Algebra II – MATH 251

Winter 2017 Course Notes

Dr. Eyal Goren, McGill University

CONTENTS

Introduction.
This course is about *vector spaces* and the maps between them, called *linear transformations* (or linear maps, or linear mappings).

Vector spaces have their origins in physics. Vectors in 3-dimensional space may represent the momentum of travelling bodies. Those can be *added* and *rescaled* by an real scalar. And all this without introducing any coordinate system as yet. Later on, the same structure was revealed in the study of functions, for example, continuous functions $f : \mathbb{R} \to \mathbb{R}$. Once more, two functions can be added: $f + g$ is the function whose value at x is $f(x) = g(x)$. They can also be rescaled, that is multiplied by a real number r: rf is the function whose value at x is $r \cdot f(x)$. We see here the emergence of a similar structure: a set whose elements can be added and multiplied by a scalar. If we now replace the field $\mathbb{R}$ by any field $\mathbb{F}$, and the set V can be any set, we obtain the notion of a vector space: a set V so that any two elements v, w of it, called "vectors", can be added and each vector can be rescaled by an element of $\mathbb{F}$. Naturally, we impose some reasonable conditions (such as those occurring for adding vectors in physical 3-dimensional space). To illustrate: we want $v + w = w + v$, $(u + v) + w = u + (v + w)$, and so on.

By introducing coordinates, the space around us can be thought of as $\mathbb{R}^3$. By abstraction we understand what are

$$\mathbb{R}, \mathbb{R}^2, \mathbb{R}^3, \mathbb{R}^4, \ldots, \mathbb{R}^n, \ldots,$$

where $\mathbb{R}^n$ is then thought of as vectors $(x_1, \ldots, x_n)$ whose coordinates x_i are real numbers. Replacing the field $\mathbb{R}$ by any field $\mathbb{F}$, we can equally conceive of the spaces

$$\mathbb{F}, \mathbb{F}^2, \mathbb{F}^3, \mathbb{F}^4, \ldots, \mathbb{F}^n, \ldots,$$

where, again, $\mathbb{F}^n$ is then thought of as vectors $(x_1, \ldots, x_n)$ whose coordinates x_i are in $\mathbb{F}$. $\mathbb{F}^n$ is called the vector space of dimension n over $\mathbb{F}$, where adding vectors is done coordinate-wise. Our goal will be, in the large, to build a theory that applies equally well to $\mathbb{R}^n$ and $\mathbb{F}^n$. We will also be interested in constructing a theory which is free of coordinates; the introduction of coordinates will be largely for computational purposes. Now, it turns out, that the theory of vector spaces in themselves is very limited. Once one introduces the notion of a basis and proves a thing or two about it, the theory is over. A basis provides us with coordinates on the space and all vector spaces look like $\mathbb{F}^n$, or a generalization of it made to accommodate vector spaces of infinite dimension. What really gives life to the subject and makes it highly applicable is the study of maps between vector spaces, the so-called linear transformations.

Here are some problems that use linear algebra and that we shall address later in this course (perhaps in the assignments):

(1) An $m \times n$ matrix over $\mathbb{F}$ is an array

$$\begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix}, \qquad a_{ij} \in \mathbb{F}.$$

We shall see that *linear transformations and matrices are essentially the same thing.*

Consider a homogenous system of linear equations with coefficients in $\mathbb{F}$:

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= 0 \end{aligned}$$

This system can be encoded by the matrix

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}.$$

We shall see that *matrix manipulations and vector spaces techniques allow us to have a very good theory of solving a system of linear equations.*

(2) Consider a smooth function of 2 real variables, $f(x, y)$. The points where

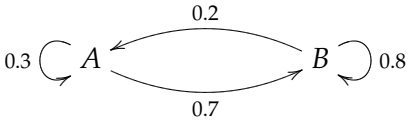
$$\frac{\partial f}{\partial x} = 0, \qquad \frac{\partial f}{\partial y} = 0,$$

are the extremum points. But is such a point a maximum, minimum or a saddle point? Perhaps none of these? To answer that one defines the **Hessian matrix**,

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}.$$

If this matrix is "negative definite", resp. "positive definite", we have a maximum, resp. minimum. Those are *algebraic concepts* that we shall define and study in this course. We will then also be able to say when we have a saddle point.

(3) The following example is a very special case of what is called a **Markov chain**. Imagine a system that has two states A, B, where the system changes its state every second (say), with given probabilities. For example,



Given that the system is initially with equal probability in any of the states, we'd like to know the long term behavior. For example, what is the probability that the system is in state B after a year. If we let

$$M = \begin{pmatrix} 0.3 & 0.2 \\ 0.7 & 0.8 \end{pmatrix},$$

then the question is what is

$$M^{60\times 60\times 24\times 365} \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}?$$

While modern computers (or even smart phones) have no problem making this calculation by brute-force, we can imagine systems with 10^6 particles[1] evolving over a billion years and then a brute-force approach is certainly not the best. Thus, we are led to ask if there is a fast method to perform such calculations.

(4) Consider a sequence defined by recurrence. A very famous example is the **Fibonacci sequence**:

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \dots$$

[1]The number of water molecules in a tea cup exceeds 10^{23} so for physical applications 10^6 is really a modest size number.

It is defined by the recurrence

$$a_0 = 1, \quad a_1 = 1, \quad a_{n+2} = a_n + a_{n+1}, \quad n \geq 0.$$

If we let

$$M = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix},$$

then

$$\begin{pmatrix} a_n \\ a_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a_{n-1} \\ a_n \end{pmatrix} = \cdots = M^n \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}.$$

We see again that the issue is to find a formula for M^n.

(5) Consider a **graph** G. By that we mean a finite set of vertices $V(G)$ and a subset $E(G) \subset V(G) \times V(G)$, which is symmetric: $(u, v) \in E(G) \Leftrightarrow (v, u) \in E(G)$. The elements of $E(G)$ are the edges. (It follows from our definition that there is at most one edge between any two vertices u, v. So we really are considering a special kind of graphs.) We shall also assume that the graph is **simple**: (u, u) is never in $E(G)$.

To a graph we can associate its **adjacency matrix**

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}, \qquad a_{ij} = \begin{cases} 1 & (i,j) \in E(G) \\ 0 & (i,j) \notin E(G). \end{cases}$$

It is a symmetric matrix whose entries are $0, 1$. The algebraic properties of the adjacency matrix teach us about the graph. For example, we shall see that one can read off if the graph is connected, or bipartite, from algebraic properties of the adjacency matrix. If we think about the vertices as trees and the edges indicate if a fire can spread from one tree to the next, the eigenvalues of the adjacency matrix inform us how quickly a fire will spread through the graph.

(6) The goal of **coding theory** is *to communicate data over a noisy channel*. The main idea is to associate to a message m a uniquely determined element $c(m)$ of a subspace $C \subseteq \mathbb{F}_2^n$. The subspace C is called a **linear code**. Typically, the number of digits required to write $c(m)$, the code word associated to m, is much larger than that required to write m itself, but by means of this redundancy something is gained.

Define the **Hamming distance** of two elements u, v in $\mathbb{F}_2^n$ as

$$d(u, v) = \text{no. of digits in which } u \text{ and } v \text{ differ.}$$

We also call $d(0, u)$ the Hamming **weight** $w(u)$ of u. Thus, $d(u, v) = w(u - v)$. We wish to find linear codes C such that

$$w(C) := \min\{w(u) : u \in C \setminus \{0\}\},$$

is large. The receiver of the message $c(m)$ can tell what was the original message m by reversing the encoding process. But, in practice, due to errors in transmission, the receiver may not receive $c(m)$ but rather a string of bits $c(m)'$ that is not far off, hopefully, from $c(m)$. The receiver is thus looking for the element of C closest to the message $c(m)'$ received; very likely that element is indeed $c(m)$. The larger $w(C)$ is, the more likely it is that the receiver found correctly the original transmission $c(m)$ and thus the message m.

The mathematics begins in finding codes C with a large weight $w(C)$ that nonetheless fill up much of the space $\mathbb{F}_2^n$ (otherwise, one can show, that the rate of data transmission is very low). Such codes enable deep space NASA missions to transmit data from a distance of billions of miles using, essentially, the power of a little battery.

(7) Let $y(t)$ be a real differentiable function and let $y^{(n)}(t) = \frac{\partial^n y}{\partial t^n}$. The **ordinary differential equation**:

$$y^{(n)}(t) = a_{n-1} \cdot y^{(n-1)}(t) + \cdots + a_1 \cdot y^{(1)}(t) + a_0 \cdot y(t)$$

where the a_i are some real numbers, can be translated into a system of linear differential equations. Let $f_i(t) = y^{(i)}(t)$ then

$$\begin{aligned} f_0' &= f_1 \\ f_1' &= f_2 \\ &\vdots \\ f_{n-1}' &= a_{n-1}f_{n-1} + \cdots + a_1 f_1 + a_0 f_0. \end{aligned}$$

More generally, given functions $g_1, \ldots, g_n$, we may consider the system of differential equations:

$$\begin{aligned} g_1' &= a_{11}g_1 + \cdots + a_{1n}g_n \\ &\vdots \\ g_n' &= a_{n1}g_1 + \cdots + a_{nn}g_n. \end{aligned}$$

It turns out that the matrix

$$\begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix}$$

determines the solutions uniquely, and effectively. In the example above, the matrix is

$$\begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & & & & \\ 0 & & \ldots & & 1 \\ a_0 & a_1 & a_2 & \ldots & a_{n-1} \end{pmatrix}.$$

Part 1. **Vector spaces**

1. Vector spaces: key notions

1.1. **Definition of vector space and subspace.** Let $\mathbb{F}$ be a field.

