_j\right) = \sum_{i=1}^{k} \lambda_i T(\mathbf{e}_i) + \sum_{j=k+1}^{n} \mu_j T(\mathbf{e}_j) = \sum_{i=1}^{k} \lambda_i \mathbf{f}_i,$$

  so $\lambda_i = 0$ since $\mathbf{f}_i$ forms a basis. Then $\sum \mu_j \mathbf{e}_j = 0_V$, so $\mu_j = 0$ too since $\mathbf{e}_j$, $k+1 \leq j \leq n$ forms a basis.

This implies the result since $\dim(\operatorname{im} T) = k, \dim(\ker T) = n - k$ and $\dim V = n$. □

From the proof, we can then draw such a picture visualising a linear transformation:



In particular, for $T : V \to W$, one can write $V = \ker T \oplus V'$, so that T annihilates its kernel while sending V' to an isomorphic copy of itself, $\operatorname{im} T$, in W.

The theorem also, finally, explains the second interpretation in the motivation back in 2.6. Recall that we discussed how the system has infinitely many solutions. In general, consider a system of linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

The system is called **homogeneous** if $b_1 = b_2 = \cdots = b_m = 0$. In this case we can write the system in matrix form, i.e. $A\mathbf{x} = 0$, and consider the map $A : F^n \to F^m$ via $\mathbf{x} \mapsto A\mathbf{x}$. Then the set of solutions is $\ker A$ and by rank-nullity theorem,

$$\dim(\ker A) = n - \dim(\operatorname{im} A) = n - \operatorname{rank} A.$$

Hence,

- if $\operatorname{rank} A = n$, there is only one solution, i.e. the trivial solution;
- if $\operatorname{rank} A < n$, then $\dim(\ker A) \geq 1$ so if F is infinite then there are infinitely many solutions.

Remark. Final side-note: For non-homogeneous systems, the result is similar assuming $\mathbf{b} \in \mathrm{CSp}(A)$. If $\mathbf{b} \notin \mathrm{CSp}(A)$, then the system is inconsistent.

3.4 Change of basis

Recall that given a linear transformation $T : V \to W$ and bases B for V, C for W, we obtain a matrix ${}_C[T]_B$:

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ {\scriptstyle [-]_B}\uparrow\downarrow & & \uparrow\downarrow{\scriptstyle [-]_C} \\ F^n & \xrightarrow{{}_C[T]_B} & F^m \end{array}$$

Focus on the case $V = W$ but the bases are different. Then these will give rise to two different isomorphisms to F^n, and the two bases can be related by a matrix map P again, i.e.

$$\begin{array}{ccc} V & \xrightarrow{\mathrm{id}} & V \\ {\scriptstyle [-]_B}\uparrow\downarrow & & \uparrow\downarrow{\scriptstyle [-]_C} \\ F^n & \xrightarrow{P} & F^n \end{array}$$

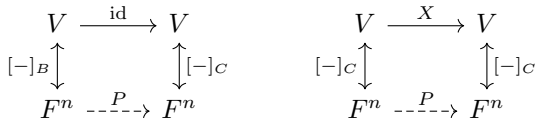
Definition 3.19 (Change of basis matrix)
The matrix P is defined to be the **change of basis matrix** from B to C.

We then have from definition that $P = {}_C[\mathrm{id}]_B$ and $P[\mathbf{v}]_B = [\mathbf{v}]_C$ for all $\mathbf{v} \in V$. Writing the matrix out explicitly, it can be expressed as follows: if $B = \{\mathbf{e}_i\}$ and $C = \{\mathbf{f}_i\}$, then we can write $\mathbf{e}_i = \sum_j \lambda_{ji}\mathbf{f}_j$ and $P := (\lambda_{ij})$, so the j-th column of P is $[\mathbf{e}_j]_C$. Another interpretation of P is as follows:

Proposition 3.20
$P = [X]_C$ where $X : V \to V$ is the unique linear transformation such that $X(\mathbf{f_i}) = \mathbf{e}_i$.

Proof. The j-th columnn of $[X]_C$ is $[X(\mathbf{f}_j)]_C = [\mathbf{e}_j]_C$, which is exactly the j-th column of P. □

In comparison of the figures, this means:



- if we fix the elements being mapped (i.e. choose the map to be id), we get P;
- but if we fix the basis and choose X to send $\mathbf{f}_i$ to $\mathbf{e_i}$, we also get P.

This is unfortunately very confusing:

Caution: In the second interpretation, X **takes elements in** C **to** B, but P is called the change of basis matrix **from** B **to** C since $P[\mathbf{v}]_B = [\mathbf{v}]_C$, i.e. it maps vectors written in basis B to them written in basis C.

This is inevitable because both interpretations have their use, so it is actually just a preference to name P as the change of basis matrix from B to C or vice versa.

Anyways, change of basis matrices have a lot of properties. The most important ones being how we can transform matrices written in one basis to another:

Proposition 3.21

Let V, B, C, P be as above, then

(i) P is invertible, and P^{-1} is the change of basis matrix from C to B;

(ii) let $T : V \to V$ be a linear transformation, then $[T]_C = P[T]_B P^{-1}$.

Proof. The proof can be done directly, or using the diagrams. Here we will do it directly:

(i). Let Q be the change of basis matrix from C to B. Then

$$QP[\mathbf{v}]_B = Q[\mathbf{v}]_C = [\mathbf{v}]_B.$$

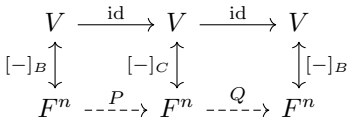
As $\mathbf{v}$ ranges over all of V, $[\mathbf{v}]_B$ ranges over all of F^n, so $QP\mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in F^n$, i.e. $QP = I$. Thus P is invertible with inverse Q.

(ii). We have $[T]_C[\mathbf{v}]_C = [T(\mathbf{v})]_C$, and

$$(P[T]_B P^{-1})[\mathbf{v}]_C = (P[T]_B P^{-1})P[\mathbf{v}]_B = P[T]_B[\mathbf{v}]_B = P[T(\mathbf{v})]_B = [T(\mathbf{v})]_C.$$

Again as this ranges over all $\mathbf{v}$, we have the desired result. □

Another method of using the diagrams are probably easier. For instance, in (i), we consider the diagram



By Proposition 3.12, QP is the matrix representing the map $\mathrm{id} \circ \mathrm{id} = \mathrm{id}$ with respect to the same basis B. Then it is clear that $QP = I$ as we clearly have $[\mathrm{id}]_B = I$. Similarly one can prove (ii) by a diagram and matrix multiplication.

Example 3.22

Let $V = \mathbb{R}^2$. Take bases $B = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ and $E = \{\mathbf{e}_1 + \mathbf{e}_2\}$ as the standard basis. Then

- the change of basis matrix P from B to E is ${}_E[\mathrm{id}]_B$. We have

$$\mathrm{id} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \mathbf{e}_1 + \mathbf{e}_2 \quad \text{and} \quad \mathrm{id} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \mathbf{e}_1 + 2\mathbf{e}_2,$$

  so $P = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ by taking the columns as the scalars;

- thus the change of basis matrix Q from E to B is the inverse of P, which is $\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$.

Remark. Suppose we have three bases B, C, D. Again, by Proposition 3.12 we have

$${}_D[\mathrm{id}]_B = {}_D[\mathrm{id}]_C \, {}_C[\mathrm{id}]_B.$$

In particular, by choosing C as the standard basis E, we obtain

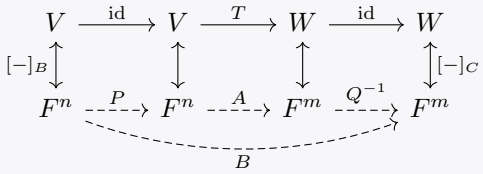
$${}_D[\mathrm{id}]_B = {}_D[\mathrm{id}]_E \, {}_E[\mathrm{id}]_B = ({}_E[\mathrm{id}]_D)^{-1} \, {}_E[\mathrm{id}]_B$$

while we notice from the above example that ${}_E[\mathrm{id}]_-$ is usually much easier to compute. So this provides a faster method to calculate an arbitrary change of basis matrix.

We conclude the section by giving a final definition of a more generalised notion:

Motivation

Instead of composing two change of basis maps, we can also slip in a linear transformation in the middle. Indeed, consider a linear transformation $T : V \to W$, and consider the diagram



where P and Q are change of basis matrices, and A represents T. Then by matrix multiplication again, $B = Q^{-1}AP$ is the matrix representing T w.r.t bases B and C. Note in particular that A and B actually both represents T, just under different bases.

Hence, we might define the following:

Definition 3.23 (Equivalent matrices)

We say $A, B \in M_{m\times n}(F)$ are **equivalent** if there are invertible matrices $P \in M_{m\times m}(F)$ and $Q \in M_{n\times n}(F)$ such that $A = QBP^{-1}$.

Of course, the name has to make sense – and the way it makes sense is exactly given by the motivation:

! Keypoint

Two matrices are equivalent if and only if they represent the same linear map w.r.t different bases.

This marks the end of our discussion for now.

4 Duality ⋆

You may have learned in high school that given a matrix

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix},$$

the trace is the sum along the diagonal $a+d$ and the determinant is $ad-bc$. As mentioned earlier, one could define such words using both the entries of a matrix, or intrinsically in terms of an actual linear map T.

While we might be tempted to naively define the trace of T to be the trace of the matrix $[T]_B$ for some basis, this raises the question: Why would these random formulas somehow not depend on the choice of the basis?

In this chapter, we will answer the question by giving an intrinsic definition of $\operatorname{tr}(T)$. This will give a coordinate-free definition as wanted. To do this, we will need to introduce two new constructions: the **tensor product** $V \otimes W$ and the **dual space** $V^\vee$.

4.1 Tensor product

Let's start off with a motivation:

Motivation

We know that $\dim(V \oplus W) = \dim V + \dim W$, even though $V \oplus W$ looks like $V \times W$. What if we wanted a real "product" of spaces, with their dimensions multiplied?

For example, again consider

$$V = \{ax^2 + bx + c : a, b, c \in \mathbb{R}\}$$

with another space

$$W = \{dy + e : d, e, \in \mathbb{R}\}.$$

If we take the direct sum, we will get some rather unnatural vector space of dimension five: the element can be thought of as pairs $(ax^2+bx+c, dy+e)$. But if we **multiply the elements** in V and W, we get a vector space which would contain elements like $4x^2y + 5xy + y + 3$. In particular, a basis of it is

$$\{x^2y, x^2, xy, x, y, 1\}$$

and thus has dimension 6.

For this, we resort to the **tensor product**, which does exactly this, except the "multiplication" is done by a scary symbol $\otimes$. We first give the rigorous definition:

Definition 4.1 (Tensor product)

Let V and W be F-vector spaces. The **tensor product** $V \otimes_F W$ is defined as the **span** of the elements of the form $\mathbf{v} \otimes \mathbf{w}$, subject to the relations

- distributive in V: $(\mathbf{v}_1 + \mathbf{v}_2) \otimes \mathbf{w} = \mathbf{v}_1 \otimes \mathbf{w} + \mathbf{v}_2 \otimes \mathbf{w}$;
- distributive in W: $\mathbf{v} \otimes (\mathbf{w}_1 + \mathbf{w}_2) = \mathbf{v} \otimes \mathbf{w}_1 + \mathbf{v} \otimes \mathbf{w}_2$;
- $(\lambda \mathbf{v}) \otimes \mathbf{w} = \mathbf{v} \otimes (\lambda \mathbf{w})$.

with scalar multiplication defined by $\lambda \cdot (\mathbf{v} \otimes \mathbf{w}) = (\lambda \mathbf{v}) \otimes \mathbf{w} = \mathbf{v} \otimes (\lambda \mathbf{w})$.

Remark. Some may also say that the $\otimes$ is **bilinear** by the given relations, since it is linear in both arguments. Also, if the context is clear, we will omit the subscript and simply write $V \otimes W$.

To understand this daunting definition, let's go back to the example:

Example 4.2

Following the motivation, the above element might be written as

$$4x^2 \otimes y + 5x \otimes y + 1 \otimes y + 3 \otimes 1,$$

and we can think of $\otimes$ as a "wall" separating the elements in the two vector spaces.

- There's no need to do everything in terms of just the monomials, so we can write

  $$(x+1) \otimes (y+1)$$

  and it should be possible to expand it as $x \otimes y + 1 \otimes y + x \otimes 1 + 1 \otimes 1$. This explains the distributivity.
- There should also be no distinction between writing $4x^2 \otimes y$ and $x^2 \otimes 4y$, or even $2x^2 \otimes 2y$, so the scalars should be free to float around. This explains the final condition.

Let's look at some other examples:

Example 4.3

- If V is any F-vector space, then $V \otimes_F F \cong V$. In this case, $\mathbf{v} \otimes f$ is just the scalar multiplication $f\mathbf{v}$.
- Consider $\mathbb{R}[x]$ and $\mathbb{R}[y]$ as $\mathbb{R}$-vector spaces. Then

  $$\mathbb{R}[x] \otimes_{\mathbb{R}} \mathbb{R}[y] \cong \mathbb{R}[x, y],$$

  i.e. tensor product of polynomials in x with polynomials in y turns out to just be two-variable polynomials (this should not be surprising).

Caution: The elements of $V \otimes W$ really are **sums** of $\mathbf{v} \otimes \mathbf{w}$: from the previous example, not every polynomial in $\mathbb{R}[x, y]$ can be written as a polynomial in x times a polynomial in y (i.e. in the form $f(x) \otimes g(y)$), but they can all be written as sums of $x^a \otimes y^b$.

As suggested by the examples, the basis of $V \otimes_F W$ is literally the "product" of the bases of V and W. In particular, this fulfills our desire that $\dim(V \otimes W) = \dim V \cdot \dim W$.

Proposition 4.4 (Basis of $V \otimes_F W$)

Let V, W be finite-dimensional F-vector spaces. If $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a basis of V and $\{\mathbf{f}_1, \ldots, \mathbf{f}_m\}$ is a basis of W, then $\mathbf{e}_i \otimes \mathbf{f}_j$ over all i, j is a basis for $V \otimes_F W$.

Surprisingly, the proof is a little bit out of reach by the knowledge we have developed until now. We will leave out the linear independence part, but nonetheless, it is quite easy to show that it is spanning:

Proof of spanning. Let $S = \{\mathbf{e}_i \otimes \mathbf{f}_j : 1 \le i \le n, 1 \le i \le m\}$. Clearly $\operatorname{Span} S \subseteq V \otimes_F W$.

Now let

$$\lambda_1(\mathbf{v}_1 \otimes \mathbf{w}_1) + \lambda_2(\mathbf{v}_2 \otimes \mathbf{w}_2) + \cdots + \lambda_k(\mathbf{v}_k \otimes \mathbf{w}_k) \in V \otimes_F W.$$

Then since $\{\mathbf{e}_i\}$ and $\{\mathbf{f}_i\}$ are respectively bases of V and W, we can write each $\mathbf{v}_a$ and $\mathbf{w}_a$ as linear combinations of elements of their corresponding basis, i.e.

$$\mathbf{v}_a = \sum_{i=1}^{n} \alpha_{ai}\mathbf{e}_i \qquad \text{and} \qquad \mathbf{w}_a = \sum_{j=1}^{m} \beta_{aj}\mathbf{f}_j,$$

which gives $\mathbf{v}_a \otimes \mathbf{w}_a = \sum\sum \alpha_{ai}\beta_{aj}(\mathbf{e}_i \otimes \mathbf{f}_j)$. Hence the sum $\sum \lambda_i(\mathbf{v}_i \otimes \mathbf{w}_i)$ would also be a linear combination of the elements $\mathbf{e}_i \otimes \mathbf{f}_j$, i.e. $V \otimes_F W \subseteq \operatorname{Span} S$, as desired. □

Example 4.5 (Explicit computation)

Let V have basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ and W have basis $\{\mathbf{f}_1, \mathbf{f}_2\}$. Let $\mathbf{v} = 3\mathbf{e}_1 + 4\mathbf{e}_2 \in V$ and $\mathbf{w} = 5\mathbf{f}_1 + 6\mathbf{f}_2 \in W$. Let's write $\mathbf{v} \otimes \mathbf{w}$ in the basis for $V \otimes_F W$:

$$\begin{aligned}\mathbf{v} \otimes \mathbf{w} &= (3\mathbf{e}_1 + 4\mathbf{e}_2) \otimes (5\mathbf{f}_1 + 6\mathbf{f}_2) \\ &= (3\mathbf{e}_1) \otimes (5\mathbf{f}_1) + (4\mathbf{e}_2) \otimes (5\mathbf{f}_1) + (3\mathbf{e}_1) \otimes (6\mathbf{f}_2) + (4\mathbf{e}_2) \otimes (6\mathbf{f}_2) \\ &= 15(\mathbf{e}_1 \otimes \mathbf{f}_1) + 20(\mathbf{e}_2 \otimes \mathbf{f}_1) + 18(\mathbf{e}_1 \otimes \mathbf{f}_2) + 24(\mathbf{e}_2 \otimes \mathbf{f}_2).\end{aligned}$$

So you can see why tensor product is a nice "product" to consider if we are interested in $V \times W$, but with more structure than just in the way defined by a direct sum.