Definition 1.1.1. A **vector space** V over $\mathbb{F}$ is a non-empty set together with two operations

$$V \times V \to V, \qquad (v_1, v_2) \mapsto v_1 + v_2,$$

and

$$\mathbb{F} \times V \to V, \qquad (\alpha, v) \mapsto \alpha v,$$

(sometimes we write $\alpha \cdot v$) such that V is an abelian group with identity element denoted 0_V and, in addition:

(1) $1v = v, \ \ \forall v \in V$;
(2) $(\alpha\beta)v = \alpha(\beta v), \ \ \forall \alpha, \beta \in \mathbb{F}, \forall v \in V$;
(3) $(\alpha + \beta)v = \alpha v + \beta v, \ \ \forall \alpha, \beta \in \mathbb{F}, \forall v \in V$;
(4) $\alpha(v_1 + v_2) = \alpha v_1 + \alpha v_2, \ \ \forall \alpha \in \mathbb{F}, v_1, v_2 \in V$.

The elements of V are called **vectors** and the elements of $\mathbb{F}$ are called **scalars**.

Here are some formal consequences:

(1) $\boxed{0_{\mathbb{F}} \cdot v = 0_V}$
This holds true because $0_{\mathbb{F}} \cdot v = (0_{\mathbb{F}} + 0_{\mathbb{F}})v = 0_{\mathbb{F}} \cdot v + 0_{\mathbb{F}} \cdot v$ and so $0_V = 0_{\mathbb{F}} \cdot v$.
(2) $\boxed{-1 \cdot v = -v}$
This holds true because $0_V = 0_{\mathbb{F}} \cdot v = (1 + (-1))v = 1 \cdot v + (-1) \cdot v = v + (-1) \cdot v$ and that shows that $-1 \cdot v$ is $-v$.
(3) $\boxed{\alpha \cdot 0_V = 0_V}$
Indeed, $\alpha \cdot 0_V = \alpha \cdot (0_V + 0_V) = \alpha \cdot 0_V + \alpha \cdot 0_V$.

Definition 1.1.2. A **subspace** W of a vector space V is a non-empty subset such that:

(1) $\forall w_1, w_2 \in W$ we have $w_1 + w_2 \in W$;
(2) $\forall \alpha \in \mathbb{F}, w \in W$ we have $\alpha w \in W$.

It follows from the definition that W is a vector space in its own right. Indeed, the consequences noted above show that W is a subgroup and the rest of the axioms follow immediately since they hold for V. We also note that we always have the trivial subspaces $\{0\}$ and V.

Example 1.1.3. The vector space $\mathbb{F}^n$. Let $n \geq 0$ be an integer.
We define

$$\mathbb{F}^n = \{(x_1, \ldots, x_n) : x_i \in \mathbb{F}\},$$

with coordinate-wise addition. Multiplication by a scalar is defined by

$$\alpha(x_1, \ldots, x_n) = (\alpha x_1, \ldots, \alpha x_n).$$

The axioms are easy to verify. (If $n = 0$ we understand by $\mathbb{F}^0$ a set of one element $\{0\}$ with $0 + 0 = 0$ and for all $\alpha \in \mathbb{F}$, $\alpha \cdot 0 = 0$.)

For example, for $n = 5$ we have that

$$W = \{(x_1, x_2, x_3, 0, 0) : x_i \in \mathbb{F}\}$$

is a subspace of $\mathbb{F}^5$. This can be generalized considerably.

Let $a_{ij} \in \mathbb{F}$ and let W be the set of vectors $(x_1, \ldots, x_n)$ such that

$$a_{11}x_1 + \cdots + a_{1n}x_n = 0,$$
$$\vdots$$
$$a_{m1}x_1 + \cdots + a_{mn}x_n = 0.$$

Then W is a subspace of $\mathbb{F}^n$.

Example 1.1.4. Polynomials of degree less than n.
Again, $\mathbb{F}$ is a field. We define $\mathbb{F}[t]_n$ to be

$$\mathbb{F}[t]_n = \{a_0 + a_1 t + \cdots + a_{n-1}t^{n-1} : a_i \in \mathbb{F}\}.$$

We also put $\mathbb{F}[t]_0 = \{0\}$. It is easy to check that this is a vector space under the usual operations on polynomials. Let $a \in \mathbb{F}$ and consider

$$W = \{f \in \mathbb{F}[t]_n : f(a) = 0\}.$$

Then W is a subspace. Another example of a subspace is given by

$$U = \{f \in \mathbb{F}[t]_n : f''(t) + 3f'(t) = 0\},$$

where if $f(t) = a_0 + a_1 t + \cdots + a_{n-1}t^{n-1}$ we let $f'(t) = a_1 + 2a_2 t + \cdots + (n-1)a_{n-1}t^{n-2}$ and similarly for f'' and so on.

Example 1.1.5. Continuous real functions.
Let V be the set of real continuous functions $f : \mathbb{R} \to \mathbb{R}$. We have the usual definitions:

$$(f+g)(x) = f(x) + g(x), \qquad (\alpha f)(x) = \alpha f(x).$$

Here are some examples of subspaces:

(1) The functions whose value at 5 is zero.
(2) The functions f satisfying $f(1) + 9f(\pi) = 0$.
(3) The functions that are differentiable.
(4) The functions f such that $\int_0^1 f(x)dx = 0$.

Proposition 1.1.6. *Let $W_1, W_2 \subset V$ be subspaces then*

$$W_1 + W_2 := \{w_1 + w_2 : w_i \in W_i\}$$

and

$$W_1 \cap W_2$$

are subspaces of V.

Proof. Let $x = w_1 + w_2, y = w_1' + w_2'$ with $w_i, w_i' \in W_i$. Then

$$x + y = (w_1 + w_1') + (w_2 + w_2').$$

We have $w_i + w_i' \in W_i$, because W_i is a subspace, so $x + y \in W_1 + W_2$. Also,

$$\alpha x = \alpha w_1 + \alpha w_2,$$

and $\alpha w_i \in W_i$, again because W_i is a subspace. It follows that $\alpha x \in W_1 + W_2$. Thus, $W_1 + W_2$ is a subspace.

As for $W_1 \cap W_2$, we already know it is a subgroup, hence closed under addition. If $x \in W_1 \cap W_2$ then $x \in W_i$ and so $\alpha x \in W_i$, $i = 1, 2$, because W_i is a subspace. Thus, $\alpha x \in W_1 \cap W_2$. $\square$

1.2. **Direct sum.** Let U and W be vector spaces over the same field $\mathbb{F}$. Let

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

We define addition and multiplication by scalar as

$$(u_1, w_1) + (u_2, w_2) = (u_1 + u_2, w_1 + w_2), \qquad \alpha(u, w) = (\alpha u, \alpha w).$$

It is easy to check that $U \oplus W$ is a vector space over $\mathbb{F}$. It is called the **direct sum** of U and W, or, if we need to be more precise, the **external direct sum** of U and W.

We consider the following situation: U, W are subspaces of a vector space V. Then, in general, $U + W$ (in the sense of Proposition 1.1.6) is different from the external direct sum $U \oplus V$, though there is a connection between the two constructions as we shall see in Theorem 3.4.1.

1.3. **Linear combinations, linear dependence and span.** Let V be a vector space over $\mathbb{F}$ and $S = \{v_i : i \in I, v_i \in V\}$ be a collection of elements of V, indexed by some index set I. Note that we may have $i \neq j$, but $v_i = v_j$.

Definition 1.3.1. A **linear combination** of the elements of S is an expression of the form

$$\alpha_1 v_{i_1} + \cdots + \alpha_n v_{i_n},$$

where the $\alpha_j \in \mathbb{F}$ and $v_{i_j} \in S$. If S is empty then the only linear combination is the empty sum, defined to be 0_V. We let the **span** of S be

$$\mathrm{Span}(S) = \left\{ \sum_{j=1}^{m} \alpha_j v_{i_j} : \alpha_j \in \mathbb{F}, i_j \in I \right\}.$$

Note that $\mathrm{Span}(S)$ is all the linear combinations one can form using the elements of S.

Example 1.3.2. Let S be the collection of vectors $\{(0,1,0), (1,1,0), (0,1,0)\}$, say in $\mathbb{R}^3$. The vector 0 is always a linear combination; in our case, $(0,0,0) = 0 \cdot (0,1,0) + 0 \cdot (1,1,0) + 0 \cdot (0,1,0)$, but also $(0,0,0) = 1 \cdot (0,1,0) + 0 \cdot (1,1,0) - 1 \cdot (0,1,0)$, which is a non-trivial linear combination. It is important to distinguish between the collection S and the collection $T = \{(0,1,0), (1,1,0)\}$. There is only one way to write $(0,0,0)$ using the elements of T, namely, $0 \cdot (0,1,0) + 0 \cdot (1,1,0)$.

Proposition 1.3.3. *The set* $\mathrm{Span}(S)$ *is a subspace of* V.

Proof. Let $\sum_{j=1}^{m} \alpha_j v_{i_j}$ and $\sum_{j=1}^{n} \beta_j v_{k_j}$ be two elements of $\mathrm{Span}(S)$. Since the α_j and β_j are allowed to be zero, we may assume that the same elements of S appear in both sums, by adding more vectors with zero coefficients if necessary. That is, we may assume we deal with two elements $\sum_{j=1}^{m} \alpha_j v_{i_j}$ and $\sum_{j=1}^{m} \beta_j v_{i_j}$. It is then clear that

$$\sum_{j=1}^{m} \alpha_j v_{i_j} + \sum_{j=1}^{m} \beta_j v_{i_j} = \sum_{j=1}^{m} (\alpha_j + \beta_j) v_{i_j},$$

is also an element of $\mathrm{Span}(S)$.

Let $\alpha \in \mathbb{F}$ then $\alpha \left(\sum_{j=1}^{m} \alpha_j \cdot v_{i_j} \right) = \sum_{j=1}^{m} \alpha\alpha_j \cdot v_{i_j}$ shows that $\alpha \left(\sum_{j=1}^{m} \alpha_j v_{i_j} \right)$ is also an element of $\mathrm{Span}(S)$. □

Definition 1.3.4. If $\mathrm{Span}(S) = V$, we call S a **spanning set**. If $\mathrm{Span}(S) = V$ and for every $T \subsetneq S$ we have $\mathrm{Span}(T) \subsetneq V$ we call S a **minimal spanning set**.