Remark. Notice that the above essentially implies that we can associate $\mathbf{v} \otimes \mathbf{w}$ to a matrix (a_{ij}) where a_{ij} is the scalar coefficient of $\mathbf{e}_i \otimes \mathbf{f}_j$. So one might actually "construct" a tensor product by firstly writing

$$\begin{aligned}[\mathbf{v}]_{(\mathbf{e}_i)} &= (v_1, v_2, \dots, v_n)^T \\ [\mathbf{w}]_{(\mathbf{f}_i)} &= (w_1, w_2, \dots, w_m)^T,\end{aligned}$$

then saying that

$$\mathbf{v} \otimes \mathbf{w} = \begin{pmatrix} v_1w_1 & v_1w_2 & \cdots & v_1w_m \\ v_2w_1 & v_2w_2 & \cdots & v_2w_m \\ \vdots & \vdots & \ddots & \vdots \\ v_nw_1 & v_nw_2 & \cdots & v_nw_m \end{pmatrix}$$

Traditionally, this is called the **outer product** or **Kronecker product** of $\mathbf{v}$ and $\mathbf{w}$.

Caution: One last warning: much like the Cartesian product $A \times B$ of sets, you can tensor together any two vector spaces V and W over the same field F; but **the relationship between V and W is completely irrelavant**.

Thinking $\otimes$ as a "wall", it can only pass scalars but otherwise keep the elements of V and W separated. So for example $\mathbf{v} \otimes \mathbf{w} \neq \mathbf{w} \otimes \mathbf{v}$ in general even if $V = W$, just like $(x, y) \neq (y, x)$ in the set A^2. In particular, we have

$$\mathbb{R}[x] \otimes_{\mathbb{R}} \mathbb{R}[x] \cong \mathbb{R}[x, y],$$

since elements in the two copies of $\mathbb{R}[x]$ do not communicate.

We conclude the section by giving what's known as the "**universal property** of tensor product"; in fact, it is often more common to see the following as the definition of tensor product, since it determines $V \otimes W$ **up to unique isomorphism** (i.e. if another space P satisfies the conditions, then $V \otimes W \cong P$ through a unique isomorphism).

We firstly clarify the definition made before:

Definition 4.6

Let V, W, P be F-vector spaces. We say that a map $f : V \times W \to P$ is **bilinear** if for all $\mathbf{v}, \mathbf{w}, \lambda$,

$$\begin{aligned}f(\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2, \mathbf{w}) &= \lambda_1 f(\mathbf{v}_1, \mathbf{w}) + \lambda_2 f(\mathbf{v}_2, \mathbf{w}) \\ f(\mathbf{v}, \lambda_1\mathbf{w}_1 + \lambda_2\mathbf{w}_2) &= \lambda_1 f(\mathbf{v}, \mathbf{w}_1) + \lambda_2 f(\mathbf{v}, \mathbf{w}_2).\end{aligned}$$

which translates into the fact that f is linear in both arguments.

Therefore, for instance, the map $\otimes : V \times W \to V \otimes W$ defined by

$$(\mathbf{v}, \mathbf{w}) \mapsto \mathbf{v} \otimes \mathbf{w}$$

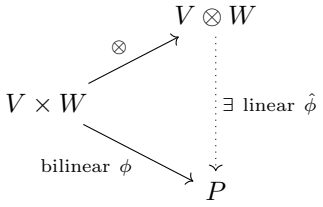
is bilinear by definition.

We bravely proceed to the main result now (but without proof, as unfortunately the proof involves too much technical machinery and hence it is omitted here):

Theorem 4.7 (Universal property of tensor products)

Let V, W, P be F-vector spaces. Then for any bilinear map $\phi : V \times W \to P$, there is a unique linear transformation $\hat{\phi} : V \otimes W \to P$, such that $\hat{\phi}(\mathbf{v} \otimes \mathbf{w}) = \phi(\mathbf{v}, \mathbf{w})$ for all $\mathbf{v} \in V$ and $\mathbf{w} \in W$.

Representing everything in a diagram, we have



so the universal property says that we can always find $\hat{\phi}$ so that the diagram commutes.

Motivation

Final motivation: this section showcases well the theme of **abstraction**. To define a tensor product $V \otimes W$, we might say, in ascending order of abstraction,

(i) $V \otimes W$ is the span of matrices constructed from the Kronecker products;

(ii) $V \otimes W$ is the span of $\mathbf{v} \otimes \mathbf{w}$ where $\otimes : V \times W \to V \otimes W$ is bilinear;

(iii) $V \otimes W$ is a vector space satisfying the universal property of tensor products.

Although abstract definitions might be hard to grasp, it often provides a better understanding of the object: if we had simply defined tensor products via a formula involving a matrix, it would be impossible to see that there is an underlying universal property that governs the whole behaviour.

This theme will appear frequently from now on, for instance when defining the determinant.

4.2 Dual space

We will now move on to the second topic, which is the notion of a **dual space**.

Definition 4.8 (Dual space)

Let V be an F-vector space. The **dual space** of V, denoted by $V^\vee$, is defined as the vector space whose elements are linear transformations from V to F, i.e.

$$V^\vee = \mathcal{L}(V, F) = \{T : V \to F : T \text{ linear}\}.$$

It should be routine to check that this is indeed a vector space, as its name suggests.

Example 4.9

Here are some basic examples:

- If $V = \mathbb{R}^3$ and $T : V \to \mathbb{R}$ sends $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ to $x_1 - x_3$, then $T \in V^\vee$.
- If $V = \mathbb{Q}[\sqrt{2}] := \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$ and $T : V \to \mathbb{Q}$ sends $a + b\sqrt{2}$ to a, then $T \in V^\vee$.
- Let $V = \mathbb{R}^X$ with elements as real-valued functions $f : X \to \mathbb{R}$. Then for any fixed x, the **evaluation map** $T : V \to \mathbb{R}$ at x, defined by $f \mapsto f(x)$ is in $V^\vee$.

Now let's try to find a basis for $V^\vee$, by computing an explicit example. Suppose $V = \mathbb{R}^3$ and we just use the canonical basis B. Then we can think of elements of V as column vectors, like

$$\mathbf{v} = \begin{pmatrix} 2 \\ 5 \\ 9 \end{pmatrix} \in V.$$

Now recall that a linear map $T \in V^\vee$ can be represented by a 1×3 matrix, i.e. a row vector, like

$$[T]_B = \begin{pmatrix} 3 & 4 & 5 \end{pmatrix}.$$

Then

$$T(\mathbf{v}) = \begin{pmatrix} 3 & 4 & 5 \end{pmatrix} \begin{pmatrix} 2 \\ 5 \\ 9 \end{pmatrix} = 71.$$

More precisely, we have from previous sections that **to specify a linear map $V \to F$, we only have to specify the matrix $[T]_B$, i.e. where each basis elements of V goes**. In the above example, T sends $\mathbf{e}_1$ to 3, $\mathbf{e}_2$ to 4 and $\mathbf{e}_3$ to 5, so f sends $2\mathbf{e}_1 + 5\mathbf{e}_2 + 9\mathbf{e}_3$ to $2 \cdot 3 + 5 \cdot 4 + 9 \cdot 5 = 71$. Hence a natural basis is the linear maps represented by $\begin{pmatrix} 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$.

This is made precise by:

Proposition 4.10 (The dual basis for $V^\vee$)

Let V be a finite-dimensional F-vector space with basis $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$. For each i, consider the linear transformations $\mathbf{e}_i^\vee : V \to F$ defined by

$$\mathbf{e}_i^\vee(\mathbf{e}_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

Then $\{\mathbf{e}_1^\vee, \mathbf{e}_2^\vee, \ldots, \mathbf{e}_n^\vee\}$ is a basis of $V^\vee$, called the **dual basis** of $V^\vee$.

Proof. Since linear maps are characterised by the values on a basis, there exists a unique choice of $\mathbf{e}_i^\vee \in V^\vee$ for all i, so they are at least defined. We now show that they form a basis.

Suppose $T \in V^\vee$. We have

$$T = \sum_{i=1}^n \lambda_i \mathbf{e}_i^\vee \iff T(\mathbf{e}_j) = \sum_{i=1}^n \lambda_i \mathbf{e}_i^\vee(\mathbf{e}_j) \tag{2}$$

for all j, since again linear maps are fixed by their image of a basis. But

$$\sum_{i=1}^n \lambda_i \mathbf{e}_i^\vee(\mathbf{e}_j) = \lambda_1 \mathbf{e}_1^\vee(\mathbf{e}_j) + \lambda_2 \mathbf{e}_2^\vee(\mathbf{e}_j) + \cdots + \lambda_n \mathbf{e}_n^\vee(\mathbf{e}_j) = \lambda_j,$$

so (2) is again equivalent to $\lambda_j = T(\mathbf{e}_j)$. Hence there exists a unique way to write any $T \in V^\vee$ as a linear combination of $\mathbf{e}_i^\vee$, thus $\{\mathbf{e}_i^\vee\}$ is a basis by Proposition 2.16. □

Example 4.11

Let's translate the above example in this notation. We have $T = 3\mathbf{e}_1^\vee + 4\mathbf{e}_2^\vee + 5\mathbf{e}_3^\vee$, since for instance

$$\begin{aligned} T(\mathbf{e}_1) &= (3\mathbf{e}_1^\vee + 4\mathbf{e}_2^\vee + 5\mathbf{e}_3^\vee)(\mathbf{e}_1) \\ &= 3\mathbf{e}_1^\vee(\mathbf{e}_1) + 4\mathbf{e}_2^\vee(\mathbf{e}_1) + 5\mathbf{e}_3^\vee(\mathbf{e}_1) \\ &= 3 \cdot 1 + 4 \cdot 0 + 5 \cdot 0 = 3, \end{aligned}$$

and similarly for $\mathbf{e}_2$ and $\mathbf{e}_3$, which is exactly what we wanted

You might also be inclined to point out that $V \cong V^\vee$ at this point, since there is an obvious isomorphism $\mathbf{e}_i \mapsto \mathbf{e}_i^\vee$. In other words, this map might be called "rotating the column vector by 90°". The issue is that this isomorphism depends very much on which basis we choose.

However, it is indeed true that V and $V^\vee$ are isomorphic for finite-dimensional V, but this is simple as we have from the above proposition that

Corollary 4.12

If V is finite dimensional, then $\dim V = \dim V^\vee$.

simply because of the fact that their basis have the same size. We furthermore know that **any** two F-vector spaces with the same dimension are isomorphic. In light of this, the fact that $V \cong V^\vee$ is not particularly impressive.

Remark. In general, this is **not** true for infinite-dimensional vector spaces, but in that case we still have

$$\dim V^\vee \geq \dim V,$$

since we gave a set of $\dim V$ linear independent vectors in the proposition (assuming that the dimension for an infinite-dimensional vector space makes sense).

4.3 The trace

We are finally ready to define the trace. But before that, we need a useful result first. The goal is to prove:

! Keypoint

If V and W are finite-dimensional F-vector spaces, then $V^\vee \otimes W$ represents linear maps $V \to W$.

In other words, we will try to find an isomorphism between $V^\vee \otimes W$ and $\mathcal{L}(V, W)$.

Remark. We didn't actually define $\mathcal{L}(V, W)$ formally, but it is often denoted by $\mathrm{Hom}(V, W)$ too. (The "Hom" stands for homomorphism.)

Motivation

The intuition is as follows: suppose V is three-dimensional and W is five-dimensional. Then the linear maps $V \to W$ can be thought as a 5×3 array of numbers. These maps form a vector space $\mathcal{L}(V, W)$, which will have dimension 15. But just saying "F^{15}" is not really that satisfying (what would the basis be in F^{15}?).

To do better, we want an isomorphism that still preserves some structure, which is precisely $V^\vee \otimes W \cong \mathcal{L}(V, W)$.

The actual isomorphism might be terrifying at first sight, so let's compile an example first:

Example 4.13

Firstly, how do we interpret an element of $V^\vee \otimes W$ as a map $V \to W$? For concreteness, suppose V has a basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and W has a basis $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \mathbf{f}_4, \mathbf{f}_5\}$. Consider an element of $V^\vee \otimes W$, say

$$\mathbf{e}_1^\vee \otimes (\mathbf{f}_2 + 2\mathbf{f}_4) + 4\mathbf{e}_2^\vee \otimes \mathbf{f}_5.$$

We want to interpret this element as a function $V \to W$, so given a $\mathbf{v} \in V$ we want to output an element of W. There is really only one way of doing this: feed in $\mathbf{v}$ into the $V^\vee$ elements on the left. That is, take the map

$$\mathbf{v} \mapsto \mathbf{e}_1^\vee(\mathbf{v}) \cdot (\mathbf{f}_2 + 2\mathbf{f}_4) + 4\mathbf{e}_2^\vee(\mathbf{v}) \cdot \mathbf{f}_5.$$

Since $\mathbf{e}_i^\vee$ are linear transformations $V \to F$, $\mathbf{e}_i^\vee(\mathbf{v})$ is indeed a scalar so the above expression makes sense.

So there is a natural way to interpret any element $\theta_1 \otimes \mathbf{w}_1 + \cdots + \theta_m \otimes \mathbf{w}_m \in V^\vee \otimes W$ as a linear map $V \to W$. The claim is that in fact, every linear map $V \to W$ has such an interpretation.

Writing down something for the general case, we have:

Theorem 4.14 ($V^\vee \otimes W$ is linear maps $V \to W$)
Let V and W be finite-dimensional vector spaces. The map

$$\Psi : V^\vee \otimes W \to \mathcal{L}(V, W)$$

defined via sending $\theta_1 \otimes \mathbf{w}_1 + \cdots + \theta_m \otimes \mathbf{w}_m$ to the linear map

$$\mathbf{v} \mapsto \theta_1(\mathbf{v})\mathbf{w}_1 + \cdots + \theta_m(\mathbf{v})\mathbf{w}_m$$

is an isomorphism of vector spaces, i.e. every linear map $V \to W$ can be uniquely represented as an element of $V^\vee \otimes W$ in this way.

Proof. This looks intimidating, but it's actually not that difficult. Ψ is clearly a linear transformation, so it suffices to show that it is bijective. We proceed in two steps:

- We firstly check that $\dim(V^\vee \otimes W) = \dim(\mathcal{L}(V, W))$. Indeed, $V^\vee \otimes W$ has dimension $\dim V \cdot \dim W$ since $\dim V^\vee = \dim V$. But by viewing $\mathcal{L}(V, W)$ as $\dim V \cdot \dim W$ matrices, it too has dimension $\dim V \cdot \dim W$.
- Now we show that Ψ is surjective. Take any $T : V \to W$. Suppose V has basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ and $T(\mathbf{e}_i) = \mathbf{w}_i$. Then the element
$$\sum_{i=1}^{n} \mathbf{e}_i^\vee \otimes \mathbf{w}_i = \mathbf{e}_1^\vee \otimes \mathbf{w}_1 + \mathbf{e}_2^\vee \otimes \mathbf{w}_2 + \cdots + \mathbf{e}_n^\vee \otimes \mathbf{w}_n$$
is mapped to T under Ψ, since the image of this element agrees with T on the basis elements $\mathbf{e}_i$.