Example 1.3.5. Consider the set $S = \{(1,0,1), (0,1,1), (1,1,2)\}$. The span of S is $W = \{(x,y,z) : x + y - z = 0\}$. Indeed, W is a subspace containing S and so $\mathrm{Span}(S) \subset W$. On the other hand, if $(x,y,z) \in W$ then $(x,y,z) = x(1,0,1) + y(0,1,1)$ and so $W \subseteq \mathrm{Span}(S)$. Note that we have actually proven that $W = \mathrm{Span}(\{(1,0,1), (0,1,1)\})$ and so S is not a minimal spanning set for W. It is easy to check that $\{(1,0,1), (0,1,1)\}$ is a minimal spanning set for W.

Example 1.3.6. Let

$$\mathrm{St} = \{(1,0,0,\dots,0),\ (0,1,0,\dots,0),\dots,\ (0,\dots,0,1)\}.$$

Namely,

$$\mathrm{St} = \{e_1, e_2, \dots, e_n\},$$

where e_i is the vector all whose coordinates are 0, except for the i-th coordinate which is equal to 1. We claim that St is a minimal spanning set for $\mathbb{F}^n$. Indeed, since we have

$$\alpha_1 e_1 + \cdots + \alpha_n e_n = (\alpha_1, \dots, \alpha_n),$$

we deduce that: (i) every vector is in the span of St, that is, St is a spanning set; (ii) if $T \subseteq \mathrm{St}$ and $e_j \notin T$ then every vector in $\mathrm{Span}(T)$ has j-th coordinate equal to zero. So in particular $e_j \notin \mathrm{Span}(T)$ and so St is a minimal spanning set for $\mathbb{F}^n$.

Definition 1.3.7. Let $S = \{v_i : i \in I\}$ be a non-empty collection of vectors of a vector space V. We say that S is **linearly dependent** if there are $\alpha_j \in \mathbb{F}$, $j = 1, \dots, m$, not all zero and $v_{i_j} \in S$, $j = 1, \dots, m$, such that

$$\alpha_1 v_{i_1} + \cdots + \alpha_m v_{i_m} = 0.$$

Thus, S is **linearly independent** if

$$\alpha_1 v_{i_1} + \cdots + \alpha_m v_{i_m} = 0$$

for some $\alpha_j \in \mathbb{F}, v_{i_j} \in S$, implies $\alpha_1 = \cdots = \alpha_m = 0$.

Example 1.3.8. If $S = \{v\}$ then S is linearly dependent if and only if $v = 0$. If $S = \{v, w\}$ then S is linearly dependent if and only if one of the vectors is a multiple of the other.

Example 1.3.9. The set $\mathrm{St} = \{e_1, e_2, \dots, e_n\}$ in $\mathbb{F}^n$ is linearly independent. Indeed, if $\sum_{i=1}^n \alpha_i e_i = 0$ then $(\alpha_1, \dots, \alpha_n) = 0$ and so each $\alpha_i = 0$.

Example 1.3.10. The set $\{(1,0,1), (0,1,1), (1,1,2)\}$ is linearly dependent. Indeed, $(1,0,1) + (0,1,1) - (1,1,2) = (0,0,0)$.

Definition 1.3.11. A collection of vectors S of a vector space V is called a **maximal linearly independent** set if S is independent and for every $v \in V$ the collection $S \cup \{v\}$ is linearly dependent.

Example 1.3.12. The set $\mathrm{St} = \{e_1, e_2, \dots, e_n\}$ in $\mathbb{F}^n$ is a maximal linearly independent. Indeed, given any vector $v = (\alpha_1, \dots, \alpha_n)$, the collection $\{e_1, \dots, e_n, v\}$ is linearly dependent as we have

$$\alpha_1 e_1 + \alpha_2 e_2 + \cdots + \alpha_n e_n - v = 0.$$

1.3.1. *Geometric interpretation.* Suppose that $S = \{v_1, \dots, v_k\}$ is a linearly independent collection of vectors in $\mathbb{F}^n$. Then $v_1 \neq 0$ (else $1 \cdot v_1 = 0$ shows S is linearly dependent). Next, $v_2 \notin \mathrm{Span}(\{v_1\})$, else $v_2 = \alpha_1 v_1$ for some α_1 and we get a linear dependence $\alpha_1 v_1 - v_2 = 0$. Then, $v_3 \notin \mathrm{Span}(\{v_1, v_2\})$, else $v_3 = \alpha_1 v_1 + \alpha_2 v_2$, etc. We conclude the following:

Proposition 1.3.13. *$S = \{v_1, \dots, v_k\}$ is linearly independent if and only if $v_1 \neq 0$ and for any i we have $v_i \notin \mathrm{Span}(\{v_1, \dots, v_{i-1}\})$.*

1.4. **Spanning and independence.** We keep the notation of the previous section 1.3. Thus, V is a vector space over $\mathbb{F}$ and $S = \{v_i : i \in I, v_i \in V\}$ is a collection of elements of V, indexed by some index set I.

Lemma 1.4.1. *We have $v \in \mathrm{Span}(S)$ if and only if $\mathrm{Span}(S) = \mathrm{Span}(S \cup \{v\})$.*

The proof of the lemma is left as an exercise.

Theorem 1.4.2. *Let V be a vector space over a field $\mathbb{F}$ and $S = \{v_i : i \in I\}$ a collection of vectors in V. The following are equivalent:*

(1) S is a minimal spanning set.
(2) S is a maximal linearly independent set.
(3) Every vector v in V can be written as a unique linear combination of elements of S:

$$v = \alpha_1 v_{i_1} + \cdots + \alpha_m v_{i_m},$$

$\alpha_j \in \mathbb{F}$, non-zero, $v_{i_j} \in S$ ($i_1, \ldots, i_m$ distinct).[2]

Proof. We shall show (1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (1).

(1) $\Rightarrow$ (2)

We first prove S is independent. Suppose that $\alpha_1 v_{i_1} + \cdots + \alpha_m v_{i_m} = 0$, with $\alpha_i \in \mathbb{F}$ and $i_1, \ldots, i_m$ distinct indices, is a linear dependence. Then some $\alpha_i \neq 0$ and we may assume w.l.o.g. that $\alpha_1 \neq 0$. Then,

$$v_{i_1} = -\alpha_1^{-1}\alpha_2 v_{i_2} - \alpha_1^{-1}\alpha_3 v_{i_3} - \cdots - \alpha_1^{-1}\alpha_m v_{i_m},$$

and so $v_{i_1} \in \mathrm{Span}(S \setminus \{v_{i_1}\})$ and thus $\mathrm{Span}(S \setminus \{v_{i_1}\}) = \mathrm{Span}(S)$ by Lemma 1.4.1. This proves S is not a minimal spanning set, which is a contradiction.

Next, we show that S is a maximal independent set. Let $v \in V$. Then we have $v \in \mathrm{Span}(S)$ so $v = \alpha_1 v_{i_1} + \cdots + \alpha_m v_{i_m}$ for some $\alpha_j \in \mathbb{F}$, $v_{i_j} \in S$ ($i_1, \ldots, i_m$ distinct). We conclude that $\alpha_1 v_{i_1} + \cdots + \alpha_m v_{i_m} - v = 0$ and so that $S \cup \{v\}$ is linearly dependent.

(2) $\Rightarrow$ (3)

Let $v \in V$. Since $S \cup \{v\}$ is linearly dependent, we have

$$\alpha_1 v_{i_1} + \cdots + \alpha_m v_{i_m} + \beta v = 0,$$

for some $\alpha_j \in \mathbb{F}, \beta \in \mathbb{F}$ not all zero, $v_{i_j} \in S$ ($i_1, \ldots, i_m$ distinct). Note that we cannot have $\beta = 0$ because that would show that S is linearly dependent. Thus, we get

$$v = -\beta^{-1}\alpha_1 v_{i_1} - \beta^{-1}\alpha_2 v_{i_2} - \cdots - \beta^{-1}\alpha_m v_{i_m}.$$

We next show such an expression is unique. Suppose that we have two such expressions, then, by adding zero coefficients, we may assume that the same elements of S are involved, that is

$$v = \sum_{j=1}^{m} \beta_j v_{i_j} = \sum_{j=1}^{m} \gamma_j v_{i_j},$$

where for some j we have $\beta_j \neq \gamma_j$. We then get that

$$0 = \sum_{i=j}^{m} (\beta_j - \gamma_j) v_{i_j}.$$

Since S is linearly independent all the coefficients must be zero, that is, $\beta_j = \gamma_j, \forall j$.

(3) $\Rightarrow$ (1)

Firstly, the fact that every vector v has such an expression is precisely saying that $\mathrm{Span}(S) = V$ and so that S is a spanning set. We next show that S is a minimal spanning set. Suppose not. Then for some index $i \in I$ we have $\mathrm{Span}(S) = \mathrm{Span}(S - \{v_i\})$. That means, that

$$v_i = \alpha_1 v_{i_1} + \cdots + \alpha_m v_{i_m},$$

[2] As the zero vector 0_V is equal to the empty sum, this also says that there is no way to write 0_V as a non-trivial linear combinations of the set S.

for some $\alpha_j \in \mathbb{F}$, $v_{i_j} \in S - \{v_i\}$ ($i_1, \ldots, i_m$ distinct). But also

$$v_i = v_i.$$

That is, we get two ways to express v_i as a linear combination of the elements of S. This is a contradiction. $\square$

2. Basis and dimension

Definition 2.0.1. Let $S = \{v_i : i \in I\}$ be a collection of vectors of a vector space V over a field $\mathbb{F}$. We say that S is a **basis** if it satisfies one of the equivalent conditions of Theorem 1.4.2. Namely, S is a minimal spanning set, or a maximal independent set, or every vector can be written uniquely as a linear combination of the elements of S.

Example 2.0.2. Let $\mathrm{St} = \{e_1, e_2, \dots, e_n\}$ be the set appearing in Examples 1.3.6, 1.3.9 and 1.3.12. Then S is a basis of $\mathbb{F}^n$ called the **standard basis**.

The main theorem is the following:

Theorem 2.0.3. *Every vector space has a basis. Any two bases of V have the same cardinality.*

Based on the theorem, we can make the following definition.

Definition 2.0.4. The cardinality of (any) basis of V is called its **dimension**.

We do not have the tools to prove Theorem 2.0.3. We are lacking knowledge of how to deal with infinite cardinalities effectively. The existence of a basis is a rather easy consequence of Zorn's lemma and is explained in an appendix. We shall only prove the following weaker theorem.

Theorem 2.0.5. *Assume that V has a basis $S = \{s_1, s_2, \dots, s_n\}$ of finite cardinality n. Then every other basis of V has n elements as well.*

Remark 2.0.6. There are definitely vector spaces of infinite dimension. For example, the vector space V of continuous real functions $\mathbb{R} \to \mathbb{R}$ (see Example 1.1.5) has infinite dimension (exercise). Also, the set of infinite vectors

$$\mathbb{F}^\infty = \{(\alpha_1, \alpha_2, \alpha_3, \dots) : \alpha_i \in \mathbb{F}\},$$

with the coordinate-wise addition and $\alpha(\alpha_1, \alpha_2, \alpha_3, \dots) = (\alpha\alpha_1, \alpha\alpha_2, \alpha\alpha_3, \dots)$ is a vector space of infinite dimension (exercise).

2.1. Steinitz's substitution lemma.

Lemma 2.1.1. *Let $A = \{v_1, v_2, v_3, \dots\}$ be a list of vectors of V. Then A is linearly dependent if and only if one of the vectors of A is a linear combination of the preceding ones.*

This lemma is essentially the same as Proposition 1.3.13. Still, we supply a proof.

Proof. Clearly, if $v_{k+1} = \alpha_1 v_1 + \cdots + \alpha_k v_k$ then A is linearly dependent since we then have the non-trivial linear dependence $\alpha_1 v_1 + \cdots + \alpha_k v_k - v_{k+1} = 0$.

Conversely, if we have a linear dependence

$$\alpha_1 v_{i_1} + \cdots + \alpha_k v_{i_k} = 0$$

with some $\alpha_i \neq 0$, we may assume that each $\alpha_i \neq 0$ and also that $i_1 < i_2 < \cdots < i_k$. We then find that

$$v_{i_k} = -\alpha_k^{-1}\alpha_1 v_{i_1} - \cdots - \alpha_k^{-1}\alpha_{k-1} v_{i_{k-1}}$$

is a linear combination of the preceding vectors. □

Lemma 2.1.2. *(Steinitz) Let $A = \{v_1, \dots, v_n\}$ be a linearly independent set and let $B = \{w_1, \dots, w_m\}$ be another linearly independent set. Suppose that $m \geq n$. Then, for every j, $0 \leq j \leq n$, we may re-number the elements of B such that the set*

$$\{v_1, v_2, \dots, v_j, w_{j+1}, w_{j+2}, \dots, w_m\}$$

is linearly independent.

Remark 2.1.3. The lemma says that for every $j \leq n$ we can *substitute* j elements of B by the elements $v_1, \ldots, v_j$ of A and retain independence.

Proof. We prove the Lemma by induction on j. For $j = 0$ the claim is just that B is linearly independent, which is given.

Assume the result for j. Thus, we have re-numbered the elements of B so that

$$\{v_1, \ldots, v_j, w_{j+1}, \ldots, w_m\}$$

is linearly independent. Suppose that $j < n$ (else,we are done). Consider the list

$$\{v_1, \ldots, v_j, v_{j+1}, w_{j+1}, \ldots, w_m\}$$

If this list is linearly independent, omit w_{j+1} and then

$$\{v_1, \ldots, v_j, v_{j+1}, w_{j+2}, \ldots, w_m\}$$

is linearly independent and we are done. Else,

$$\{v_1, \ldots, v_j, v_{j+1}, w_{j+1}, \ldots, w_m\}$$

is linearly dependent and so by Lemma 2.1.1 one of the vectors is a linear combination of the preceding ones. Since $\{v_1, \ldots, v_{j+1}\}$ is linearly independent that vector must be one of the w's. Thus, there is some minimal $r \geq 1$ such that w_{j+r} is a linear combination of the previous vectors. So,

$$w_{j+r} = \alpha_1 v_1 + \cdots + \alpha_{j+1} v_{j+1} + \beta_{j+1} w_{j+1} + \cdots + \beta_{j+r-1} w_{j+r-1}. \tag{1}$$

We claim that

$$\{v_1, \ldots, v_j, v_{j+1}, w_{j+1}, \ldots, \widehat{w_{j+r}}, \ldots, w_m\}$$

is linearly independent.

If not, then we have a non-trivial linear relation:

$$\gamma_1 v_1 + \cdots + \gamma_{j+1} v_{j+1} + \delta_{j+1} w_{j+1} + \cdots + \widehat{\delta_{j+r} w_{j+r}} + \cdots + \delta_m w_m = 0. \tag{2}$$

Note that $\gamma_{j+1} \neq 0$, because we know that $\{v_1, v_2, \ldots, v_j, w_{j+1}, w_{j+2}, \ldots, w_m\}$ is linearly independent. Thus, using Equation (2), we get for some ϵ_i, η_i that

$$v_{j+1} = \epsilon_1 v_1 + \cdots + \epsilon_j v_j + \eta_{j+1} w_{j+1} + \cdots + \widehat{\eta_{j+r} w_{j+r}} + \cdots + \eta_m w_m.$$

Substitute this in Equation (1) and obtain that w_{j+r} is a linear combination of the vectors $\{v_1, \ldots, v_j, w_{j+1}, \ldots, \widehat{w_{j+r}}, \ldots, w_m\}$ which is a contradiction to the induction hypothesis. Thus,

$$\{v_1, \ldots, v_j, v_{j+1}, w_{j+1}, \ldots, \widehat{w_{j+r}}, \ldots, w_m\}$$

is linearly independent. Now, rename the elements of B so that w_{j+r} becomes w_{j+1}. □

Remark 2.1.4. The use of Lemma 2.1.1 in the proof of Steinitz's substitution lemma is not essential. It is convenient in that it tells us exactly which vector needs to be taken out in order the continue the construction. For a concrete application of Steinitz's lemma see Example 2.2.3 below.

2.2. **Proof of Theorem 2.0.5.**

Proof. Let $S = \{s_1, \ldots, s_n\}$ be a basis of finite cardinality of V. Let T be another basis and suppose that there are more than n elements in T. Then we may choose $t_1, \ldots, t_{n+1}$, elements of T, such that $t_1, \ldots, t_{n+1}$ are linearly independent. By Steinitz's Lemma, we can re-number the t_i such that $\{s_1, \ldots, s_n, t_{n+1}\}$ is linearly independent, which implies that S is not a maximal independent set. Contradiction. Thus, any basis of V has at most n elements. However, suppose that T has less than n elements. Reverse the role of S and T in the argument above. We get again a contradiction. Thus, all bases have the same cardinality. □

The proof of the theorem also shows the following

Lemma 2.2.1. *Let V be a vector space of finite dimension n. Let $T = \{t_1, \ldots, t_a\}$ be a linearly independent set. Then $a \leq n$.*

(Take S to be a basis of V and run through the argument above.) We conclude:

Corollary 2.2.2. *Any independent set of vectors of V (a vector space of finite dimension n) can be completed to a basis.*

Proof. Let $S = \{s_1, \ldots, s_a\}$ be an independent set. Then $a \leq n$. If $a < n$ then S cannot be a maximal independent set and so there's a vector s_{a+1} such that $\{s_1, \ldots, s_a, s_{a+1}\}$ is linearly independent. And so on. The process stops when we get an independent set $\{s_1, \ldots, s_a, \ldots, s_n\}$ of n vectors. Such a set must be maximal independent set (else we would get that there is a set of $n+1$ independent vectors) and so a basis. □

Example 2.2.3. Consider the vector space $\mathbb{F}^n$ and a set $B = \{b_1, \ldots, b_a\}$ of linearly independent vectors. We know that B can be completed to a basis of $\mathbb{F}^n$, but is there a more explicit method of doing that? Steinitz's Lemma does just that. Take the standard basis $\mathrm{St} = \{e_1, \ldots, e_n\}$ (or any other basis if you like). Then, Steinitz's Lemma implies the following. There is a choice of $n - a$ indices $i_1, \ldots, i_{n-a}$ such that

$$b_1, \ldots, b_a, e_{i_1}, \ldots, e_{i_{n-a}},$$

is a basis for $\mathbb{F}^n$. More than that, the Lemma tells us how to choose the basis elements to be added. Namely,

(1) Let $B = \{s_1, \ldots, s_a, e_1\}$ and $S = \{e_1, \ldots, e_n\}$.
(2) If $\{b_1, \ldots, b_a, e_1\}$ is linearly independent (this happens if and only if $e_1 \notin \mathrm{Span}(\{s_1, \ldots, s_a\})$ then let $B = \{b_1, \ldots, b_a, e_1\}$ and $S = \{e_2, \ldots, e_n\}$ and repeat this step with the new B, S and the first vector in S.
(3) If $\{b_1, \ldots, b_a, e_1\}$ is linearly dependent let $S = \{e_2, \ldots, e_n\}$ and, keeping the same B go to the previous step and perform it with these B, S and the first vector in S.

Corollary 2.2.4. *Let $W \subset V$ be a subspace of a finite dimensional vector space V. Then* $\dim(W) \leq \dim(V)$ *and*

$$W = V \Leftrightarrow \dim(W) = \dim(V).$$

Proof. Any independent set T of vectors of W is an independent set of vectors of V and so can be completed to a basis of V. In particular, a basis of W can be completed to a basis of V and so $\dim(W) \leq \dim(V)$.

Now, clearly $W = V$ implies $\dim(W) = \dim(V)$. Suppose that $W \neq V$ and choose a basis for W, say $\{t_1, \ldots, t_m\}$. Then, there's a vector $v \in V$ which is not a linear combination of the $\{t_i\}$ and we see that $\{t_1, \ldots, t_m, v\}$ is a linearly independent set in V. It follows that $\dim(V) \geq m + 1 > m = \dim(W)$. □

Example 2.2.5. Let $V_i, i = 1, 2$ be two finite dimensional vector spaces over $\mathbb{F}$. Then (exercise)

$$\dim(V_1 \oplus V_2) = \dim(V_1) + \dim(V_2).$$

2.3. **Coordinates and change of basis.**

Definition 2.3.1. Let V be a finite dimensional vector space over $\mathbb{F}$. Let

$$B = \{b_1, \ldots, b_n\}$$

be a basis of V. Then any vector v can be written uniquely in the form

$$v = \alpha_1 b_1 + \cdots + \alpha_n b_n,$$

where $\alpha_i \in \mathbb{F}$ for all i. The α_i are called the **coordinates** of v with respect to the basis B and we use the notation

$$[v]_B = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

Note that the coordinates depend on the order of the elements of the basis. Thus, whenever we talk about a basis $\{b_1, \ldots, b_n\}$ we think about that as a list of vectors.