Finally, by the rank-nullity theorem,

$$\dim(\operatorname{im}\Psi) + \dim(\ker\Psi) = \dim(V^\vee \otimes W).$$

Yet Ψ is surjective, so $\operatorname{im}\Psi = \mathcal{L}(V, W)$. By the first step, this implies $\dim(\ker\Psi) = 0$, i.e. $\ker\Psi = \{0_{V^\vee \otimes W}\}$. Hence Ψ is injective too, and thus an isomorphism. □

The above is perhaps a bit dense, so here is a concrete example:

Example 4.15
Let $V = \mathbb{R}^2$ and take a basis $B = \{\mathbf{e}_1, \mathbf{e}_2\}$ of V. Define $T : V \to V$ by

$$[T]_B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

Then we have

$$\Psi(\mathbf{e}_1^\vee \otimes \mathbf{e}_1 + 2\mathbf{e}_2^\vee \otimes \mathbf{e}_1 + 3\mathbf{e}_1^\vee \otimes \mathbf{e}_2 + 4\mathbf{e}_2^\vee \otimes \mathbf{e}_2) = T.$$

The beauty is that the definition of Ψ is **basis-free**, i.e. even if we change the basis, although the above expression will look completely different, **the actual element in $V^\vee \otimes V$ doesn't change**. (Even though we did choose a basis in proving that it works.)

We are now ready to give the definition of a trace. Recall that a square matrix T can be thought of a map $T : V \to V$. According to the above theorem,

$$\mathcal{L}(V, V) \cong V^\vee \otimes V,$$

so every map $V \to V$ can be thought of as an element of $V^\vee \otimes V$.

Then:

Definition 4.16 (Trace)

Define the **evaluation map** $\mathrm{ev} : V^\vee \otimes V \to F$ by "collapsing" each tensor (and extending linearly), i.e. $f \otimes v \mapsto f(v)$. Consider the composed map

$$\mathrm{tr} : \mathcal{L}(V,V) \xrightarrow{\cong} V^\vee \otimes V \xrightarrow{\mathrm{ev}} F.$$

The **trace** of a linear map $T : V \to V$ is then defined as the image $\mathrm{tr}(T)$ under the above composed map.

We are often taught that the trace of a matrix is the sum of its diagonal entries. This is explained by the following example:

Example 4.17

Continuing the last example, we know T is represented by

$$\mathbf{e}_1^\vee \otimes \mathbf{e}_1 + 2\mathbf{e}_2^\vee \otimes \mathbf{e}_1 + 3\mathbf{e}_1^\vee \otimes \mathbf{e}_2 + 4\mathbf{e}_2^\vee \otimes \mathbf{e}_2$$

in $V^\vee \otimes V$. Hence under the evaluation map,

$$\mathrm{tr}\, T = \mathbf{e}_1^\vee(\mathbf{e}_1) + 2\mathbf{e}_2^\vee(\mathbf{e}_1) + 3\mathbf{e}_1^\vee(\mathbf{e}_2) + 4\mathbf{e}_2^\vee(\mathbf{e}_2) = 1 + 0 + 0 + 4 = 5.$$

And that is why the trace is the sum of the diagonal entries.

Again, this definition of trace is basis-free: even if we choose another basis $\mathbf{f}_i$, we still know that it is the sum of the diagonal entries, since in that case $\mathbf{f}_i^\vee(\mathbf{f}_i) = 1$ and $\mathbf{f}_i^\vee(\mathbf{f}_j) = 0$ for $i \neq j$.

We conclude this section by proving a well-known result:

Proposition 4.18

Let $T : V \to W$ and $S : W \to V$ be linear transformations between finite-dimensional vector spaces V and W. Then

$$\mathrm{tr}(T \circ S) = \mathrm{tr}(S \circ T).$$

Proof. Consider two tensors $f \otimes \mathbf{v} \in W^\vee \otimes V$ and $g \otimes \mathbf{w} \in V^\vee \otimes W$. Recall the map Ψ from Theorem 4.14, then $\Psi(f \otimes \mathbf{v}) : W \to V$ and $\Psi(g \otimes \mathbf{w}) : V \to W$. For arbitrary $\mathbf{u} \in V$ we have

$$\Psi(f \otimes \mathbf{v}) \circ \Psi(g \otimes \mathbf{w})(\mathbf{u}) = \Psi(f \otimes \mathbf{v})(g(\mathbf{u})\mathbf{w}) = f(g(\mathbf{u})\mathbf{w})\mathbf{v} = f(\mathbf{w})g(\mathbf{u})\mathbf{v} = \Psi(f(\mathbf{w}) \cdot g \otimes \mathbf{v})(\mathbf{u}),$$

Hence $\Psi(f \otimes \mathbf{v}) \circ \Psi(g \otimes \mathbf{w}) = \Psi(f(\mathbf{w}) \cdot g \otimes \mathbf{v})$. By taking trace, this means

$$\mathrm{tr}(\Psi(f \otimes \mathbf{v}) \circ \Psi(g \otimes \mathbf{w})) = \mathrm{tr}(\Psi(f(\mathbf{w}) \cdot g \otimes \mathbf{v})) = f(\mathbf{w})\,\mathrm{ev}(g \otimes \mathbf{v}) = f(\mathbf{w})g(\mathbf{v})$$

which is clearly symmetric if we swap the order of composition.

Now, as T and S can be represented by a linear combination of elements of the form $g_i \otimes \mathbf{w}_i \in V^\vee \otimes W$ and $f_i \otimes \mathbf{v}_i \in W^\vee \otimes V$ respectively, by linearity we have $S \circ T$ as a linear combination of elements of the form $\Psi(f_i \otimes \mathbf{v}_i) \circ \Psi(g_j \otimes \mathbf{w}_j) \in \mathcal{L}(V,V)$ (and similarly $T \circ S$). This means that

$$\mathrm{tr}(S \circ T) = \sum_{i,j} \mathrm{tr}(\Psi(f_i \otimes \mathbf{v}_i) \circ \Psi(g_j \otimes \mathbf{w}_j)) = \sum_{i,j} \mathrm{tr}(\Psi(g_j \otimes \mathbf{w}_j) \circ \Psi(f_i \otimes \mathbf{v}_i)) = \mathrm{tr}(T \circ S),$$

as desired. □

5 Determinant

If you came from the last section, you might be fascinated (or frightened) by the basis-free definition of trace. We could do the same thing for determinants as well, but instead let's take a completely different route here: we are going to define determinants simply by the formula everyone knows, to show different levels of abstraction.

Remark. To make this section complete, I have also included an optional part of defining an intrinsic definition of $\det T$ at the end.

5.1 Definitions and properties

Let's cut to the chase and give the definition first:

Definition 5.1 (Minors and Determinants)

Let $A \in M_n(F)$, the space of $n \times n$ matrices with entries in F. The **ij-minor** of A, denoted by A_{ij}, is the $(n-1) \times (n-1)$ matrix obtained by deleting the i-th row and the j-th column from A.

Now write $A = (a_{ij})$. The **determinant** of A, denoted by $\det(A)$ or $|A|$, is defined inductively:

- For $n = 1$: $\det(A) = a_{11}$.
- Suppose the determinant of an $(k-1) \times (k-1)$ matrix has already been defined. Then for $n = k$, define:

$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(A_{1j}) = a_{11} \det(A_{11}) - a_{12} \det(A_{12}) + \cdots + (-1)^{n+1} a_{1n} \det(A_{1n}).$$

Note that the definition makes sense since A_{1j} is of dimension $(k-1) \times (k-1)$.

This definition is sometimes referred to as the expansion along the first row of A. For instance,

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \qquad \text{and} \qquad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix},$$

which you should already be familiar with.

Remark. This is not the only way to define the determinant: we will see another way later, but there is yet another definition using the notion of "the sign of a permutation".

Example 5.2

Consider the following 3×3 matrix over $\mathbb{R}$:

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 0 & -1 \\ -1 & 2 & 1 \end{pmatrix}.$$

Then its determinant can be computed by

$$\begin{aligned} \det(A) &= 1 \det \begin{pmatrix} 0 & -1 \\ 2 & 1 \end{pmatrix} - 2 \det \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} + 0 \\ &= (0 \cdot 1 - (-1) \cdot 2) - 2(2 \cdot 1 - (-1) \cdot (-1)) \\ &= 2 - 2 = 0 \end{aligned}$$

We will now develop some crucial properties of determinants which will make calculations much simpler. Essentially this reduces to considering how does the determinant change when we apply row operations.

Proposition 5.3 (Basic properties of determinants)

The determinant satisfies the following:

(i) (Linearity in row) If $\lambda \in F$, $\mathbf{v} \in F^n$ and the following matrices are equal except in the denoted row, then

$$\det \begin{pmatrix} \vdots \\ \lambda \mathbf{v} \\ \vdots \end{pmatrix} = \lambda \det \begin{pmatrix} \vdots \\ \mathbf{v} \\ \vdots \end{pmatrix} \qquad \text{and} \qquad \det \begin{pmatrix} \vdots \\ \mathbf{v}_1 \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ \mathbf{v}_2 \\ \vdots \end{pmatrix} = \det \begin{pmatrix} \vdots \\ \mathbf{v}_1 + \mathbf{v}_2 \\ \vdots \end{pmatrix}.$$

(ii) (Identical consecutive rows) If two consecutive rows of a matrix is equal, then the determinant is 0, i.e.

$$\det \begin{pmatrix} \vdots \\ \mathbf{v} \\ \mathbf{v} \\ \vdots \end{pmatrix} = 0.$$

(iii) (Determinant of identity matrix) $\det(I_n) = 1$.

Proof. The proof of these are simple but extremely messy, so we shall only show the first half of (i):

We will induct on n. The case $n = 1$ is trivial. Suppose the result holds for $(n-1) \times (n-1)$ matrices, and let

$$A = \begin{pmatrix} \vdots \\ \lambda \mathbf{v} \\ \vdots \end{pmatrix} \qquad \text{and} \qquad B = \begin{pmatrix} \vdots \\ \mathbf{v} \\ \vdots \end{pmatrix}$$

where the two matrices differ in row l. We now split into two cases:

$l > 1$ The first row of A is the same as that of B, so

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

But for each j, the $(l-1)$-th row of A_{1j} is λ times the $(l-1)$-th row of B_{1j} while all the other rows are the same. By induction hypothesis we have $\det(A_{1j}) = \lambda \det(B_{1j})$. This gives $\det(A) = \lambda \det(B)$ as needed.

$l = 1$ The first row of A is $(\lambda b_{11}, \lambda b_{12}, \ldots, \lambda b_{1n})$, so by the definition of determinants,

$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} \lambda b_{1j} \det(A_{1j}) = \lambda \sum_{j=1}^{n} (-1)^{1+j} b_{1j} \det(A_{1j}).$$

But the minors A_{1j} and B_{1j} are the same, which gives the result. □

Now recall the elementary row operations:

- multiply a row by a non-zero factor;
- add a multiple of a row to another;
- swap two rows.

We have handled the first operation, so we shall cover the rest as follows.

Proposition 5.4 (Row operations on determinants)
Let $A, B \in M_n(F)$. Then

(i) if B is obtained from A by swapping two consecutive rows, then $\det(B) = -\det(A)$;

(ii) if B is obtained from A by swapping any two rows, then $\det(B) = -\det(A)$;

(iii) if B is obtained from A by adding a multiple of a row to another, then $\det(B) = \det(A)$.

Proof. Clearly (ii) implies (i), but we need to prove (i) first in order to obtain (ii).

(i) Denote the two rows of A by $\mathbf{v}_i$ and $\mathbf{v}_{i+1}$. Then

$$0 = \det \begin{pmatrix} \vdots \\ \mathbf{v}_i + \mathbf{v}_{i+1} \\ \mathbf{v}_i + \mathbf{v}_{i+1} \\ \vdots \end{pmatrix} = \det \begin{pmatrix} \vdots \\ \mathbf{v}_i \\ \mathbf{v}_i \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ \mathbf{v}_i \\ \mathbf{v}_{i+1} \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ \mathbf{v}_{i+1} \\ \mathbf{v}_i \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ \mathbf{v}_{i+1} \\ \mathbf{v}_{i+1} \\ \vdots \end{pmatrix} = \det(A) + \det(B).$$

(ii) We first show that if A has any two rows equal, then the determinant is 0.

Indeed, we can repeatedly interchange consecutive rows of A to end up with a matrix B with two consecutive rows equal. By (i), $\det(B) = \pm\det(A)$. But then $\det(B) = 0$.

Now by redoing the proof of (i) with the two equal rows arbitrarily placed, the result follows.

(iii) Suppose B is obtained from adding λ times row i in A to row j. Then

$$\det(B) = \det \begin{pmatrix} \vdots \\ \mathbf{v}_i \\ \vdots \\ \lambda\mathbf{v}_i + \mathbf{v}_j \\ \vdots \end{pmatrix} = \det \begin{pmatrix} \vdots \\ \mathbf{v}_i \\ \vdots \\ \lambda\mathbf{v}_i \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ \mathbf{v}_i \\ \vdots \\ \mathbf{v}_j \\ \vdots \end{pmatrix} = \lambda \det \begin{pmatrix} \vdots \\ \mathbf{v}_i \\ \vdots \\ \mathbf{v}_i \\ \vdots \end{pmatrix} + \det(A) = \det(A),$$

as desired. □

As a result, this summarises to

! Keypoint
If A, B are row-equivalent, then $\det(A) = \mu\det(B)$ for some non-zero $\mu \in F$.

In fact, we even have the following stronger theorem, which is one of the most useful application of determinants:

Theorem 5.5
A is row-equivalent to I_n iff $\det(A) \neq 0$.

Proof. The above keypoint shows the ($\Rightarrow$) direction, since $\det(I_n) = 1$.

For the other direction, suppose A is not row-equivalent to I_n. Then it is row-equivalent to a matrix B with a row of zeroes, since it has rank less than n. WLOG assume this is the first row, then $\det(B) = 0$ and it follows from the keypoint again that $\det(A) = 0$. □

Motivation

Note that the statement "A is row-equivalent to I_n" has multiple equivalent formulations. Indeed, from previous discussions, all of the following are equivalent to it:

- A is invertible;
- A is non-singular (i.e. the only solution to $A\mathbf{x} = 0$ is $\mathbf{x} = 0$);
- A has rank n;
- the rows/columns of A are linearly independent.

Having established the notion of determinants, we now have yet another a systematic method to say whether these conditions hold true by simply computing the determinant.

Another corollary after we have determined the effects of row operations on determinants is that we can now expand along any row:

Proposition 5.6 (Expansion along the i-th row)

Let $A \in M_n(F)$ and $1 \leq i \leq n$. Then

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

Proof. Let B be the matrix A with rows 1 and i exchanged, then

$$\det(A) = -\det(B) = -\sum_{j=1}^{n} (-1)^{1+j} a_{ij} \det(B_{1j}) = \sum_{j=1}^{n} (-1)^{j} a_{ij} \det(B_{1j}).$$

Comparing A_{ij} with B_{1j}, they look like

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & & & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix} \qquad \begin{pmatrix} a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & & & \vdots & & \vdots \\ a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & & & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix}$$

so most of their entries are the same, except for a different permutation of the rows. In particular we can swap $(i-2)$ pairs of rows in B_{1j} to obtain A_{ij}, so $\det(B_{1j}) = (-1)^{i-2} \det(A_{ij}) = (-1)^{i} \det(A_{ij})$. This implies the result. □

Example 5.7 (Determinant of upper triangular matrix)

Consider an $n \times n$ upper triangular matrix:

$$A = \begin{pmatrix} a_{11} & & & & \\ & a_{22} & & * & \\ & & a_{33} & & \\ & 0 & & \ddots & \\ & & & & a_{nn} \end{pmatrix}.$$

Then $\det(A) = a_{11}a_{22}\dots a_{nn}$. Indeed, we can expand along the last row, then $\det(A) = a_{nn}\det(A_{nn})$. Inductively, this implies the result.

We now move on to showing that determinant is **multiplicative**. This will eventually help us establish more properties of determinants.