Example 2.3.2. We may think about the vector space $\mathbb{F}^n$ as

$$\mathbb{F}^n = \left\{ \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} : \alpha_i \in \mathbb{F} \right\}.$$

Addition is done coordinate wise and in this notation we have

$$\alpha \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \alpha\alpha_1 \\ \vdots \\ \alpha\alpha_n \end{pmatrix}.$$

Let St be the standard basis $\{e_1, \ldots, e_n\}$, where

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \leftarrow i.$$

If $v = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$ is an element of $\mathbb{F}^n$ then of course $v = \alpha_1 e_1 + \cdots + \alpha_n e_n$ and so

$$[v]_{\mathrm{St}} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

Example 2.3.3. Let $V = \mathbb{R}^2$ and $B = \{(1, 1), (1, -1)\}$. Let $v = (5, 1)$. Then $v = 3(1, 1) + 2(1, -1)$. Thus,

$$[v]_B = \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

Conversely, if

$$[v]_B = \begin{pmatrix} 2 \\ 12 \end{pmatrix}$$

then $v = 2(1,1) + 12(1,-1) = (14,-10)$.

2.3.1. *Change of basis.* Suppose that B and C are two bases, say

$$B = \{b_1, \dots, b_n\}, \quad C = \{c_1, \dots, c_n\}.$$

We would like to determine the relation between $[v]_B$ and $[v]_C$. Let,

$$\begin{aligned} b_1 &= m_{11}c_1 + \cdots + m_{n1}c_n \\ &\vdots \\ b_j &= m_{1j}c_1 + \cdots + m_{nj}c_n \\ &\vdots \\ b_n &= m_{1n}c_1 + \cdots + m_{nn}c_n, \end{aligned} \tag{3}$$

and let

$${}_CM_B = \begin{pmatrix} m_{11} & \dots & m_{1n} \\ \vdots & & \vdots \\ m_{n1} & \dots & m_{nn} \end{pmatrix}.$$

Theorem 2.3.4. *We have*

$$[v]_C = {}_CM_B[v]_B.$$

We first prove a lemma.

Lemma 2.3.5. *We have the following identities:*

$$[v]_B + [w]_B = [v+w]_B, \qquad [\alpha v]_B = \alpha[v]_B.$$

Proof. This follows immediately from the fact that if

$$v = \sum \alpha_i b_i, \quad w = \sum \beta_i b_i,$$

then

$$v + w = \sum(\alpha_i + \beta_i)b_i, \quad \alpha v = \sum \alpha\alpha_i \cdot b_i.$$

□

Proof. (Of theorem). It follows from the Lemma that it is enough to prove

$$[v]_C = {}_CM_B[v]_B$$

for v running over a basis of V. We take the basis B itself. Then,

$$\begin{aligned} {}_CM_B[b_j]_B &= {}_CM_Be_j \\ &= j\text{-th column of } {}_CM_B \\ &= \begin{pmatrix} m_{1j} \\ \vdots \\ m_{nj} \end{pmatrix} \\ &= [b_j]_C \end{aligned}$$

(cf. Equation 3). □

Lemma 2.3.6. *Let M be a matrix such that*

$$[v]_C = M[v]_B,$$

for every $v \in V$. Then,

$$M = {}_C M_B.$$

Proof. Since

$$[b_j]_C = M[b_j]_B = Me_j = j\text{-th column of } M,$$

the columns of M are uniquely determined. □

Corollary 2.3.7. *Let B, C, D be bases. Then:*

(1) ${}_B M_B = I_n$ (the identity $n \times n$ matrix).
(2) ${}_D M_B = {}_D M_C \, {}_C M_B$.
(3) The matrix ${}_C M_B$ is invertible and ${}_C M_B = {}_B M_C^{-1}$.

Proof. For (1) we note that

$$[v]_B = I_n [v]_B,$$

and so, by Lemma 2.3.6, $I_n = {}_B M_B$.

We use the same idea for (2). We have

$$[v]_D = {}_D M_C [v]_C = {}_D M_C ({}_C M_B [v]_B) = ({}_D M_C \, {}_C M_B)[v]_B,$$

and so ${}_D M_B = {}_D M_C \, {}_C M_B$.

For (3) we note that by (1) and (2) we have

$${}_C M_B \, {}_B M_C = {}_C M_C = I_n, \qquad {}_B M_C \, {}_C M_B = {}_B M B = I_n,$$

and so ${}_C M_B$ and ${}_B M_C$ are invertible and are each other's inverse. □

Example 2.3.8. Here is a general principle: if $B = \{b_1, \ldots, b_n\}$ is a basis of $\mathbb{F}^n$ then each b_i is already given by coordinates relative to the standard basis. Say,

$$b_j = \begin{pmatrix} m_{1j} \\ \vdots \\ m_{nj} \end{pmatrix}.$$

Then the matrix $M = (m_{ij})$ obtained by writing the basis elements of B as column vectors one next to the other is the matrix ${}_S M_B$. Since,

$${}_B M_S = {}_{\mathrm{St}} M_B^{-1},$$

this gives a useful method to pass from coordinates relative to the standard basis to coordinates relative to the basis B.

For example, consider the basis $B = \{(5,1), (3,2)\}$ of $\mathbb{R}^2$. Then

$${}_B M_{\mathrm{St}} = ({}_{\mathrm{St}} M_B)^{-1} = \begin{pmatrix} 5 & 3 \\ 1 & 2 \end{pmatrix}^{-1} = \frac{1}{7} \begin{pmatrix} 2 & -3 \\ -1 & 5 \end{pmatrix}.$$

Thus, the vector $(2,3)$ has coordinates $\frac{1}{7} \begin{pmatrix} 2 & -3 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -5/7 \\ 13/7 \end{pmatrix}$. Indeed, $\frac{-5}{7}(5,1) + \frac{13}{7}(3,2) = (2,3)$.

Let $C = \{(2,2), (1,0)\}$ be another basis. To pass from coordinates relative to the basis C to coordinates relative to the basis B we use the matrix

$${}_B M_C = {}_B M_{\mathrm{St}} \, {}_{\mathrm{St}} M_C = \frac{1}{7} \begin{pmatrix} 2 & -3 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 2 & 0 \end{pmatrix} = \frac{1}{7} \begin{pmatrix} -2 & 2 \\ 8 & -1 \end{pmatrix}.$$

3. Linear transformations

Definition 3.0.1. Let V and W be two vector spaces over a field $\mathbb{F}$. A **linear transformation**

$$T : V \longrightarrow W,$$

is a function $T : V \to W$ such that

(1) $T(v_1 + v_2) = T(v_1) + T(v_2)$ for all $v_1, v_2 \in V$;
(2) $T(\alpha v) = \alpha T(v)$ for all $v \in V, \alpha \in \mathbb{F}$.

(A linear transformation is also called a linear map, or mapping, or application.)

Here are some formal consequences of the definition:

(1) $\boxed{T(0_V) = 0_W}$ Indeed, since T is a homomorphism of (abelian groups) we already know that. For the same reason we know that:
(2) $\boxed{T(-v) = -T(v)}$
(3) $\boxed{T(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 T(v_1) + \alpha_2 T(v_2)}$

Lemma 3.0.2. $\mathrm{Ker}(T) = \{v \in V : T(v) = 0_W\}$ *is a subspace of* V *and* $\mathrm{Im}(T)$ *is a subspace of* W.

Proof. We already know $\mathrm{Ker}(T), \mathrm{Im}(T)$ are subgroups and so closed under addition. Next, if $\alpha \in \mathbb{F}, v \in \mathrm{Ker}(T)$ then $T(\alpha v) = \alpha T(v) = \alpha 0_W = 0_W$ and so $\alpha v \in \mathrm{Ker}(T)$ as well. If $w \in \mathrm{Im}(T)$ then $w = T(v)$ for some $v \in V$. It follows that $\alpha w = \alpha T(v) = T(\alpha v)$ is also in $\mathrm{Im}(T)$. □

Remark 3.0.3. From the theory of groups we know that T is injective if and only if $\mathrm{Ker}(T) = \{0_V\}$.

Example 3.0.4. The zero map $T : V \to W$, $T(v) = 0_W$ for every $v \in V$, is a linear map with kernel V and image $\{0_W\}$.

Example 3.0.5. The identity map $\mathrm{Id} : V \to V$, $\mathrm{Id}(v) = v$ for all $v \in V$, is a linear map with kernel $\{0\}$ and image V. More generally, if $V \subset W$ is a subspace and $i : V \to W$ is the inclusion map, $i(v) = v$, then i is a linear map with kernel $\{0\}$ and image V.

Example 3.0.6. Let $B = \{b_1, \ldots, b_n\}$ be a basis for V and let fix some $1 \leq j \leq n$. Let

$$T : V \to V, \qquad T(\alpha_1 b_1 + \cdots + \alpha_n b_n) = \alpha_{j+1} b_{j+1} + \alpha_{j+2} b_{j+2} + \cdots + \alpha_n b_n.$$

(To understand the definition for $j = n$, recall that the empty sum is by definition equal to 0.) The kernel of T is $\mathrm{Span}(\{b_1, \ldots, b_j\})$ and $\mathrm{Im}(T) = \mathrm{Span}(\{b_{j+1}, b_{j+2}, \ldots, b_n\})$.