Lemma 5.8
Let $A \in M_n(F)$ and let $E \in M_n(F)$ be an $n \times n$ elementary matrix. Then $\det(EA) = \det(E)\det(A)$.

Proof. Recall that EA is the matrix obtained by applying the corresponding row operation of E onto A. Hence by above, $\det(EA) = \mu \det(A)$ for some $\mu \in F$. But at the same time $\det(E) = \det(EI_n) = \mu \det(I_n) = \mu$ since μ only depends on the row operation. The result follows. □

You might notice that this implies $\det(AB) = \det(A)\det(B)$ already, by a simple induction argument. But we still have to deal with the case when $\det(A)$ or $\det(B)$ is 0.

It boils down to the following lemma, which we shall not prove since it is quite simple:

Lemma 5.9
Suppose $A, B \in M_n(F)$. Then AB is singular if and only if at least one of A and B is singular.

Using both lemmas, we finally deduce the desired result.

Theorem 5.10 (Determinant is multiplicative)
Let $A, B \in M_n(F)$. Then $\det(AB) = \det(A)\det(B)$.

Proof. If $\det(AB) = 0$, then this is Lemma 5.9. Otherwise, we can write

$$A = E_1 E_2 \cdots E_r \quad \text{and} \quad B = E_1' E_2' \cdots E_s'$$

as products of elementary matrices. This gives $\det(A) = \det(E_1) \cdots \det(E_r)$ and $\det(B) = \det(E_1') \cdots \det(E_s')$. Multiplying them gives the desired result. □

This has a myriad of consequences:

Proposition 5.11
Let $A \in M_n(F)$. We have

(i) $\det(A^{-1}) = 1/\det(A)$ if A is invertible;

(ii) $\det(A^T) = \det(A)$;

(iii) expansion along columns works: $\det(A) = \sum_i (-1)^{i+j} a_{ij} \det(A_{ij})$.

Proof. All of these are quite trivial:

(i). Follows from $\det(A)\det(A^{-1}) = \det(I_n) = 1$.

(ii). One could check that $\det(E^T) = \det(E)$ for elementary matrix E, by listing the three cases of E. Write $A = E_1 \dots E_r$, then

$$\det(A^T) = \det(E_r^T) \cdots \det(E_1^T) = \det(E_1) \cdots \det(E_r) = \det(A).$$

(iii). Transpose A and apply the corresponding result for expansion along the j-th row of the transpose. □

5.2 Some applications

We will go through two applications of determinants here: one is a useful special case, and the other one provides another way to compute inverses and solutions of a system.

Proposition 5.12

Let $n \geq 2$ and $x_1, \ldots, x_n \in F$. Consider the following $n \times n$ matrix:

$$\begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{pmatrix}.$$

The determinant of this matrix is then

$$\prod_{1 \leq i < j \leq n} (x_j - x_i),$$

also called the **Vandermonde determinant**.

Proof. We can now use row and column operations. So we apply column operations

$$C_n \mapsto C_n - x_1 C_{n-1}, \quad C_{n-1} \mapsto C_{n-1} - x_1 C_{n-2}, \quad \ldots \quad, C_2 \mapsto C_2 - x_1 C_1,$$

all of which does not change the determinant. This gives

$$\det \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{pmatrix} = \det \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & x_2 - x_1 & x_2(x_2 - x_1) & \cdots & x_2^{n-2}(x_2 - x_1) \\ \vdots & & \vdots & \ddots & \vdots \\ 1 & x_n - x_1 & x_n(x_n - x_1) & \cdots & x_n^{n-2}(x_n - x_1) \end{pmatrix}.$$

Expanding along the top row and pulling out factors, this is equal to

$$(x_2 - x_1) \cdots (x_n - x_1) \cdot \det \begin{pmatrix} 1 & x_2 & \cdots & x_2^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^{n-2} \end{pmatrix}$$

but the remaining determinant is an $(n-1) \times (n-1)$ Vandermonde determinant too. Thus the result follows by induction, and the base case $n = 2$ is clear. □

In particular, notice that

$$\prod_{1 \leq i < j \leq n} (x_j - x_i) = 0 \iff x_i = x_j \text{ for some } i \neq j.$$

This implies the following cute corollary:

Corollary 5.13

Any polynomial $p(x) = a_0 + a_1 x + \cdots + a_{n-1} x^{n-1}$ with coefficients in F has at most $n - 1$ distinct roots in F.

Proof. Suppose $x_1, \ldots, x_n \in F$ are roots of p, then we have

$$a_0 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + a_1 \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \cdots + a_{n-1} \begin{pmatrix} x_1^{n-1} \\ \vdots \\ x_n^{n-1} \end{pmatrix} = 0.$$

Since $a_0, \ldots, a_{n-1}$ are not all zero, the columns of the Vandermonde matrix are linearly dependent, and thus its determinant is 0. Hence $x_i = x_j$ for some $i \neq j$. □

Determinants can also be used to find the inverse of a matrix. This relies on the following definition:

Definition 5.14
Let $A = (a_{ij}) \in M_n(F)$. For $1 \leq i, j \leq n$, the **ij-th cofactor** of A is

$$c_{ij} = (-1)^{i+j} \det(A_{ij}).$$

The matrix $C = (c_{ij}) \in M_n(F)$ is then called the **matrix of cofactors** of A.

Notice that if we expand $\det(A)$ along the j-th column, we have

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}) = \sum_{i=1}^{n} c_{ij} a_{ij},$$

but this is precisely the j-th diagonal term of $C^T A$. Turns out these are the only non-zero terms:

Proposition 5.15
Let $A \in M_n(F)$ and C its matrix of cofactors. Then

$$C^T A = \det(A) I_n.$$

In particular, if $\det(A) \neq 0$, then $A^{-1} = \frac{1}{\det(A)} C^T$.

Proof. We want to compute

$$(C^T A)_{jk} = \sum_{i=1}^{n} c_{ij} a_{ik}$$

for $j \neq k$. Consider a matrix A' same with A but with the j-th column replaced by the k-th column, i.e. $a'_{ij} = a_{ik}$.

Notice that the calculation of c_{ij} doesn't involve the j-th column of A: the ij-minor of A removes the j-th column. Hence the ij-th cofactor c'_{ij} of A' is equal to c_{ij}. Thus

$$\sum_{i=1}^{n} c_{ij} a_{ik} = \sum_{i=1}^{n} c'_{ij} a'_{ij} = \det(A') = 0$$

since A' has two identical columns. This completes the proof. □

Remark. The matrix C^T is sometimes called the **adjugate matrix** of A, denoted by $\operatorname{adj}(A)$.

This provides a more computational method to find the inverse:

Example 5.16
Consider the matrix $A = \begin{pmatrix} -2 & 3 & 2 \\ 6 & 0 & 3 \\ 4 & 1 & -1 \end{pmatrix}$. One can compute that

$$\operatorname{adj}(A) = \begin{pmatrix} -3 & 18 & 6 \\ 5 & -6 & 14 \\ 9 & 18 & -18 \end{pmatrix}^T = \begin{pmatrix} -3 & 5 & 9 \\ 18 & -6 & 18 \\ 6 & 14 & -18 \end{pmatrix},$$

so it remains to compute the determinant of A. But instead of expanding again, we can utilise the identity $\operatorname{adj}(A)A = \det(A) I_n$ and compare the 11-term on both sides. This gives

$$\det(A) = (-3)(-2) + (5)(6) + (9)(4) = 72,$$

and so the inverse of A is obtained by directly dividing $\operatorname{adj}(A)$ by 72.

Although the method of cofactors is not very useful for larger matrices, it does serve a theoretical consequence:

! Keypoint

Entries of the inverse of a matrix are rational functions, i.e. of the form $p(\mathbf{x})/q(\mathbf{x})$ for some polynomials p, q.

Here, the variable $\mathbf{x}$ are the entries of the matrix, and indeed, this follows from the fact that the determinant of an $n \times n$ matrix is a polynomial function of the n^2 entries. In particular, if $F = \mathbb{R}$ or $\mathbb{C}$, then this map is continuous.

We end the section by giving one last application, called the **Cramer's Rule** for finding solutions of a system.

Proposition 5.17 (Cramer's Rule)

Let $A \in M_n(F)$ and $\mathbf{b} = (b_1, \dots, b_n)^T \in F^n$. Consider the equation $A\mathbf{x} = \mathbf{b}$. Suppose A is invertible, so this has a unique solution

$$\mathbf{x} = (x_1, \dots, x_n)^T = A^{-1}\mathbf{b}.$$

For $1 \leq i \leq n$, denote A_i as the result of replacing the i-th column of A by $\mathbf{b}$. Then we have

$$x_i = \det(A_i)/\det(A).$$

Proof. Write $A^{-1} = (a'_{ij})$, so

$$x_i = \sum_{j=1}^{n} a'_{ij} b_j.$$

But by $A^{-1} = \frac{1}{\det(A)} C^T$ and the definition of C, we have

$$\det(A) x_i = \sum_{j=1}^{n} c_{ji} b_j = \sum_{j=1}^{n} (-1)^{i+j} \det(A_{ji}) b_j = \det(A_i)$$

where the last equality comes from expanding $\det(A_i)$ down column i. □

5.3 Wedge product ⋆

As before, we now want to define the determinant of a linear transformation. A naïve way to do this is to define something as follows:

Definition 5.18 (Determinant of linear map (naïve))

Suppose V is a finite dimensional F-vector space with basis B and $T : V \to V$ is a linear transformation. The determinant of T is then $\det([T]_B)$.

To check that this works, we have to answer the question we raised many times: why does this definition of the determinant not depend on the choice of the basis? This is answered by some of the tools developed before:

Proposition 5.19

The determinant $\det(T)$ does not depend on the choice of the basis.

Proof. Let C be another basis of V, so we want to check $\det([T]_B) = \det([T]_C)$. But we know that

$$[T]_C = {}_C[\mathrm{id}]_B [T]_B {}_B[\mathrm{id}]_C = P[T]_B P^{-1}$$

where $P = {}_C[\mathrm{id}]_B$ is the change of basis matrix from B to C. Hence,

$$\det([T]_C) = \det(P) \det([T]_B) \det(P^{-1}) = \det([T]_B)$$

as desired. □

But of course we are not so easily satisfied. The goal of this section is then to give the **basis-free** definition of the determinant. Turns out, this will trivialise several properties of the determinant (e.g. that the determinant is multiplicative).

This requires something called the **wedge product**, which looks at first like the tensor product $V \otimes V$ but with some extra relations.

Definition 5.20 (2-wedge product)

Let V be an F-vector space. The **2-wedge product** $\Lambda^2(V)$ is defined as the span of the elements of the form $\mathbf{v} \wedge \mathbf{w}$ (where $\mathbf{v}, \mathbf{w} \in V$), subject to the same relations as $\otimes$:

- distributive in V: $(\mathbf{v}_1 + \mathbf{v}_2) \wedge \mathbf{w} = \mathbf{v}_1 \wedge \mathbf{w} + \mathbf{v}_2 \wedge \mathbf{w}$;
- distributive in W: $\mathbf{v} \wedge (\mathbf{w}_1 + \mathbf{w}_2) = \mathbf{v} \wedge \mathbf{w}_1 + \mathbf{v} \wedge \mathbf{w}_2$;
- $(\lambda \mathbf{v}) \wedge \mathbf{w} = \mathbf{v} \wedge (\lambda \mathbf{w})$.

but with two additional relations:

- $\mathbf{v} \wedge \mathbf{v} = 0$;
- $\mathbf{v} \wedge \mathbf{w} = -\mathbf{w} \wedge \mathbf{v}$.

Again, scalar multiplication is defined by $\lambda \cdot (\mathbf{v} \wedge \mathbf{w}) = (\lambda \mathbf{v}) \wedge \mathbf{w} = \mathbf{v} \wedge (\lambda \mathbf{w})$.

Remark. In fact, the relation $\mathbf{v} \wedge \mathbf{w} = -\mathbf{w} \wedge \mathbf{v}$ is extraneous: expanding $(\mathbf{v} + \mathbf{w}) \wedge (\mathbf{v} + \mathbf{w}) = 0$ gives the desired relation, so $\mathbf{v} \wedge \mathbf{v} = 0$ is the only new requirement.

You might wonder, what are the two new relations? And how is this related to the determinant? The following example might give a hint:

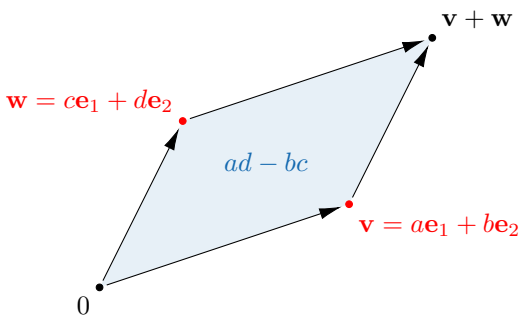
Example 5.21 (Explicit computation)

Let $V = \mathbb{R}^2$, and let $\mathbf{v} = a\mathbf{e}_1 + b\mathbf{e}_2, \mathbf{w} = c\mathbf{e}_1 + d\mathbf{e}_2$. Now let's compute $\mathbf{v} \wedge \mathbf{w}$ in $\Lambda^2(V)$:

$$\begin{aligned}\mathbf{v} \wedge \mathbf{w} &= (a\mathbf{e}_1 + b\mathbf{e}_2) \wedge (c\mathbf{e}_1 + d\mathbf{e}_2) \\ &= ac(\mathbf{e}_1 \wedge \mathbf{e}_1) + bd(\mathbf{e}_2 \wedge \mathbf{e}_2) + ad(\mathbf{e}_1 \wedge \mathbf{e}_2) + bc(\mathbf{e}_2 \wedge \mathbf{e}_1) \\ &= ad(\mathbf{e}_1 \wedge \mathbf{e}_2) + bc(\mathbf{e}_2 \wedge \mathbf{e}_1) \\ &= (ad - bc)(\mathbf{e}_1 \wedge \mathbf{e}_2).\end{aligned}$$

What is $ad - bc$? You might already recognize it

- as the area of the parallelogram formed by v and w; or
- as the determinant of $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$. In fact, the determinant is meant to interpret hypervolumes.



This is absolutely no coincidence: the wedge product is designed to interpret **signed areas**. You can see why the condition $(\lambda\mathbf{v}) \wedge \mathbf{w} = \mathbf{v} \wedge (\lambda\mathbf{w})$ would make sense now, and now of course you know why $\mathbf{v} \wedge \mathbf{v}$ should be 0: it is an area zero parallelogram. The miracle here is that the only additional requirement is $\mathbf{v} \wedge \mathbf{v} = 0$, and suddenly the wedge product will do all our work of interpreting volumes.

Recall from before that the basis for $V \otimes W$ is $\mathbf{e}_i \otimes \mathbf{f}_j$ over all i, j. In analog to this, we have:

Proposition 5.22 (Basis of $\Lambda^2(V)$)

Let V be a finite-dimensional vector space with basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. Then $\mathbf{e}_i \wedge \mathbf{e}_j$ over all $i < j$ is a basis for $\Lambda^2(V)$.

Remark. In particular, this implies that $\dim \Lambda^2(V) = \binom{n}{2}$.

This is natural since $\mathbf{e}_i \wedge \mathbf{e}_j = -\mathbf{e}_j \wedge \mathbf{e}_i$. Hence by removing all the extra basis elements, we are left with the ones with $i < j$. We will not prove this result since it is quite similar with the one before.

Now, we extend our results by defining a multi-dimensional wedge product:

Definition 5.23

Let V be an F-vector space and m a positive integer. The **m-wedge product** $\Lambda^m(V)$ is defined as the span of elements of the form

$$\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \cdots \wedge \mathbf{v}_m$$

subject to the relations

- distributive in any term: $\cdots \wedge (\mathbf{v}_1 + \mathbf{v}_2) \wedge \cdots = (\cdots \wedge \mathbf{v}_1 \wedge \cdots) + (\cdots \wedge \mathbf{v}_2 \wedge \cdots)$;
- $\cdots \wedge (\lambda\mathbf{v}_1) \wedge \mathbf{v}_2 \wedge \cdots = \cdots \wedge \mathbf{v}_1 \wedge (\lambda\mathbf{v}_2) \wedge \cdots$;
- $\cdots \wedge \mathbf{v} \wedge \mathbf{v} \wedge \cdots = 0$;
- $\cdots \wedge \mathbf{v} \wedge \mathbf{w} \wedge \cdots = -(\cdots \wedge \mathbf{w} \wedge \mathbf{v} \wedge \cdots)$.

Again, scalar multiplication is defined similar to before.

Although the definition is quite wordy, it just says

- we should be able to add products like before;
- you can put constants onto any of the m components; and
- switching **any two adjacent wedges** negates the whole wedge.

Note that any element of the form

$$\cdots \wedge \mathbf{v} \wedge \cdots \wedge \mathbf{v} \wedge \cdots$$

is still zero. Similar to be before, we again have (but will not prove):

Proposition 5.24 (Basis of $\Lambda^m(V)$)

Let V be a finite-dimensional vector space with basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. Then

$$\mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} \wedge \cdots \wedge \mathbf{e}_{i_m}$$

where

$$1 \le i_1 < i_2 < \cdots < i_m \le n$$

is a basis for $\Lambda^m(V)$.

Remark. In particular, this implies that $\dim \Lambda^m(V) = \binom{n}{m}$.

We are finally ready to define the determinant. Suppose $T : V \to V$ is a linear transformation. We claim that the map $\Lambda^m(V) \to \Lambda^m(V)$ given on wedges by

$$\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \cdots \wedge \mathbf{v}_m \mapsto T(\mathbf{v}_1) \wedge T(\mathbf{v}_2) \wedge \cdots \wedge T(\mathbf{v}_m)$$

and extending linearly to all of $\Lambda^m(V)$ is a linear map. We call this map $\Lambda^m(T)$.

Example 5.25

Let $V = \mathbb{R}^4$ with standard basis $\mathbf{e}_i$. Consider a map with $T(\mathbf{e}_1) = \mathbf{e}_2, T(\mathbf{e}_2) = 2\mathbf{e}_3, T(\mathbf{e}_3) = \mathbf{e}_3$ and $T(\mathbf{e}_4) = 2\mathbf{e}_2 + \mathbf{e}_3$. Then, for example, $\Lambda^2(T)$ sends

$$\begin{aligned}\mathbf{e}_1 \wedge \mathbf{e}_2 + \mathbf{e}_3 \wedge \mathbf{e}_4 &\mapsto T(\mathbf{e}_1) \wedge T(\mathbf{e}_2) + T(\mathbf{e}_3) \wedge T(\mathbf{e}_4) \\ &= \mathbf{e}_2 \wedge 2\mathbf{e}_3 + \mathbf{e}_3 \wedge (2\mathbf{e}_2 + \mathbf{e}_3) \\ &= 2(\mathbf{e}_2 \wedge \mathbf{e}_3 + \mathbf{e}_3 \wedge \mathbf{e}_2) \\ &= 0.\end{aligned}$$

However, it turns out that this map is special for $m = n$:

Proposition 5.26

Let $T : V \to V$ be a linear transformation, then $\Lambda^{\dim V}(T)$ is a multiplication by some constant.

Proof. Suppose V has dimension n. Then $\Lambda^n(V)$ has dimension $\binom{n}{n} = 1$, so it is isomorphic to the base field F.

Hence $\Lambda^n(T)$ can be thought of as a linear map from F to F. However any linear map f from F to F is precisely a multiplication by a constant: $f(x) = f(x \cdot 1) = xf(1) = cx$ where $c = f(1)$. The result follows. □

Thus, it makes sense to define:

Definition 5.27 (Determinant of linear map (basis-free))

Suppose V is an n-dimensional vector space. Then $\Lambda^n(T)$ is a multiplication by some constant c. The **determinant** of T is then defined as $c = \det(T)$.

Again, in this way the determinant is basis-free, since we defined it in terms of $\Lambda^n(T)$. We also have

$$\Lambda^n(S \circ T) = \Lambda^n(S) \circ \Lambda^n(T)$$

by definition, so we get $\det(S \circ T) = \det(S)\det(T)$ for free.

Example 5.28

Let $V = \mathbb{R}^2$ again with basis $B = \{\mathbf{e}_1, \mathbf{e}_2\}$. Let $T : V \to V$ be represented by

$$[T]_B = \begin{pmatrix} a & c \\ b & d \end{pmatrix}.$$

In other words, $T(\mathbf{e}_1) = a\mathbf{e}_1 + b\mathbf{e}_2$ and $T(\mathbf{e}_2) = c\mathbf{e}_1 + d\mathbf{e}_2$.

Now let's consider $\Lambda^2(V)$. It has a basis $\mathbf{e}_1 \wedge \mathbf{e}_2$. Then $\Lambda^2(T)$ sends it to

$$\Lambda^2(T)(\mathbf{e}_1 \wedge \mathbf{e}_2) = T(\mathbf{e}_1) \wedge T(\mathbf{e}_2) = (ad - bc)(\mathbf{e}_1 \wedge \mathbf{e}_2).$$

So $\Lambda^2(T) : \Lambda^2(V) \to \Lambda^2(V)$ is a multiplication by $\det(T) = ad - bc$.

Remark. More generally, if we replace 2 by n, and write out the result of expanding

$$(a_{11}\mathbf{e}_1 + a_{21}\mathbf{e}_2 + \cdots + a_{n1}\mathbf{e}_n) \wedge \cdots \wedge (a_{1n}\mathbf{e}_1 + a_{2n}\mathbf{e}_2 + \cdots + a_{nn}\mathbf{e}_n),$$

we will indeed get the formula for $n \times n$ determinants.

6 Eigen-things

We know that a linear map T from V to V is really just a square matrix. So what is the simplest type of linear map? It would be multiplication by some scalar λ, which would have corresponding matrix (in any basis!)

$$[T]_B = \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda \end{pmatrix}.$$

That's perhaps *too* simple though. If we had a fixed basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ then another very simple operation would just be scaling each basis element $\mathbf{e}_i$ by λ_i, i.e.

$$[T]_B = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

These maps are more general. Indeed, you can, for example, compute T^{100} easily: the maps sends $\mathbf{e}_i \mapsto \lambda_i^{100}\mathbf{e}_i$. Doing this with an arbitrary matrix would have been a disaster.

Of course, most linear maps are probably not that nice. *Or are they*?

Motivation

Let V be some two-dimensional vector space with $\mathbf{e}_1$ and $\mathbf{e}_2$ as basis elements. Consider a map $T : V \to V$ by $\mathbf{e}_1 \mapsto 2\mathbf{e}_1$ and $\mathbf{e}_2 \mapsto \mathbf{e}_1 + 3\mathbf{e}_2$, i.e.

$$[T]_{\{\mathbf{e}_1, \mathbf{e}_2\}} = \begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix}.$$

This doesn't look appealing until we realise we can rewrite it as

$$\begin{aligned} \mathbf{e}_1 &\mapsto 2\mathbf{e}_1 \\ \mathbf{e}_1 + \mathbf{e}_2 &\mapsto 3(\mathbf{e}_1 + \mathbf{e}_2). \end{aligned}$$

So suppose we change to the basis $\{\mathbf{e}_1, \mathbf{e}_1 + \mathbf{e}_2\}$, then

$$[T]_{\{\mathbf{e}_1, \mathbf{e}_1 + \mathbf{e}_2\}} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

and so our random-looking map, under a suitable change of basis, looks like the nice maps we described above!

In this section, we will be *making* our luck, so that arbitrary matrices can be written into our desired form. This requires the notion of **eigenvalues** and **eigenvectors**, as we shall introduce now.

6.1 Definitions

In the above example, we saw that there were two very nice vectors, $\mathbf{e}_1$ and $\mathbf{e}_1 + \mathbf{e}_2$, for which T did something very simple. Naturally, these vectors have a name:

Definition 6.1 (Eigenvalues and eigenvectors)

Suppose V is a vector space over F and $T : V \to V$ is a linear map. We say that $\lambda \in F$ is an **eigenvalue** of T if there is a **non-zero** vector $\mathbf{v} \in V$ with $T(\mathbf{v}) = \lambda\mathbf{v}$. Such a vector $\mathbf{v}$ is called an **eigenvector** of T.

Of course, one could then define the same notion for matrices: given a matrix $A \in M_n(F)$, we require $A\mathbf{v} = \lambda\mathbf{v}$ instead. It is easy to show that the eigenvalues of T are the same as the eigenvalues of $[T]_B$, where B is a basis. So, from the above example,

Example 6.2

Let $T : V \to V$ be a linear map, with $[T]_{\{\mathbf{e}_1,\mathbf{e}_2\}} = \begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix}$ as before. Then

- $\mathbf{e}_1$ and $\mathbf{e}_1 + \mathbf{e}_2$ are eigenvectors with eigenvalues 2 and 3 respectively.
- Of course, $5\mathbf{e}_1$ is also a 2-eigenvector.
- And, $7\mathbf{e}_1 + 7\mathbf{e}_2$ is also a 3-eigenvector.

Continuing the example, if we change to the basis $B = \{\mathbf{e}_1, \mathbf{e}_1 + \mathbf{e}_2\}$, we have the relation

$$[T]_B = {}_B[\mathrm{id}]_E \, {}_E[T]_E \, {}_E[\mathrm{id}]_B = P^{-1}AP$$

where E is the basis $\{\mathbf{e}_1, \mathbf{e}_2\}$. But $D := [T]_B$ is a diagonal matrix, so from $A = PDP^{-1}$ we have

$$A^k = (PDP^{-1})(PDP^{-1}) \cdots (PDP^{-1}) = PD^kP^{-1}$$

and D^k is much easier to compute.

This raises a few questions:

- Does every linear transformation (or matrix) have eigenvectors, or more generally, n eigenvectors?
- If there are n eigenvectors, must they form a basis of V?
- If so, after changing the basis, must we obtain a diagonal matrix?

Unfortunately, the first question is already wrong:

Example 6.3 (Eigenvectors need not exist)

Let $V = \mathbb{R}^2$ and let T be the map represented by $[T]_E = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Then

$$T(\mathbf{v}) = \lambda\mathbf{v} \implies -v_2 = \lambda v_1 \quad \text{and} \quad v_1 = \lambda v_2$$

which gives $\mathbf{v} = 0_V$. Note that 0_V is **not** an eigenvector, so this map has no eigenvectors and eigenvalues.

In fact, geometrically, this maps correspond to rotating a vector by 90° around the origin, so it is clear that $T(\mathbf{v})$ is not a multiple of $\mathbf{v}$ for any $\mathbf{v} \in V$.

However, in some cases, we can still guarnatee the existence of eigenvectors. To classify the precise situations when this happens, we need a new tool:

Definition 6.4

Let V be a finite-dimensional F-vector space, and B a basis for V. Let $T : V \to V$ be a linear map, then the **characteristic polynomial** for T is

$$\chi_T(x) = \det(x \cdot \mathrm{id} - T) = \det(xI_n - [T]_B).$$

You should be able to deduce the similar definition for a matrix. And again, notice that this does not depend on the choice of basis B: if we have another basis C then $[T]_C = P^{-1}[T]_B P$, so

$$\begin{aligned} \det(xI_n - [T]_C) &= \det(P^{-1}(xI_n - [T]_B)P) \\ &= \det(P^{-1}) \det(xI_n - [T]_B) \det(P) \\ &= \det(xI_n - [T]_B). \end{aligned}$$

The importance of this new tool is that eigenvalues are closely related to it:

Proposition 6.5

Suppose $T : V \to V$ is a linear transformation and $\lambda \in F$. Then λ is an eigenvalue of T if and only if $\chi_T(\lambda) = 0$.

Proof. For $\lambda \in F$, note that λ is an eigenvalue of T if and only if there is a non-zero vector $\mathbf{v} \in V$ such that $(\lambda \operatorname{id} - T)\mathbf{v} = 0$. But this means that the matrix $(\lambda I_n - [T]_B)$ is singular (where B is any basis), so by previous discussions this is equivalent to $\det(\lambda I_n - [T]_B) = 0$, i.e. $\chi_T(\lambda) = 0$. □

We also have that the characteristic polynomial have degree at most n (since the term with highest degree comes from expanding the determinant and the product of the diagonal terms). So we immediately obtain

Corollary 6.6

If $T : V \to V$ is a linear map, then T has at most $\dim V$ eigenvalues in F.

We will also make use of the following notation:

Definition 6.7

Given a linear transformation $T : V \to V$ (or a matrix $A \in M_n(F)$, for which $V = F^n$), for any $\lambda \in F$, the **λ-eigenspace** E_λ is the set of λ-eigenvectors, together with 0_V, i.e.

$$E_\lambda = \{\mathbf{v} \in V : T(\mathbf{v}) = \lambda \mathbf{v}\}$$

Notice that this is a subspace of V: it is a kernel of the transformation $\lambda \cdot \operatorname{id} - T$, and λ is an eigenvalue if and only if this is not the zero-subspace. This notation then allows us to state succinctly sentences such as "2 is an eigenvalue of T with one-dimensional eigenspace spanned by $\mathbf{e}_1$".

Example 6.8

Let's try to compute the eigenvalues and eigenvectors of the 2×2 matrix:

$$A = \begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix}.$$

We have its characteristic polynomial as

$$\chi_A(x) = \det \begin{pmatrix} x-2 & -1 \\ 1 & x \end{pmatrix} = x^2 - 2x + 1 = (x-1)^2,$$

so the only eigenvalue of A is 1. Now solving $A\mathbf{v} = 1\mathbf{v}$, we have

$$(A - I_2)\mathbf{v} = 0 \implies \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix} \mathbf{v} = 0 \implies \mathbf{v} \in \operatorname{Span} \begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

so the eigenvectors are the non-zero scalar multiples of $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

6.2 Diagonalisation

As the name of this section suggests, we now explore the other questions mentioned before. Namely, we will classify when can we "diagonalise" a linear map (or a matrix), i.e. it becomes a diagonal map after a change of basis.

Definition 6.9

A linear map $T : V \to V$ is **diagonalisable** if there is a basis of V consisting of eigenvectors of T.

We have to explain the terminology, so we have the following theorem:

Theorem 6.10

Suppose V is a finite-dimensional F-vector space and $T : V \to V$ is a linear map. Then T is diagonalisable if and only if there is a basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of V such that $D = [T]_B$ is diagonal.

Proof. Suppose that $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is any basis of V. Note that $\mathbf{v}_i \neq 0$. Let $D = [T]_B$. Then, by the definition of $[T]_B$, it is a diagonal matrix if and only if $T(\mathbf{v}_i) = d_{ii}\mathbf{v}_i$ for each $1 \leq i \leq n$. But this is equivalent to $\mathbf{v}_i$ being an eigenvector of T with eigenvalue d_{ii}, as needed. □

Combining with the definition of diagonalisable, this answers the third question from above:

! Keypoint

If the eigenvectors of T form a basis B of V, then $[T]_B$ is diagonal.

In the case of matrices, we can even compute the desired change of basis matrix:

Corollary 6.11

Let $A \in M_n(F)$ be a matrix. If its eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of A form a basis B of F^n, then $P^{-1}AP$ is a diagonal matrix, where the columns of P are $\mathbf{v}_i$.

Proof. Write T_A as the linear map $\mathbf{v} \mapsto A\mathbf{v}$. Notice that by definition, $P = {}_E[\mathrm{id}]_B$. This gives

$$P^{-1}AP = {}_B[\mathrm{id}]_E[T_A]_E\,{}_E[\mathrm{id}]_B = [T_A]_B$$

which is indeed a diagonal matrix by the keypoint. Moreover, the diagonal matrix have entries λ_i on its diagonal. □

Recall from above the following example. We can extend the example via this result now:

Example 6.12 (Eigenvectors need not exist, but they do in $\mathbb{C}$)

Recall the example where $V = \mathbb{R}^2$ and T is the map represented by $[T]_E = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. We have shown that this has no eigenvalues in $\mathbb{R}$, and hence no eigenvectors in $\mathbb{R}^2$.