Example 3.0.7. Let $V = \mathbb{F}^n, W = \mathbb{F}^m$, written as column vectors. Let $A = (a_{ij})$ be an $m \times n$ matrix with entries in $\mathbb{F}$. Define

$$T : \mathbb{F}^n \to \mathbb{F}^m,$$

be the following formula

$$T \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = A \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

Then T is a linear map. This follows from identities for matrix multiplication:

$$A \begin{pmatrix} \alpha_1 + \beta_1 \\ \vdots \\ \alpha_n + \beta_n \end{pmatrix} = A \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} + A \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}, \qquad A \begin{pmatrix} \alpha\alpha_1 \\ \vdots \\ \alpha\alpha_n \end{pmatrix} = \alpha A \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

Those identities are left as an exercise. We note that $\mathrm{Ker}(T)$ are the solutions for the following homogenous system of linear equations:

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= 0. \end{aligned}$$

The image of T is precisely the vectors $\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}$ for which the following inhomogenous system of linear equations has a solution:

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= \beta_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= \beta_m. \end{aligned}$$

Example 3.0.8. Let $V = \mathbb{F}[t]_n$, the space of polynomials of degree less than n. Define

$$T : V \to V, \qquad T(f) = f',$$

the formal derivative of f. Then T is a linear map. We leave the description of the kernel and image of T as an exercise.

The following Proposition is very useful. Its proof is left as an exercise.

Proposition 3.0.9. *Let V and W be vector spaces over $\mathbb{F}$. Let $B = \{b_1, \ldots, b_n\}$ be a basis for V and let $t_1, \ldots, t_n$ be any elements of W. There is a unique linear map*

$$T : V \to W,$$

such that

$$T(b_i) = t_i, \qquad i = 1, \ldots, n.$$

The following lemma is left as an exercise.

Lemma 3.0.10. *Let V, W be vector spaces over $\mathbb{F}$. Let*

$$\mathrm{Hom}(V, W) = \{T : V \to W : T \text{ is a linear map}\}.$$

Then $\mathrm{Hom}(V, W)$ is a vector space in its own right where we define for two linear transformations S, T and scalar α the linear transformations $S + T, \alpha S$ as follows:

$$(S + T)(v) = S(v) + T(v), \qquad (\alpha S)(v) = \alpha S(v).$$

In addition, if $T : V \to W$ and $R : W \to U$ are linear maps, where U is a third vector space over $\mathbb{F}$, then

$$R \circ T : V \to U$$

is a linear map.

3.1. **Isomorphisms.** Let $T : V \to W$ be an injective linear map. One also says that T is non-singular. If T is not injective, one says also that it is singular. T is called an **isomorphism** if it is bijective. In that case, the inverse map

$$S = T^{-1} : W \to V$$

is also an isomorphism. Indeed, from the theory of groups we already know it is a group isomorphism. Next, to check that $S(\alpha w) = \alpha S(w)$ it is enough to check that $T(S(\alpha w)) = T(\alpha S(w))$. But, $T(S(\alpha w)) = \alpha w$ and $T(\alpha S(w)) = \alpha T(S(w)) = \alpha w$ too.

As in the case of groups it follows readily from the properties above that being isomorphic is an equivalence relation on vector spaces. We use the notation

$$V \cong W$$

to denote that V is isomorphic to W.

Theorem 3.1.1. *Let V be a vector space of dimension n over a field $\mathbb{F}$ then*

$$V \cong \mathbb{F}^n.$$

Proof. Let $B = \{b_1, \ldots, b_n\}$ be any basis of V. Define a function

$$T : V \longrightarrow \mathbb{F}^n, \qquad T(v) = [v]_B.$$

The formulas we have established in Lemma 2.3.5, $[v + w]_B = [v]_B + [w]_B, [\alpha v]_B = \alpha[v]_B$, are precisely the fact that T is a linear map. The linear map T is injective since $[v]_B = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ implies that $v = 0 \cdot b_1 + \cdots + 0 \cdot b_n = 0_V$ and T is clearly surjective as $[\alpha_1 b_1 + \ldots \alpha_n b_n]_B = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$. □

Proposition 3.1.2. *If $T : V \to W$ is an isomorphism and $B = \{b_1, \ldots, b_n\}$ is a basis of V then $\{T(b_1), \ldots, T(b_n)\}$ is a basis of W. In particular,* $\dim(V) = \dim(W)$.

Proof. We prove first that $\{T(b_1), \ldots, T(b_n)\}$ is linearly independent. Indeed, if $\sum \alpha_i T(b_i) = 0$ then $T(\sum \alpha_i b_i) = 0$ and so $\sum \alpha_i b_i = 0$, since T is injective. Since B is a basis, each $\alpha_i = 0$ and so $\{T(b_1), \ldots, T(b_n)\}$ is a linearly independent set.

Now, if $\{T(b_1), \ldots, T(b_n)\}$ is not maximal linearly independent then for some $w \in W$ we have that $\{T(b_1), \ldots, T(b_n), w\}$ is linearly independent. Applying what we have already proven to the map T^{-1}, we find that $\{b_1, \ldots, b_n, T^{-1}(w)\}$ is a linearly independent set in V, which is a contradiction because B is a maximal independent set. □

Corollary 3.1.3. *Every finite dimensional vector space V over $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$ for a unique n; this n is* $\dim(V)$. *Two vector spaces are isomorphic if and only if they have the same dimension.*

3.2. **The theorem about the kernel and the image.**

Theorem 3.2.1. *Let $T : V \to W$ be a linear map where V is a finite dimensional vector space. Then* $\mathrm{Im}(T)$ *is finite dimensional and*

$$\dim(V) = \dim(\mathrm{Ker}(T)) + \dim(\mathrm{Im}(T)).$$

Proof. Let $\{v_1, \ldots, v_n\}$ be a basis for $\mathrm{Ker}(T)$ and extend it to a basis for V,

$$B = \{v_1, \ldots, v_n, w_1, \ldots, w_r\}.$$

So $\dim(V) = n + r$ and $\dim(\mathrm{Ker}(T)) = n$. Thus, the only thing we need to prove is that $\{T(w_1), \ldots, T(w_r)\}$ is a basis for $\mathrm{Im}(T)$. We shall show it is a minimal spanning set. First, let $v \in V$ and write v as

$$v = \sum_{i=1}^{n} \alpha_n v_n + \sum_{i=1}^{r} \beta_i w_i.$$

Then

$$\begin{aligned} T(v) &= T(\sum_{i=1}^{n} \alpha_n v_n + \sum_{i=1}^{r} \beta_i w_i) \\ &= \sum_{i=1}^{n} \alpha_n T(v_n) + \sum_{i=1}^{r} \beta_i T(w_i) \\ &= \sum_{i=1}^{r} \beta_i T(w_i), \end{aligned}$$

since $T(v_i) = 0$ for all i. Hence, $\{T(w_1), \ldots, T(w_r)\}$ is a spanning set.

To show it's minimal, suppose to the contrary that for some i we have that the set $\{T(w_1), \ldots, \widehat{T(w_i)}, \ldots, T(w_r)\}$ is a spanning set. W.l.o.g., $i = r$. Then, for some β_i,

$$T(w_r) = \sum_{i=1}^{r-1} \beta_i T(w_i),$$

whence,

$$T(\sum_{i=1}^{r-1} \beta_i w_i - w_r) = 0.$$

Thus, $\sum_{i=1}^{r-1} \beta_i w_i - w_r$ is in $\mathrm{Ker}(T)$ and so there are α_i such that

$$\sum_{i=1}^{r-1} \beta_i w_i - w_r - \sum_{i=1}^{n} \alpha_i v_i = 0.$$

This is a linear dependence between elements of the basis B and hence gives a contradiction. □

Remark 3.2.2. Suppose that $T : V \to W$ is surjective. Then we get that $\dim(V) = \dim(W) + \dim(\mathrm{Ker}(T))$. Note that for every $w \in W$ we have that the fibre $T^{-1}(w)$ is a coset of $\mathrm{Ker}(T)$, a set of the form $v + \mathrm{Ker}(T)$ where $T(v) = w$, and so it is natural to think about the dimension of $T^{-1}(w)$ as $\dim(\mathrm{Ker}(T))$.

Thus, we get that the dimension of the source is the dimension of the image plus the dimension of a general (in fact, any) fibre. This is an example of a general principle that holds true in many other circumstances in mathematics where there is a notion of dimension.

3.3. **Quotient spaces.** Let V be a vector space and U a subspace. Then V/U has a structure of abelian groups. We also claim that it has a structure of a vector space where we define

$$\alpha(v + U) = \alpha v + U,$$

or, in simpler notation,

$$\alpha \cdot \bar{v} = \overline{\alpha v}.$$

It is easy to check this is well defined and makes V/U into a vector space, called a **quotient space**. The natural map

$$\pi : V \to V/U$$

is a surjective linear map with kernel U. The following Corollary holds by applying Theorem 3.2.1 to the map $\pi : V \to V/U$.

Corollary 3.3.1. $\dim(V/U) = \dim(V) - \dim(U)$.

Theorem 3.3.2. *(First isomorphism theorem) Let $T : V \to W$ be a surjective linear map then*

$$V/\mathrm{Ker}(T) \cong W.$$

Proof. We already know that T induces an isomorphism $\bar{T}$ of abelian groups

$$\bar{T} : V/\mathrm{Ker}(T) \to W, \qquad \bar{T}(\bar{v}) := T(v).$$

We only need to check that $\bar{T}$ is a linear map, that is, that also $\bar{T}(\alpha\bar{v}) = \alpha\bar{T}(\bar{v})$. Indeed, $\bar{T}(\alpha\bar{v}) = \bar{T}(\overline{\alpha v}) = T(\alpha v) = \alpha T(v) = \alpha\bar{T}(\bar{v})$. □

3.4. **Applications of Theorem 3.2.1.**

Theorem 3.4.1. *Let W_1, W_2 be subspaces of a vector space V. Then,*

$$\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2).$$

Proof. Consider the function

$$T : W_1 \oplus W_2 \to W_1 + W_2,$$

given by $T(w_1, w_2) = w_1 + w_2$. Clearly T is a linear map and surjective. We thus have

$$\dim(W_1 + W_2) = \dim(W_1 \oplus W_2) - \dim(\mathrm{Ker}(T)).$$

However, $\dim(W_1 \oplus W_2) = \dim(W_1) + \dim(W_2)$ by Example 2.2.5. Our proof is thus complete if we show that

$$\mathrm{Ker}(T) \cong W_1 \cap W_2.$$

Let $u \in W_1 \cap W_2$ then $(u, -u) \in W_1 \oplus W_2$ and $T(u, -u) = 0$. We may thus define a map

$$L : W_1 \cap W_2 \to \mathrm{Ker}(T), \qquad L(u) = (u, -u),$$

which is clearly an injective linear map. Let $(w_1, w_2) \in \mathrm{Ker}(T)$ then $w_1 + w_2 = 0$ and so $w_1 = -w_2$. This shows that $w_1 \in W_2$ and so that $w_1 \in W_1 \cap W_2$. Thus, $(w_1, w_2) = L(w_1)$ and so L is surjective. □

Corollary 3.4.2. *If $\dim(W_1) + \dim(W_2) > \dim(V)$ then $W_1 \cap W_2$ contains a non-zero vector.*

The proof is left as an exercise. Here is a concrete example:

Example 3.4.3. Any two planes W_1, W_2 through the origin in $\mathbb{R}^3$ are either equal or intersect in a line.