However, by extending to $V = \mathbb{C}^2$ and viewing T as a linear map from $\mathbb{C}$ to $\mathbb{C}$,

$$\chi_T(x) = \det \begin{pmatrix} x & 1 \\ -1 & x \end{pmatrix} = x^2 + 1 = (x + i)(x - i)$$

so we have eigenvalues $\pm i$. Corresponding eigenvectors are $\begin{pmatrix} i \\ 1 \end{pmatrix}$ for i and $\begin{pmatrix} -i \\ 1 \end{pmatrix}$ for $-i$. Since they form a basis for $\mathbb{C}$, T is diagonalisable over $\mathbb{C}$. By the corollary, we can then write

$$[T]_E = \begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix} \underbrace{\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}}_{\text{diagonal}} \begin{pmatrix} -\frac{i}{2} & \frac{1}{2} \\ \frac{i}{2} & \frac{1}{2} \end{pmatrix} = P[T]_B P^{-1}.$$

Remark. For experts: In general, this means

Theorem 6.13 (Eigenvalues always exist over algebraically closed fields)

Suppose F is an **algebraically closed** field (i.e. every non-constant polynomial has a root). Let V be a finite-dimensional F-vector space. Then if $T : V \to V$ is a linear map, there exists an eigenvalue $\lambda \in F$.

which is trivial since eigenvalues *are* roots of a polynomial, namely, the characteristic polynomial. Of course, there might not be n distinct eigenvectors, see Example 6.8 where there is only one eigenvalue and one eigenvector, independent of the base field.

Now it's time to answer the second and final question. The following result is useful: essentially, it states eigenvectors for different eigenvalues are linearly independent.

Theorem 6.14
Suppose V is an F-vector space and $T : V \to V$ is a linear map. Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be the eigenvectors of T with $T(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$. If the λ_i are distinct then $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent.

Proof. We prove this by induction on n. The base case is by $\mathbf{v}_1 \neq 0$ since $\mathbf{v}_1$ is an eigenvector.

For the induction step, assume that the result is true for fewer than n eigenvectors. Suppose $\alpha_1, \dots, \alpha_n$ and

$$\alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n = 0.$$

If $\alpha_i = 0$ for some i, then by induction hypothesis all α_i must be zero. So now assume that the α_i's are all non-zero. Dividing the above equation by α_1 and applying T, we have

$$0 = T(0) = T\left(\mathbf{v}_1 + \frac{\alpha_2}{\alpha_1}\mathbf{v}_2 + \cdots + \frac{\alpha_n}{\alpha_1}\mathbf{v}_n\right) = \lambda_1 \mathbf{v}_1 + \sum_{i=2}^{n} \frac{\lambda_i \alpha_i}{\alpha_1}\mathbf{v}_i.$$

Subtracting $\frac{\lambda_1}{\alpha_1}$ times the original equation from this, we have

$$0 = \left(\lambda_1 \mathbf{v}_1 + \sum_{i=2}^{n} \frac{\lambda_i \alpha_i}{\alpha_1}\mathbf{v}_i\right) - \left(\lambda_1 \mathbf{v}_1 + \sum_{i=2}^{n} \frac{\lambda_1 \alpha_i}{\alpha_1}\mathbf{v}_i\right) = \sum_{i=2}^{n} \frac{(\lambda_i - \lambda_1)\alpha_i}{\alpha_1}\mathbf{v}_i.$$

By induction hypothesis, we must have $(\lambda_i - \lambda_1)\alpha_i = 0$ for all $2 \leq i \leq n$. Since the λ_i's are distinct, we have $\alpha_i = 0$ for $2 \leq i \leq n$, contradiction to our assumption. This implies the result. □

This immediately implies that

! Keypoint
If $T : V \to V$ has $\dim V$ distinct eigenvalues, then the eigenvectors of T form a basis of V (i.e. T is diagonalisable).

Another way to put this is that if $\chi_T(x)$ has n distinct roots, then T is diagonalisable. We might generalise the result, to classify all cases when a linear map T is diagonalisable:

Proposition 6.15
Let V be a finite-dimensional vector space and $T : V \to V$ be a linear map. Suppose that T has (distinct) eigenvalues $\lambda_1, \dots, \lambda_r$, and let E_{λ_i} be the λ_i-eigenspace. Denote B_i to be a basis of E_{λ_i}. If $\sum |B_i| = \dim V$, then the union $B = B_1 \cup \cdots \cup B_r$ is a basis of V, i.e. T is diagonalisable.

Proof. Write $B_i = \{\mathbf{v}_{i1}, \dots, \mathbf{v}_{in(i)}\}$ (so $\dim E_{\lambda_i} = n(i)$). It suffices to show that the vectors $\mathbf{v}_{ij}$ are linear independent. Hence suppose that

$$\sum_{i=1}^{r} \sum_{j=1}^{n(i)} \alpha_{ij} \mathbf{v}_{ij} = 0$$

for some $\alpha_{ij} \in F$. Let

$$\mathbf{w}_i = \sum_{j=1}^{n(i)} \alpha_{ij} \mathbf{v}_{ij} \in \operatorname{Span}(\mathbf{v}_{i1}, \dots, \mathbf{v}_{in(i)}) = E_{\lambda_i},$$

so $\mathbf{w}_1 + \mathbf{w}_2 + \cdots + \mathbf{w}_r = 0$.

As the λ_i's are distinct, Theorem 6.14 gives $\mathbf{w}_i = 0$ for all $i \leq r$, since any non-zero $\mathbf{w}_i$ would give us a linear dependence between these. So for each fixed i, $\sum \alpha_{ij} \mathbf{v}_{ij} = 0$. But the vectors $\mathbf{v}_{ij}$ form a basis for E_{λ_i} and hence are linearly independent, so $\alpha_{ij} = 0$ for all i, j as needed. □

So we have completely answered the questions. We also note here that if $\sum |B_i| < \dim V$, then T is not diagonalisable: we cannot have a basis consisting of eigenvectors. Combining with the above,

! Keypoint

A linear map $T : V \to V$ with eigenvalues λ_i is diagonalisable if and only if $\sum \dim E_{\lambda_i} = \dim V$.

Remark. Final remark: Proposition 6.15 actually has a more succinct formulation; we can show that

$$E_{\lambda_1} + E_{\lambda_2} + \cdots + E_{\lambda_r} = \bigoplus_{i=1}^{r} E_{\lambda_i}$$

is a direct sum, i.e. each element is represented uniquely. Then T is diagonalisable iff $V = \bigoplus E_{\lambda_i}$.

6.3 Interlude: A glimpse in inner product spaces

We know now one equivalent formulation for when a linear map is diagonalisable. But how that will be achieved is still unclear. We will now try to move on to showing that a class of linear maps (and matrices) are diagonalisable using the above results, but that requires a bit of technicalities, which we shall cover here.

Caution: Throughout this chapter, **all vector spaces are over** $\mathbb{R}$, unless otherwise specified.

Definition 6.16 (Inner product)

Let V be a real vector space. An **inner product** is a function

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$$

which satisfies the following properties:

- symmetry: $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$;
- billinearity: $\langle \mathbf{v}_1 + \mathbf{v}_2, \mathbf{w} \rangle = \langle \mathbf{v}_1, \mathbf{w} \rangle + \langle \mathbf{v}_2, \mathbf{w} \rangle$ and $\langle \lambda \mathbf{v}, \mathbf{w} \rangle = \lambda \langle \mathbf{v}, \mathbf{w} \rangle$ and similarly in the second argument;
- positive-definite: $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ for any $\mathbf{v}$, and equality holds only if $\mathbf{v} = 0_V$.

As we have seen before, billinearity simply means that $\langle \cdot, \cdot \rangle$ is linear in both arguments.

Example 6.17 ($\mathbb{R}^n$)

As you might already know, one can define an inner product on $\mathbb{R}^n$ as the **dot product**. Let $\mathbf{e}_i$ be the usual basis, then we let

$$\langle \alpha_1 \mathbf{e}_1 + \cdots + \alpha_n \mathbf{e}_n, \beta_1 \mathbf{e}_1 + \cdots + \beta_n \mathbf{e}_n \rangle := \alpha_1 \beta_1 + \cdots + \alpha_n \beta_n.$$

It is easy to see that this is indeed symmetric and billinear. To see it is positive definite, note that if $a_i = b_i$ then the dot product is $a_1^2 + \cdots + a_n^2$, which is exactly zero when all a_i are zero.

Remark. You might wonder: why do we limit ourselves to $\mathbb{R}$? This is because inner products only exist in some spaces, namely $\mathbb{R}$- or $\mathbb{C}$-vector spaces. We didn't include the definition of complex inner products since (i) it is not relavant to later discussions and (ii) it is quite different from real inner products.

With an inner product, we can now consider "length" of vectors and "distances":

Definition 6.18 (Norm)

Let V be a real vector space with an inner product. The **norm** of $\mathbf{v} \in V$ is defined by

$$||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

Note that this only makes sense since the inner product is assumed to be positive-definite.

We now have this super famous result, which will become a stepping stone for later:

Lemma 6.19 (Cauchy-Schwarz)

Let V be a real vector space with an inner product. For any $\mathbf{v}, \mathbf{w} \in V$, we have

$$|\langle \mathbf{v}, \mathbf{w} \rangle| \leq ||\mathbf{v}|| \, ||\mathbf{w}||$$

with equality if and only if $\mathbf{v}$ and $\mathbf{w}$ are linearly dependent.

Proof. The theorem is immediate if $||\mathbf{v}|| = 0$, so henceforth we assume $\mathbf{v} \neq 0_V$.

The key step is to consider the equality case: we will use the inequality $\langle \lambda \mathbf{v} - \mathbf{w}, \lambda \mathbf{v} - \mathbf{w} \rangle \geq 0$. Deferring the choice of λ for later, we compute

$$\begin{aligned} 0 &\leq \langle \lambda \mathbf{v} - \mathbf{w}, \lambda \mathbf{v} - \mathbf{w} \rangle \\ &= \langle \lambda \mathbf{v}, \lambda \mathbf{v} \rangle - \langle \lambda \mathbf{v}, \mathbf{w} \rangle - \langle \lambda \mathbf{w}, \lambda \mathbf{v} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle \\ &= \lambda^2 \langle \mathbf{v}, \mathbf{v} \rangle - \lambda \langle \mathbf{v}, \mathbf{w} \rangle - \lambda \langle \mathbf{w}, \mathbf{v} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle \\ 2\lambda \langle \mathbf{v}, \mathbf{w} \rangle &\leq \lambda^2 ||\mathbf{v}||^2 + ||\mathbf{w}||^2. \end{aligned}$$

At this point, a good choice of λ is

$$\lambda = \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{||\mathbf{v}||^2},$$

since then

$$2 \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{||\mathbf{v}||^2} \langle \mathbf{v}, \mathbf{w} \rangle \leq \frac{\langle \mathbf{v}, \mathbf{w} \rangle^2}{||\mathbf{v}||^4} ||\mathbf{v}||^2 + ||\mathbf{w}||^2 \implies \langle \mathbf{v}, \mathbf{w} \rangle^2 \leq ||\mathbf{v}||^2 \, ||\mathbf{w}||^2,$$

as desired. The equality holds if and only if $\mathbf{w} = \lambda \mathbf{v}$ from the start of the inequality. □

Remark. The choice of λ might seem arbitrary, but it is in fact natural: we can view the inequality above as a quadratic in λ, then this particular choice of λ minimises the quadratic.

We can also show:

Theorem 6.20 (Triangle inequality)

Let V be a real vector space with an inner product. For any $\mathbf{v}, \mathbf{w} \in V$, we have

$$||\mathbf{v}|| + ||\mathbf{w}|| \geq ||\mathbf{v} + \mathbf{w}||$$

Proof. Omitted; proved by squaring both sides applying Cauchy-Schwarz. □

Just as the case in $\mathbb{R}^n$, a natural notion we would be interested in is orthogonality:

Definition 6.21 (Orthogonal)

Two non-zero vectors $\mathbf{v}, \mathbf{w}$ in a real vector space with an inner product are **orthogonal** if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$.

More generally, a set of non-zero vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ form an **orthogonal set** if they are pairwise orthogonal, and in addition if each $\mathbf{v}_i$ have norm 1 then they form an **orthonormal set**.

As we expect from our geometric intuition in $\mathbb{R}^n$, this implies independence:

Lemma 6.22

Orthogonal sets are linearly independent.

Proof. If $\sum \alpha_i \mathbf{v}_i = 0$ where $\alpha_i \in \mathbb{R}$, then

$$0_V = \left\langle \mathbf{v}_1, \sum \alpha_i \mathbf{v}_i \right\rangle = \alpha_1 ||v_1||^2,$$

and so $a_1 = 0$ since v_1 is non-zero. Similarly a_i are all zero. □

It turns out that any vector space with an inner product has a basis that is also orthonormal:

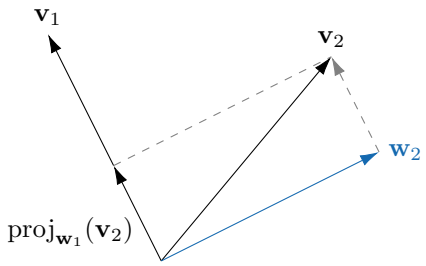
Theorem 6.23 (Gram-Schmidt)

Let $\mathbf{v}_1, \dots, \mathbf{v}_r$ be linearly independent vectors in V. Let $\text{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u}$, and recursively define

$$\begin{aligned}
\mathbf{w}_1 &= \mathbf{v}_1 \\
\mathbf{w}_2 &= \mathbf{v}_2 - \text{proj}_{\mathbf{w}_1}(\mathbf{v}_2) \\
\mathbf{w}_3 &= \mathbf{v}_3 - \text{proj}_{\mathbf{w}_1}(\mathbf{v}_3) - \text{proj}_{\mathbf{w}_2}(\mathbf{v}_3) \\
&\vdots \\
\mathbf{w}_r &= \mathbf{v}_r - \text{proj}_{\mathbf{w}_1}(\mathbf{v}_r) - \cdots - \text{proj}_{\mathbf{w}_{r-1}}(\mathbf{v}_r).
\end{aligned}$$

Then the set of vectors $\{\mathbf{w}_1, \dots, \mathbf{w}_r\}$ are an orthogonal set of vectors, and for $1 \leq i \leq r$ we have $\text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_i) = \text{Span}(\mathbf{w}_1, \dots, \mathbf{w}_i)$. In particular, if $\mathbf{v_i}$ form a basis, then $\mathbf{w}_i / ||\mathbf{w}_i||$ form an orthonormal basis.

Notice that if we choose $V = \mathbb{R}^n$ and the dot product as the inner product, then $\text{proj}_{\mathbf{u}}(\mathbf{v})$ is the projection of $\mathbf{v}$ onto $\mathbf{u}$. Hence the picture (for $n = 2$) is:



so you can see intuitively that $\mathbf{w}_i$ will indeed be orthogonal.

Proof. The idea is induction on i that $\mathbf{w}_i \neq 0$, $\text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_i) = \text{Span}(\mathbf{w}_1, \dots, \mathbf{w}_i)$ and $\langle \mathbf{w}_k, \mathbf{w}_i \rangle$ for all $k < i$. The base case is trivial for all three of these statements. We now assume that all three statements are true for $1, \dots, i-1$.