Indeed, $W_1 \cap W_2$ is a non-zero vector space by the Corollary. If $\dim(W_1 \cap W_2) = 2$ then, since $W_1 \cap W_2 \subseteq W_i, i = 1, 2$ we have that $W_1 \cap W_2 = W_i$ and so $W_1 = W_2$. The only other option is that $\dim(W_1 \cap W_2) = 1$, that is, $W_1 \cap W_2$ is a line.

Another application of Theorem 3.2.1 is the following.

Corollary 3.4.4. *Let $T : V \to W$ be a linear map and assume $\dim(V) = \dim(W)$.*

(1) If T is injective it is an isomorphism.
(2) If T is surjective it is an isomorphism.

Proof. We prove that first part, leaving the second part as an exercise. We have that $\dim(\mathrm{Im}(T)) = \dim(V) - \dim(\mathrm{Ker}(T)) = \dim(V) = \dim(W)$, which implies by Corollary 2.2.4, that $\mathrm{Im}(T) = W$. Thus, T is surjective and the proof is complete. □

3.5. **Inner direct sum.** Let $U_1, \dots, U_n$ be subspaces of V such that:

(1) $V = U_1 + U_2 + \cdots + U_n$;

(2) For each i we have $U_i \cap (U_1 + \cdots + \widehat{U_i} + \cdots + U_n) = \{0\}$.

Then V is called an **inner direct sum** of the subspaces $U_1, \dots, U_n$.

Proposition 3.5.1. *V is the inner direct sum of the subspaces $U_1, \dots, U_n$ if and only if the map*

$$T : U_1 \oplus \cdots \oplus U_n \to V, \qquad (u_1, \dots, u_n) \mapsto u_1 + \cdots + u_n,$$

is an isomorphism.

Proof. The image of T is precisely the subspace $U_1 + U_2 + \cdots + U_n$. Thus, T is surjective iff condition (1) holds. We now show that T is injective iff condition (2) holds.

Suppose that T is injective. If $u \in U_i \cap (U_1 + \cdots + \widehat{U_i} + \cdots + U_n)$ for some i, say $u = u_1 + \cdots + \widehat{u_i} + \cdots + u_n$ then $0 = T(0, \dots, 0) = T(u_1, \dots, u_{i-1}, -u, u_{i+1}, \dots, u_n)$ and so $(u_1, \dots, u_{i-1}, -u, u_{i+1}, \dots, u_n) = 0$ and in particular $u = 0$. So condition (2) holds.

Suppose now that condition (2) holds and $T(u_1, \dots, u_n) = 0$. Then $-u_i = u_1 + \cdots + \widehat{u_i} + \cdots + u_n$ and so $u_i \in U_i \cap (U_1 + \cdots + \widehat{U_i} + \cdots + U_n) = \{0\}$. Thus, $u_i = 0$ and that holds for every i. We conclude that $\mathrm{Ker}(T) = \{(0, \dots, 0)\}$ and so T is injective. □

When V is the inner direct sum of the subspaces $U_1, \dots, U_n$ we shall use the notation

$$V = U_1 \oplus \cdots \oplus U_n.$$

This abuse of notation is justified by the Proposition.

Proposition 3.5.2. *The following are equivalent:*

(1) V is the inner direct sum of the subspaces $U_1, \dots, U_n$;

(2) $V = U_1 + \cdots + U_n$ and $\dim(V) = \dim(U_1) + \cdots + \dim(U_n)$;

(3) Every vector $v \in V$ can be written as $v = u_1 + \cdots + u_n$, with $u_i \in U_i$, in a unique way.

The proof of the Proposition is left as an exercise.

3.6. **Nilpotent operators.** A linear map $T : V \to V$ from a vector space to itself is often called a **linear operator**.

Definition 3.6.1. Let V be a finite dimensional vector space and $T : V \to V$ a linear operator. T is called **nilpotent** if for some $N \geq 1$ we have $T^N \equiv 0$. (Here $T^N = T \circ T \circ \cdots \circ T$, N-times.)

The following Lemma is left as an exercise:

Lemma 3.6.2. *Let T be a nilpotent operator on an n-dimensional vector space then $T^n \equiv 0$.*

Example 3.6.3. Here are some examples of nilpotent operators. Of course, the trivial example is $T \equiv 0$, the zero map. For another example, let V be a vector space of dimension n and let $B = \{b_1, \dots, b_n\}$ be a basis. Let T be the unique linear transformation (cf. Proposition 3.0.9) satisfying

$$\begin{aligned} T(b_1) &= 0 \\ T(b_2) &= b_1 \\ T(b_3) &= b_2 \\ &\vdots \\ T(b_n) &= b_{n-1}. \end{aligned}$$

We see that $T^n \equiv 0$.

Example 3.6.4. Let $T : \mathbb{F}[t]_n \to \mathbb{F}[t]_n$ be defined by $T(f) = f'$. Then T is nilpotent and $T^n = 0$.

The following theorem, called Fitting's Lemma, is important in mathematics because the statement and the method of proof generalize to many other situations. We remark that later on we shall prove much stronger "structure theorems" (for example, the Jordan canonical form, cf. § 9.2) from which Fitting's Lemma follows immediately, but this is very special to vector spaces.

Theorem 3.6.5 (Fitting's Lemma)**.** *Let V be a finite dimensional vector space and let $T : V \to V$ be a linear operator. Then there is a decomposition*

$$V = U \oplus W$$

such that

(1) U, W are T-invariant subspaces of V, that is $T(U) \subseteq U, T(W) \subseteq W$;
(2) $T|_U$ is nilpotent;
(3) $T|_W$ is an isomorphism.

Remark 3.6.6. About notation. $T|_U$, read "T restricted to U", is the linear map

$$U \to U, \qquad u \mapsto T(u).$$

Namely, it is just the map T considered on the subspace U.

Proof. Let us define

$$U_i = \mathrm{Ker}(T^i), \quad W_i = \mathrm{Im}(T^i).$$

We note the following facts:

(1) U_i, W_i are subspaces of V;
(2) $\dim(U_i) + \dim(W_i) = \dim(V)$;
(3) $\{0\} \subseteq U_1 \subseteq U_2 \subseteq \dots$;
(4) $V \supseteq W_1 \supseteq W_2 \supseteq \dots$.

It follows from Fact (4) that $\dim(V) \geq \dim(W_1) \geq \dim(W_2) \geq \dots$ and so, for some N we have

$$\dim(W_N) = \dim(W_{N+1}) = \dim(W_{N+2}) = \dots$$

It then follows from Fact (2) that also

$$\dim(U_N) = \dim(U_{N+1}) = \dim(U_{N+2}) = \dots$$

Hence, using Corollary 2.2.4, we obtain

$$W_N = W_{N+1} = W_{N+2} = \dots, \quad U_N = U_{N+1} = U_{N+2} = \dots$$

We let

$$W = W_N, \qquad U = U_N.$$

We note that $T(W_N) = W_{N+1}$ and so $T|_W : W \to W$ is an isomorphism since the dimension of the image is the dimension of the source (Corollary 3.4.4). Also $T(\mathrm{Ker}(T^N)) \subseteq \mathrm{Ker}(T^{N-1}) \subseteq \mathrm{Ker}(T^N)$ and so $T|_U : U \to U$ and $(T|_U)^N = (T^N)|_U = 0$. That is, T is nilpotent on U.

It remains to show that $V = U \oplus W$. First, $\dim(U) + \dim(W) = \dim(V)$. Second, if $v \in U \cap W$ is a non-zero vector, then $T(v) \neq 0$ because $T|_W$ is an isomorphism and so $T^N(v) \neq 0$, but on the other hand $T^N(v) = 0$ because $v \in U$. Thus, $U \cap W = \{0\}$. It follows that the map

$$U \oplus W \to V, \qquad (u, w) \mapsto u + w,$$

which has kernel $U \cap W$ (cf. the proof of Theorem 3.4.1) is injective. The information on the dimension gives that it is an isomorphism (Corollary 3.4.4) and so $V = U \oplus W$ by Proposition 3.5.1. □

3.7. **Projections.**

Definition 3.7.1. Let V be a vector space. A linear operator, $T : V \to V$, is called a **projection** if $T^2 = T$.

Theorem 3.7.2. *Let V be a vector space over $\mathbb{F}$.*

(1) Let U, W be subspaces of V such that $V = U \oplus W$. Define a map

$$T : V \to V, \qquad T(v) = u \qquad \text{if } v = u + w, \ \ u \in U, w \in W.$$

Then T is a projection, $\mathrm{Im}(T) = U, \mathrm{Ker}(T) = W$.

(2) Let $T : V \to V$ be a projection and $U = \mathrm{Im}(T), W = \mathrm{Ker}(T)$. Then $V = U \oplus W$ and $T(u + w) = u$, that is, T is the operator constructed in (1).

Definition 3.7.3. The operator constructed in (1) of the Theorem is called the **projection on** U **along** W.