- If $\mathbf{w}_i = 0$, then $\mathbf{v}_i \in \text{Span}(\mathbf{w}_1, \dots, \mathbf{w}_{i-1}) = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_{i-1})$ by definition of $\mathbf{w}_i$ and induction hypothesis. Contradiction to the linear independence of $\mathbf{v}_i$'s.
- $\text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_i) = \text{Span}(\mathbf{w}_1, \dots, \mathbf{w}_i)$ comes directly from the definition of $\mathbf{w}_i$ and induction hypothesis.
- Pick any $k < i$, and we want to show $\langle \mathbf{w}_k, \mathbf{w}_i \rangle = 0$. We have

$$\begin{aligned}
\langle \mathbf{w}_k, \mathbf{w}_i \rangle &= \langle \mathbf{w}_k, \mathbf{v}_i \rangle - \sum_{j=1}^{i-1} \langle \mathbf{w}_k, \text{proj}_{\mathbf{w}_j}(\mathbf{v}_i) \rangle \\
&= \langle \mathbf{w}_k, \mathbf{v}_i \rangle - \sum_{j=1}^{i-1} \left\langle \mathbf{w}_k, \frac{\langle \mathbf{v}_i, \mathbf{w}_j \rangle}{\langle \mathbf{w}_j, \mathbf{w}_j \rangle} \mathbf{w}_j \right\rangle \\
&= \langle \mathbf{w}_k, \mathbf{v}_i \rangle - \sum_{j=1}^{i-1} \frac{\langle \mathbf{v}_i, \mathbf{w}_j \rangle}{\langle \mathbf{w}_j, \mathbf{w}_j \rangle} \langle \mathbf{w}_k, \mathbf{w}_j \rangle
\end{aligned}$$

  but by induction hypothesis, since $j, k < i$, if $j \neq k$ then $\langle \mathbf{w}_k, \mathbf{w}_j \rangle = 0$. Hence this simplifies to

$$\langle \mathbf{w}_k, \mathbf{w}_i \rangle = \langle \mathbf{w}_k, \mathbf{v}_i \rangle - \frac{\langle \mathbf{v}_i, \mathbf{w}_k \rangle}{\langle \mathbf{w}_k, \mathbf{w}_k \rangle} \langle \mathbf{w}_k, \mathbf{w}_k \rangle = 0,$$

  as desired. □

Hence, we can generally assume our bases are orthonormal.

Example 6.24

Let V be a finite-dimensional real vector space with an inner product, and consider **any** orthonormal basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. Then we have

$$\langle \alpha_1 \mathbf{e}_1 + \cdots + \alpha_n \mathbf{e}_n, \beta_1 \mathbf{e}_1 + \cdots + \beta_n \mathbf{e}_n \rangle = \sum_{i,j=1}^{n} \alpha_i \beta_j \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \sum_{i=1}^{n} \alpha_i \beta_i$$

since $\{\mathbf{e}_i\}$ are orthonormal. So you can conclude that the dot product is the "only" inner product, if we WLOG assume that every basis of $\mathbb{R}^n$ is orthonormal.

Note that the word orthogonal has been shared with matrices:

Definition 6.25

A matrix $A \in M_n(\mathbb{R})$ is orthogonal if $A^T A = I_n$.

and they are indeed related; we have that a matrix A is orthogonal if and only if its columns form an **orthonormal** set in $\mathbb{R}^n$. Indeed, the ij-th entry of $A^T A$ is the dot product of columns i and j of A.

Moreover, as a corollary to the Gram-Schmidt process, we have

Corollary 6.26

Let V be a real vector space with an inner product, and $\mathbf{u} \in V$ be a unit vector (i.e. with norm 1). Then there is an orthogonal matrix with its first column as $\mathbf{u}$.

Proof. Extend $\mathbf{u}$ to a basis, and apply Gram-Schmidt to obtain an orthonormal basis $\mathbf{u}, \mathbf{u}_2, \dots, \mathbf{u}_n$. Note that the first vector is unchanged by the process. We can then take the matrix having these vectors as its columns. □

This marks the end of our technical requirements to proceed.

6.4 Real symmetric matrices

In this section, we will cover the big result in diagonalisation: if a matrix $A \in M_n(\mathbb{R})$ is symmetric, then it is diagonalisable. This is sometimes called the **spectral theorem**. In fact, the result is even better: there is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.

Remark. Throughout this chapter, to simplify calculations, we will only focus on the real vector space $\mathbb{R}^n$ with the dot product as the inner product.

The key property about a symmetric matrix which makes the proof works is how it iteracts with the inner product:

Lemma 6.27 (Symmetric matrices are self-adjoint)

If $A \in M_n(\mathbb{R})$ is symmetric and $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, then

$$\langle A\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, A\mathbf{v} \rangle.$$

The proof is simple: just notice that $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \mathbf{v}$. The following is the first lemma we need:

Lemma 6.28 (Real symmetric matrices have real eigenvalues)

Suppose $A \in M_n(\mathbb{R})$ is symmetric and $\lambda \in \mathbb{C}$ is a root of $\chi_A(x)$. Then $\lambda \in \mathbb{R}$.

Proof. We may regard A as a matrix in $M_n(\mathbb{C})$, so λ is an eigenvalue of A, and there exists non-zero $\mathbf{v} \in \mathbb{C}^n$ such that $A\mathbf{v} = \lambda \mathbf{v}$. Write $\mathbf{v} = (v_1, \dots, v_n)^T$ and let $\overline{\mathbf{v}} = (\overline{v_1}, \dots, \overline{v_n})^T$. Then $\overline{\mathbf{v}}^T A \mathbf{v} = \overline{\mathbf{v}}^T \lambda \mathbf{v} = \lambda \overline{\mathbf{v}}^T \mathbf{v}$.

But on the other hand, notice that A has real entries, so $A = \overline{A^T}$. Thus $\overline{\mathbf{v}}^T A \mathbf{v} = \overline{(A\mathbf{v})^T} \mathbf{v} = \overline{\lambda} \overline{\mathbf{v}}^T \mathbf{v}$. Since $\mathbf{v} \neq 0$, we have $\overline{\mathbf{v}}^T \mathbf{v} = \sum |v_i|^2 \neq 0$, so $\lambda = \overline{\lambda}$, i.e. $\lambda \in \mathbb{R}$. □

We can also quickly deduce that eigenvectors of distinct eigenvalues are orthogonal:

Lemma 6.29
Suppose $A \in M_n(\mathbb{R})$ is symmetric and $\lambda, \mu \in \mathbb{R}$ are distinct eigenvalues of A with corresponding eigenvectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Then $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

Proof. We have $\lambda\langle \mathbf{u}, \mathbf{v} \rangle = \langle A\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, A\mathbf{v} \rangle = \mu\langle \mathbf{u}, \mathbf{v} \rangle$. As $\lambda \neq \mu$, $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. □

Now comes the advertised theorem. We will show that we can pick an orthonormal basis consisting of eigenvectors of A, which implies that A is diagonalisable.

Theorem 6.30 (Spectral theorem)
Suppose $A \in M_n(\mathbb{R})$ is symmetric. Then there exists an orthonormal matrix $P \in M_n(\mathbb{R})$ with $P^{-1}AP$ a diagonal matrix.

Proof. The proof is by induction on n and the base case $n = 1$ is trivial. Suppose the result is true for $n - 1$.

Notice that $\chi_A(x)$ has at least one root $\lambda_1 \in \mathbb{C}$, by Fundamental theorem of algebra. But then by Lemma 6.28, we must have $\lambda_1 \in \mathbb{R}$, so there is an eigenvalue of A. Let $\mathbf{v}_1$ be a corresponding eigenvector, and WLOG assume $||\mathbf{v}_1|| = 1$. Then there is an orthogonal matrix $P_1 \in M_n(\mathbb{R})$ with first column $\mathbf{v}_1$, by Corollary 6.26. Write the columns as $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Then $P_1^{-1} = P_1^T$ and

$$P_1^{-1}AP_1 = \begin{pmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{pmatrix} \begin{pmatrix} A\mathbf{v}_1 & A\mathbf{v}_2 & \cdots & A\mathbf{v}_n \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{pmatrix} \begin{pmatrix} \lambda_1\mathbf{v}_1 & A\mathbf{v}_2 & \cdots & A\mathbf{v}_n \end{pmatrix}$$
$$= \begin{pmatrix} \lambda_1 & \mathbf{v}_1^T A\mathbf{v}_2 & \cdots & \mathbf{v}_1^T A\mathbf{v}_n \\ 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{pmatrix}$$

where $A' \in M_{n-1}(\mathbb{R})$. But notice that $P_1^{-1}AP_1 = P_1^T AP_1$, which is symmetric. So

$$P_1^{-1}AP_1 = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{pmatrix}$$

and A' is symmetric. By induction hypothesis, there is an orthogonal $P' \in M_{n-1}(\mathbb{R})$ with $(P')^{-1}A'P'$ diagonal. Let

$$P_2 := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & P' & \\ 0 & & & \end{pmatrix},$$

then P_2 is orthogonal since P' is orthogonal. We also have

$$P_2^{-1}(P_1^{-1}AP_1)P_2 = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & (P')^{-1}A'P' & \\ 0 & & & \end{pmatrix}$$

which is diagonal since $(P'^{-1})A'P'$ is diagonal. Then choosing $P = P_1P_2$ yields the desired result. □

We end the section by giving an explicit computation:

Example 6.31

Consider the 3×3 matrix

$$A = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}$$

so that A is symmetric. Let's try to find an orthogonal matrix P such that $P^{-1}AP$ is diagonal. We firstly have

$$\chi_A(x) = \det \begin{pmatrix} x-1 & 1 & 1 \\ 1 & x-1 & 1 \\ 1 & 1 & x-1 \end{pmatrix} = (x+1)(x-2)^2$$

so the eigenvalues are -1 and 3. Although there are not 3 distinct eigenvalues, we still know it is diagonalisable since A is real symmetric. In the usual way, we find

- $E_{-1} = \operatorname{Span} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, so an orthonormal basis for this is $\frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.
- $E_2 = \{(x, y, z)^T : x + y + z = 0\}$, so a basis for this is $(1, -1, 0)^T$ and $(0, 1, -1)^T$. To make this orthonormal, we apply Gram-Schmidt and we have

$$\mathbf{w}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{w}_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}.$$

Finally, we can put these basis vectors in the columns of P and obtain

$$P = \frac{1}{\sqrt{6}} \begin{pmatrix} \sqrt{2} & \sqrt{3} & 1 \\ \sqrt{2} & -\sqrt{3} & 1 \\ \sqrt{2} & 0 & -2 \end{pmatrix}.$$

In particular, another advantage of diagonalising A with an orthogonal matrix is that we don't have to go through the trouble of finding P^{-1}; it is just the transpose of P!

Remark. One could rewrite the whole section to using the language of linear maps. But that would require a linear-map version of "transpose", which we haven't covered yet. There is also a more general version of spectral theorem over $\mathbb{C}$-vector spaces with an inner product.

6.5 The Jordan form and multiplicities ⋆

Of course, maps with diagonal representation is one of the best situation we can get, as seen previously. We have also seen that a large class of maps are diagonal under a suitable basis – but still, not every map is diagonalisable.

In this section, we will generalise the notion of diagonalisable, in hope that matrices with a looser condition can be represented by something not too hard to understand as well.

Definition 6.32

A **Jordan block** is an $n \times n$ matrix of the following shape:

$$\begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix}.$$

In other words, a Jordan block has λ on the diagonal, and 1 above it. We allow $n = 1$, so $\begin{pmatrix}\lambda\end{pmatrix}$ is a Jordan block.

Before we dive into the technicalities, let's state the main theorem:

Theorem 6.33 (Jordan canonical form)

Let $T : V \to V$ be a linear map of finite-dimensional vector spaces over an **algebraically closed field** F. Then we can choose a basis B of V such that $[T]_B$ is "block-diagonal" with each block being a Jordan block. Such a matrix is said to be in **Jordan form**, and is unique up to rearranging the order of the blocks.

As an example, this means that the matrix should look something like

$$\begin{pmatrix}
\lambda_1 & 1 & & & & & & & \\
0 & \lambda_1 & & & & & & & \\
& & \lambda_2 & & & & & & \\
& & & \lambda_3 & 1 & 0 & & & \\
& & & 0 & \lambda_3 & 1 & & & \\
& & & 0 & 0 & \lambda_3 & & & \\
& & & & & & \ddots & & \\
& & & & & & & \lambda_m & 1 \\
& & & & & & & 0 & \lambda_m
\end{pmatrix}$$

Notice that diagonal matrices are the special case when each block is 1×1.

Motivation

What does this mean? This mean that **all matrices (with entries in $\mathbb{C}$, for instance) are Jordan-block-diagonalisable**! This is good news since

- Suppose we have a Jordan form of T, with Jordan blocks $J_1, \dots, J_m$. Then any power of the map is still quite simple, it is just
$$[T^n]_B = \begin{pmatrix} J_1^n & 0 & \cdots & 0 \\ 0 & J_2^n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & J_m^n \end{pmatrix}.$$
- But on the other hand, powers of a Jordan block is not hard to compute too, as we will see later.

So we essentially can compute powers of any map over $\mathbb{C}$.

Example 6.34 (Concrete example of Jordan form)

We can still "read off" how T acts on the eigenvectors, given a Jordan form. For instance, take

$$[T]_B = \begin{pmatrix} 5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

under a certain basis $B = \{\mathbf{v}_1, \dots, \mathbf{v}_6\}$. Then we can compute all the eigenvectors and eigenvalues:

$$\begin{aligned} T(\mathbf{v}_1) &= 5\mathbf{v}_1 \\ T(\mathbf{v}_2) &= 2\mathbf{v}_2 \\ T(\mathbf{v}_4) &= 7\mathbf{v}_4 \\ T(\lambda\mathbf{v}_5 + \mu\mathbf{v}_6) &= 3(\lambda\mathbf{v}_5 + \mu\mathbf{v}_6). \end{aligned}$$

Alright, let's dive into the proof of the theorem now. The most important definition for that is:

Definition 6.35
A map $T : V \to V$ is **nilpotent** if T^m is the zero map for some integer m. (Here T^m means T applied m times.)

What's an example of a nilpotent map?

Example 6.36 (The "descending staircase")
Let $V = F^3$ have basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. Then the map T which sends

$$\mathbf{e}_3 \mapsto \mathbf{e}_2 \mapsto \mathbf{e}_1 \mapsto 0$$

is nilpotent, since $T(\mathbf{e}_1) = T^2(\mathbf{e}_2) = T^3(\mathbf{e}_3) = 0$, so $T^3(\mathbf{v}) = 0$ for all $\mathbf{v} \in V$.

The above 3×3 descending staircase has matrix representation

$$[T]_B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Notice that this is a Jordan block. We can also see that T above has 0 as its only eigenvalue.

As another example, we can have multiple such staircases:

Example 6.37 (Double staircase)
Let $V = F^5$ have basis $\mathbf{e}_1, \dots, \mathbf{e}_5$. Then the map S which sends

$$\mathbf{e}_3 \mapsto \mathbf{e}_2 \mapsto \mathbf{e}_1 \mapsto 0 \quad \text{and} \quad \mathbf{e}_5 \mapsto \mathbf{e}_4 \mapsto 0$$

is nilpotent.

Notice that S this time has Jordan form:

$$[S]_B = \begin{pmatrix} 0 & 1 & 0 & & \\ 0 & 0 & 1 & & \\ 0 & 0 & 0 & & \\ & & & 0 & 1 \\ & & & 0 & 0 \end{pmatrix}.$$

You can see this is not really that different from the previous example; it is just the same idea repeated multiple times. In fact, we now claim that **all** nilpotent maps have essentially that form:

Theorem 6.38 (Nilpotent Jordan)
Let V be a finite-dimensional F-vector space, where F is algebraically closed, and $T : V \to V$ be a nilpotent map. Then $V = \bigoplus_{i=1}^m V_i$ where each V_i has a basis of the form $\mathbf{v}_i, T(\mathbf{v}_i), \dots, T^{\dim V_i - 1}(\mathbf{v}_i)$ for some $\mathbf{v}_i \in V_i$.

This looks horribly daunting, but we can understand the statement better by looking at an example:

Example 6.39
Using the double staircase example from above, we notice that

$$V = \operatorname{Span}(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) \oplus \operatorname{Span}(\mathbf{e}_4, \mathbf{e}_5) = \operatorname{Span}(\mathbf{e}_3, T(\mathbf{e}_3), T(T(\mathbf{e}_3))) \oplus \operatorname{Span}(\mathbf{e}_5, T(\mathbf{e}_5))$$

since of course each element in V can be uniquely written as a sum of a linear combination of $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ and a linear combination of $\mathbf{e}_4, \mathbf{e}_5$. In particular, by choosing the basis $B = \{S(S(\mathbf{e}_3)), S(\mathbf{e}_3), \mathbf{e}_3, S(\mathbf{e}_5), \mathbf{e}_5\}$, we have the matrix representation as above.

So the theorem is merely stating that any nilpotent maps gives such a basis, i.e. we might choose

$$\begin{aligned} B = \{&T^{\dim V_1 - 1}(\mathbf{v}_1), \dots, T(\mathbf{v}_1), \mathbf{v}_1, \\ &T^{\dim V_2 - 1}(\mathbf{v}_2), \dots, T(\mathbf{v}_2), \mathbf{v}_2, \\ &\dots, \\ &T^{\dim V_m - 1}(\mathbf{v}_m), \dots, T(\mathbf{v}_m), \mathbf{v}_m\} \end{aligned}$$

to be a basis of V, so that the matrix representation will become m blocks of staircases. In other words,

! Keypoint

Every nilpotent map has a matrix representation of independent staircases.

Proof. We induct on $\dim V$. The case when $\dim V = 1$ is trivial. Assume $\dim V \geq 1$, and let $W = \operatorname{im} T$. Since T is nilpotent, $W \neq V$. Moreover, if $W = \{0\}$ (i.e. T is the zero map) then we are already done. So assume $\{0\} \subset W \subset V$.

Hence $\dim W < \dim V$. By the induction hypothesis, there is a basis of W in the form

$$\begin{aligned} B' = \{&\mathbf{w}_1, T(\mathbf{w}_1), T(T(\mathbf{w}_1)), \dots \\ &\mathbf{w}_2, T(\mathbf{w}_2), T(T(\mathbf{w}_2)), \dots \\ &\dots, \\ &\mathbf{w}_k, T(\mathbf{w}_k), T(T(\mathbf{w}_k)), \dots\} \end{aligned}$$

for some $\mathbf{w}_i \in W$. But $W = \operatorname{im} T$, so we can rewrite the basis as

$$\begin{aligned} B' = \{&T(\mathbf{v}_1), T(T(\mathbf{v}_1)), T(T(T(\mathbf{v}_1))), \dots \\ &T(\mathbf{v}_2), T(T(\mathbf{v}_2)), T(T(T(\mathbf{v}_2))), \dots \\ &\dots, \\ &T(\mathbf{v}_k), T(T(\mathbf{v}_k)), T(T(T(\mathbf{v}_k))), \dots\}. \end{aligned}$$

Now note that there are exactly k elements of B' which are in $\ker T$ (namely the last element of each of the k staircases). We can thus complete it to a basis of $\ker T$ by adding some vectors $\mathbf{v}_{k+1}, \dots, \mathbf{v}_m \in \ker T$ (where $m = \dim \ker T$). Now we consider

$$\begin{aligned} B = \{&\mathbf{v}_1, T(\mathbf{v}_1), T(T(\mathbf{v}_1)), T(T(T(\mathbf{v}_1))), \dots \\ &\mathbf{v}_2, T(\mathbf{v}_2), T(T(\mathbf{v}_2)), T(T(T(\mathbf{v}_2))), \dots \\ &\dots, \\ &\mathbf{v}_k, T(\mathbf{v}_k), T(T(\mathbf{v}_k)), T(T(T(\mathbf{v}_k))), \dots \\ &\mathbf{v}_{k+1}, \mathbf{v}_{k+2}, \dots, \mathbf{v}_m\}. \end{aligned}$$

Then there are exactly $k + \dim W + (\dim \ker T - k) = \dim \ker T + \dim \operatorname{im} T = \dim V$ elements, by rank-nullity theorem. Moreover, B is linearly independent since if there is a linear dependence, then via taking T,

- the staircases $\{\mathbf{v}_i, T(\mathbf{v}_i), T(T(\mathbf{v}_i)), T(T(T(\mathbf{v}_i))), \dots\}$ gets sent to the corresponding staircase in B',
- the elements $\{\mathbf{v}_{k+1}, \dots, \mathbf{v}_m\}$ is sent to 0 since they are in $\ker T$.

Hence this would result in a linear dependence of elements in B', contradiction.

Therefore B is a basis of the desired form (note that $\mathbf{v}_{k+1}, \dots, \mathbf{v}_m$ form a staircase by themselves, since they are sent to 0 by applying T one time). □

Motivation

Incredibly, we are almost done! If we take the double staircase again, we can compute

$$[S+\lambda\cdot\mathrm{id}]_B = [S]_B + \lambda I_5 = \begin{pmatrix} \lambda & 1 & 0 & & \\ 0 & \lambda & 1 & & \\ 0 & 0 & \lambda & & \\ & & & \lambda & 1 \\ & & & 0 & \lambda \end{pmatrix}$$

which is just two λ Jordan blocks! This gives us a plan to proceed: we need to break V into a bunch of subspaces such that $T-\lambda\cdot\mathrm{id}$ is nilpotent over each subspace. Then nilpotent Jordan will give us the desired Jordan block.

Hence, we will now try to reduce to the nilpotent case. We first need a lemma:

Lemma 6.40

Let V be a finite-dimensional vector space, and $T:V\to V$ be a linear map. Denote T^n as T applied n times, then there exists an integer N such that

$$V = \ker T^N \oplus \operatorname{im} T^N.$$

Proof. Consider

$$\{0\} \subset \ker T \subseteq \ker T^2 \subseteq \ker T^3 \subseteq \cdots$$
$$V \supset \operatorname{im} T \supseteq \operatorname{im} T^2 \supseteq \operatorname{im} T^3 \supseteq \cdots.$$

Notice that each $\subset$ means an increase of dimension (and similarly $\supset$ is a decrease of dimension). Since V is finite-dimensional, the two above chains must eventually stabalise, i.e. for some N we have

$$\ker T^N = \ker T^{N+1} = \cdots \quad \text{and} \quad \operatorname{im} T^N = \operatorname{im} T^{N+1} = \cdots.$$

When this happens, we have $\ker T^N \cap \operatorname{im} T^N = \{0\}$, for if $\mathbf{w}\in\ker T^N\cap\operatorname{im} T^N$ then $\mathbf{w}=T^N(\mathbf{v})$ for some $\mathbf{v}$, but

$$\mathbf{w}=T^N(\mathbf{v})\in\ker T^N \implies T^{2N}(\mathbf{v})=0 \implies \mathbf{v}\in\ker T^{2N}=\ker T^N$$

so $\mathbf{w}=T^N(\mathbf{v})=0$.

On the other hand, by rank-nullity theorem, we also have $\dim\ker T^N+\dim\operatorname{im} T^N=\dim V$. But then

$$\dim(\ker T^N+\operatorname{im} T^N)=\dim\ker T^N+\dim\operatorname{im} T^N-\dim(\ker T^N\cap\operatorname{im} T^N)=\dim V$$

so $\ker T^N+\operatorname{im} T^N=V$ too. This gives $V=\ker T^N\oplus\operatorname{im} T^N$ as desired. □

Remark. One might notice that we didn't actually prove before that $U+V$ is a direct sum if $U\cap V=\{0\}$. But this is quite simple: if $\mathbf{u}_1+\mathbf{v}_1=\mathbf{u}_2+\mathbf{v}_2$ in $U+V$, then

$$\mathbf{u}_1-\mathbf{u}_2=\mathbf{v}_2-\mathbf{v}_1$$

but the left side is in U and the right side is in V. So they are both equal to $\{0\}$, i.e. $\mathbf{u}_1=\mathbf{u}_2$ and $\mathbf{v}_1+\mathbf{v}_2$ as needed.

Now we just need a few more results to get to the end. Bear with me now:

Definition 6.41 (Invariant subspaces)

Let $T:V\to V$. A subspace $W\subseteq V$ is called **T-invariant** if $T(\mathbf{w})\in W$ for any $\mathbf{w}\in W$. In this way, T can be thought of as a map $W\to W$.

Definition 6.42 (Indecomposable)

A map $T : V \to V$ is called **indecomposable** if it is impossible to write $V = W_1 \oplus W_2$ where both W_1 and W_2 are non-trivial T-invariant subspaces.

In this way, the Jordan form is a decomposition of V into invariant subspaces:

Example 6.43

Suppose a map T has Jordan form

$$[T]_B = \begin{pmatrix} 3 & 1 & 0 & & \\ 0 & 3 & 1 & & \\ 0 & 0 & 3 & & \\ & & & 2 & 1 \\ & & & 0 & 2 \end{pmatrix}$$

with some basis $B = \{\mathbf{e}_1, \dots, \mathbf{e}_5\}$. Then $\operatorname{Span}(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ is T-invariant, since

$$T(\mathbf{e}_1) = 3\mathbf{e}_1, \quad T(\mathbf{e}_2) = \mathbf{e}_1 + 3\mathbf{e}_2, \quad T(\mathbf{e}_3) = \mathbf{e_2} + 3\mathbf{e}_3$$

and similarly $\operatorname{Span}(\mathbf{e}_4, \mathbf{e}_5)$ is T-invariant. Hence T is decomposable.

As you might expect, we can break a space apart into "indecomposable" parts.

Proposition 6.44 (Invariant subspace decomposition)

Let V be a finite-dimensional vector space. Given any map $T : V \to V$, we can write

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_m$$

where each V_i is T-invariant, and for any i the restriction map $T : V_i \to V_i$ is indecomposable.

Proof. Same as the proof that every integer is the product of primes: if V is not decomposable, we are done. Otherwise, by definition write $V = W_1 \oplus W_2$ and repeat on each of W_1 and W_2. By dimension reasons this must terminate. □

With this, we can finally prove the big theorem. We will show that given such a decomposition, there is a basis B of V such that $[T]_B$ has Jordan blocks on the diagonal, each corresponding to a T-invariant subspace V_i.

Proof of Theorem 6.33. Consider a decomposition as above, so that the restriction $T : V_1 \to V_1$ is an indecomposable map. Then T has an eigenvalue λ_1 since F is algebraically closed, so let $S = T - \lambda_1 \cdot \mathrm{id}$, hence $\ker S \neq \{0\}$.

Now we claim that V_1 is also S-invariant. Indeed, given $\mathbf{v} \in V_1$,

$$S(\mathbf{v}) = T(\mathbf{v}) - \lambda_1 \cdot \mathrm{id}(\mathbf{v}) = T(\mathbf{v}) - \lambda_1 \mathbf{v} \in V_1.$$

So it makes sense to consider the restriction $S : V_1 \to V_1$. In fact, T-invariant and S-invariant is equivalent by the exact same argument. This also implies that S is indecomposable since T is assumed to be indecomposable.

By Lemma 6.40, we then have

$$V_1 = \ker S^N \oplus \operatorname{im} S^N$$

for some N. It is easy to see that $\ker S^N$ and $\operatorname{im} S^N$ are themselves S-invariant subspaces. But S is indecomposable, so this can only happen if $\operatorname{im} S^N = \{0\}$ and $\ker S^N = V_1$ (since $\ker S^N$ contains our eigenvector).

Hence S is nilpotent, so it can be written as a collection of staircases $S = \bigoplus S_i$ by nilpotent Jordan. But since S is indecomposable, there is only one staircase. Hence T has a Jordan block on the columns corresponding to the basis of V_1 and similarly for V_i, as desired. □

That was a lot to take in, so let's do an explicit computation:

Example 6.45

Consider $T : F^3 \to F^3$ with basis B of F such that T has representation

$$[T]_B = \begin{pmatrix} 3 & -2 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Let's try to find the Jordan form of this matrix and the corresponding change of basis matrix. Firstly, we have

$$\chi_T(x) = \det \begin{pmatrix} x-3 & 2 & 0 \\ -1 & x & 0 \\ -1 & 0 & x-1 \end{pmatrix} = x(x-1)(x-3) + 2(x-1) = (x-1)^2(x-2).$$

So the eigenvalues are 1 and 2.

We need the eigenspaces of T:

- If $\lambda = 1$, then $\begin{pmatrix} 2 & -2 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \mathbf{v} = 0$, so $\mathbf{v} \in \operatorname{Span} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.
- Similarly, if $\lambda = 2$, then $\begin{pmatrix} 1 & -2 & 0 \\ 1 & -2 & 0 \\ 1 & 0 & -1 \end{pmatrix} \mathbf{v} = 0$, so $\mathbf{v} \in \operatorname{Span} \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$.

In particular notice that this is **not** diagonalisable. From here, we already know that under some basis, T has the Jordan form

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

since (i) Jordan forms are upper-triangular, i.e. the eigenvalues are its diagonal entries (and eigenvalues are unchanged under a change of basis); and (ii) T is not diagonalisable.

We now want to compute a basis that transforms T to this. Suppose that basis is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. Then

$$T(\mathbf{v}_1) = \mathbf{v}_1, \quad T(\mathbf{v}_2) = \mathbf{v}_1 + \mathbf{v}_2, \quad T(\mathbf{v}_3) = 2\mathbf{v}_3$$

or equivalently

$$(T - \mathrm{id})(\mathbf{v}_1) = 0, \quad (T - \mathrm{id})(\mathbf{v}_2) = \mathbf{v}_1, \quad (T - 2\,\mathrm{id})(\mathbf{v}_3) = 0$$

Clearly we can choose $\mathbf{v}_3$ as a 2-eigenvector. For $\mathbf{v_2}$ and $\mathbf{v}_1$, we want them to form a staircase $\mathbf{v}_2 \mapsto \mathbf{v}_1 \mapsto 0$. So we compute

$$[(T - \mathrm{id})^2]_B = \begin{pmatrix} 2 & -2 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix}^2 = \begin{pmatrix} 2 & -2 & 0 \\ 1 & -1 & 0 \\ 2 & -2 & 0 \end{pmatrix}$$

which has kernel $\operatorname{Span}\left\{(0,0,1)^T, (1,1,0)^T\right\}$. We need to pick $\mathbf{v}_2$ out of $\ker(T - \mathrm{id})$, i.e. not a 1-eigenvector, so there is only one possibility:

$$\mathbf{v}_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}.$$

Putting them together, we finally have

$$P = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{pmatrix}.$$

Interested reader might try to work out the formula for the powers of Jordan blocks.

As you can probably see, although this section proved the existence of such a form for any linear transformation, the computation is still extremely bothersome. In fact, the calculation of the change of basis matrix for larger matrices might even require a computation of a large power of $(A-I)^n$; but that is inevitable as a consequence of this more general notion.

To end this section, we shall introduce a final convenient notation:

Definition 6.46

Let $T : V \to V$ be a linear map and λ a scalar.

- The **geometric multiplicity** of λ is the dimension $\dim E_\lambda$ of the λ-eigenspace.
- Define the **generalised eigenspace** E^λ to be the subspace of V for which $(T-\lambda\cdot\mathrm{id})^n(\mathbf{v})=0$ for some $n\geq 1$. The **algebraic multiplicity** of λ is the dimension $\dim E^\lambda$.

Let's understand this via an example:

Motivation

Consider the following matrix in Jordan form:

$$[T]_B = \begin{pmatrix} 7 & 1 & & & & \\ 0 & 7 & & & & \\ & & 9 & & & \\ & & & 7 & 1 & 0 \\ & & & 0 & 7 & 1 \\ & & & 0 & 0 & 7 \end{pmatrix}.$$

We focus on the eigenvalue 7, which appears multiiple times, so it is certainly "repeated". However, there are two different senses in which you could say it is repeated:

- Algebraic: You could say it is repeated five times, since it appears five times on the diagonal.
- Geometric: You could say it really only appears twice since there are only two eigenvectors with eigenvalue 7, namely $\mathbf{e}_1$ and $\mathbf{e}_4$.

Indeed, the vector $\mathbf{e}_2$ for instance has $T(\mathbf{e}_2)=7\mathbf{e}_2+\mathbf{e}_1$, so it is not really an eigenvector; but applying $T-7\cdot\mathrm{id}$ two times on $\mathbf{e}_1$ we do get zero. Hence $\mathbf{e}_1\in E^7$.

But in fact with our understanding of eigenvalues and characteristic polynomial, we can also conclude that the **algebraic multiplicity is the amount of λ appears as a root of $\chi_T(x)$.**

With our understanding of Jordan form, we can also say:

Proposition 6.47

Let $T : V \to V$ be a linear map of finite-dimensional vector spaces, written in Jordan form. Let $\lambda \in F$, then

- The geometric multiplicity of λ is the number of Jordan blocks with eigenvalue λ.
- The algebraic multiplicity of λ is the sum of the dimensions of the Jordan blocks with eigenvalue λ.

By this observation, we also conclude the following:

Proposition 6.48

T is diagonalisable if and only if for any λ, its geometric and algebraic multiplicity coincide.

And this marks the end of our discussion towards eigenvalues and eigenvectors.