Proof. Consider the first claim. If $u \in U$ then u is written as $u + 0$ and so $T(u) = u$ and so $T^2(v) = T(u) = u = T(v)$ and so $T^2 = T$. Now, $v \in \mathrm{Ker}(T)$ if and only if $v = 0 + w$ for some $w \in W$ and so $\mathrm{Ker}(T) = W$. Also, since for $u \in U$, $T(u) = u$, we also get that $\mathrm{Im}(T) = U$.

We now consider the second claim. Note that

$$v = T(v) + (v - T(v)).$$

$T(v) \in \mathrm{Im}(T)$ and $T(v - T(v)) = T(v) - T^2(v) = T(v) - T(v) = 0$ and so $v - T(v) \in \mathrm{Ker}(T)$. It follows that $U + W = V$. Theorem 3.2.1 gives that $\dim(V) = \dim(U) + \dim(W)$ and so Proposition 3.5.2 gives that

$$V = U \oplus W.$$

Now, writing $v = u + w = T(v) + (v - T(v))$ and comparing these expressions, we see that $u = T(v)$. □

3.8. **Linear maps and matrices.** Let $\mathbb{F}$ be a field and V, W vector spaces over $\mathbb{F}$ of dimension n and m respectively.

Theorem 3.8.1. *Let $T : V \to W$ be a linear map.*

(1) Let B be a basis for V and C a basis for W. There is a unique $m \times n$ matrix, denoted ${}_C[T]_B$ and called ***the matrix representing*** *T, with entries in $\mathbb{F}$ such that*

$$[Tv]_C = {}_C[T]_B[v]_B, \quad \forall v \in V.$$

(2) If $S : V \to W$ is another linear transformation then

$${}_C[S + T]_B = {}_C[S]_B + {}_C[T]_B, \quad {}_C[\alpha T]_B = \alpha \cdot {}_C[T]_B.$$

(3) For every matrix $M \in M_{m\times n}(\mathbb{F})$ there is a linear map $T : V \to W$ such that ${}_C[T]_B = M$. We conclude that the map

$$T \mapsto {}_C[T]_B$$

is an isomorphism of vector spaces

$$\mathrm{Hom}(V, W) \cong M_{m\times n}(\mathbb{F}).$$

(4) If $R : W \to U$ is another linear map, where U is a vector space over $\mathbb{F}$, and D is a basis for U then

$${}_D[R \circ T]_B = {}_D[R]_C {}_C[T]_B.$$

Proof. We begin by proving the first claim. Let $B = \{s_1, \dots, s_n\}, C = \{t_1, \dots, t_m\}$. Write

$$[T(s_1)]_C = \begin{pmatrix} d_{11} \\ \vdots \\ d_{m1} \end{pmatrix}, \dots, [T(s_n)]_C = \begin{pmatrix} d_{1n} \\ \vdots \\ d_{mn} \end{pmatrix}.$$

Let $M = (d_{ij})_{1 \le i \le m, 1 \le j \le n}$. We prove that

$$M[v]_B = [T(v)]_C.$$

Indeed, write $v = \alpha_1 s_1 + \cdots + \alpha_n s_n$ and calculate

$$\begin{aligned} M[v]_B &= M[\alpha_1 s_1 + \cdots + \alpha_n s_n]_B \\ &= M \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \\ &= M \left(\alpha_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \alpha_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \right) \\ &= \alpha_1 M \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \alpha_2 M \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \alpha_n M \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \\ &= \alpha_1 [T(s_1)]_C + \cdots + \alpha_n [T(s_n)]_C \\ &= [\alpha_1 T(s_1) + \cdots + \alpha_n T(s_n)]_C \\ &= [T(\alpha_1 s_1 + \cdots + \alpha_n s_n)]_C \\ &= [T(v)]_C. \end{aligned}$$

Now suppose that $N = (\delta_{ij})_{1 \le i \le m, 1 \le j \le n}$ is another matrix such that

$$[T(v)]_C = N[v]_B, \quad \forall v \in V.$$

Then,

$$[T(v_i)]_C = \begin{pmatrix} d_{1i} \\ \vdots \\ d_{mi} \end{pmatrix}, \qquad [T(v_i)]_C = Ne_i = \begin{pmatrix} \delta_{1i} \\ \vdots \\ \delta_{mi} \end{pmatrix}.$$

That shows that $N = M$.

We now show the second claim is true. We have for every $v \in V$ the following equalities:

$$\begin{aligned} ({}_C[S]_B + {}_C[T]_B)\,[v]_B &= {}_C[S]_B[v]_B + {}_C[T]_B[v]_B \\ &= [S(v)]_C + [T(v)]_C \\ &= [S(v) + T(v)]_C \\ &= [(S+T)(v)]_C. \end{aligned}$$

Namely, if we call M the matrix ${}_C[S]_B + {}_C[T]_B$ then $M[v]_B = [(S+T)(v)]_C$, which proves that $M = {}_C[S+T]_B$.

Similarly, $\alpha \cdot_C [T]_B[v]_B = \alpha[T(v)]_C = [\alpha \cdot T(v)]_C = [(\alpha T)(v)]_C$ and that shows that $\alpha \cdot_C [T]_B = {}_C[\alpha T]_B$.

The third claim follows easily from the previous results. We already know that the maps

$$H_1 : V \to \mathbb{F}^n, \quad v \mapsto [v]_B, \qquad H_3 : W \mapsto \mathbb{F}^m, w \mapsto [w]_C,$$

and

$$H_2 : \mathbb{F}^n \to \mathbb{F}^m, x \mapsto Mx$$

are linear maps. It follows that the composition $T = H_3^{-1} \circ H_2 \circ H_1$ is a linear map. Furthermore,

$$[T(v)]_C = H_3(T(v)) = M(H_1(v)) = M[v]_B,$$

and so

$$M = {}_C[T]_B.$$

This shows that the map

$$\mathrm{Hom}(V, W) \to M_{m\times n}(\mathbb{F})$$

is surjective. The fact that it's a linear map is the second claim. The map is also injective, because if ${}_C[T]_B$ is the zero matrix then for every $v \in V$ we have $[T(v)]_C = {}_C[T]_B[v]_B = 0$ and so $T(v) = 0$ which shows that T is the zero transformation.

It remains to prove that last claim. For every $v \in V$ we have

$$\begin{aligned}({}_D[R]_{CC}[T]_B)[v]_B &= {}_D[R]_C({}_C[T]_B[v]_B)\\ &= {}_D[R]_C[T(v)]_C\\ &= [R(T(v))]_D\\ &= [(R \circ T)(v)]_D.\end{aligned}$$

It follows then that ${}_D[R]_{CC}[T]_B = {}_D[R \circ T]_B$. □

Corollary 3.8.2. *We have* $\dim(\mathrm{Hom}(V, W)) = \dim(V) \cdot \dim(W)$.

Example 3.8.3. Consider the identity $\mathrm{Id} : V \to V$, but with two different basis B and C. Then

$${}_C[\mathrm{Id}]_B[v]_B = [v]_C,$$

Namely, ${}_C[\mathrm{Id}]_B$ is just the change of basis matrix,

$${}_C[\mathrm{Id}]_B = {}_CM_B.$$

Example 3.8.4. Let $V = \mathbb{F}[t]_{n+1}$ and take the basis $B = \{1, t, \ldots, t^n\}$. Let $T : V \to V$ be the formal differentiation map $T(f) = f'$. Then

$${}_B[T]_B = \begin{pmatrix} 0 & 1 & 0 & & \ldots & 0\\ 0 & 0 & 2 & & \ldots & \\ 0 & 0 & 0 & 3 & \cdots & \\ \vdots & \vdots & & & & \\ & & & & & n\\ 0 & 0 & 0 & & \cdots & 0 \end{pmatrix}.$$

Example 3.8.5. Let $V = \mathbb{F}[t]_{n+1}$, $W = \mathbb{F}^2$, $B = \{1, t, \ldots, t^n\}$, St the standard basis of W, and

$$T : V \to W, \qquad T(f) = (f(1), f(2)).$$

Then

$${}_{\mathrm{St}}[T]_B = \begin{pmatrix} 1 & 1 & 1 & \ldots & 1\\ 1 & 2 & 4 & \ldots & 2^n \end{pmatrix}.$$

3.9. **Change of basis.** It is often useful to pass from a representation of a linear map in one basis to a representation in another basis. In fact, the applications of this are hard to overestimate! We shall later see many examples. For now, we just give the formulas.

Proposition 3.9.1. *Let $T : V \to V$ be a linear transformation and B and C two bases of V. Then*

$$ {}_B[T]_B = {}_BM_{CC}[T]_{CC}M_B. $$

Proof. Indeed, for every $v \in V$ we have

$$ \begin{aligned} {}_BM_{CC}[T]_{CC}M_B[v]_B &= {}_BM_{CC}[T]_C[v]_C \\ &= {}_BM_C[Tv]_C \\ &= [Tv]_B. \end{aligned} $$

Thus, by uniqueness, we have ${}_B[T]_B = {}_BM_{CC}[T]_{CC}M_B$. □

Remark 3.9.2. More generally, the same idea of proof, gives the following. Let $T : V \to W$ be a linear map and $B, \widetilde{B}$ bases for V, $C\widetilde{C}$ bases for W. Then

$$ {}_{\widetilde{C}}[T]_{\widetilde{B}} = {}_{\widetilde{C}}M_{CC}[T]_{BB}M_{\widetilde{B}}. $$

Example 3.9.3. We want to find the matrix representing in the standard basis the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ which is the projection on the plane $\{(x_1, x_2, x_3) : x_1 + x_3 = 0\}$ along the line $\{t(1, 0, 1) : t \in \mathbb{R}\}$.

We first check that $\{(1, 0, -1), (1, 1, -1)\}$ is a minimal spanning set for the plane. We complete it to a basis by adding the vector $(1, 0, 1)$ (since that vector is not in the plane it is independent of the two preceding vectors and so we get an independent set of 3 elements, hence a basis). Thus,

$$ B = \{(1, 0, -1), (1, 1, -1), (1, 0, 1)\} $$

is a basis of $\mathbb{R}^3$. It is clear that

$$ {}_B[T]_B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. $$

Now, ${}_{\mathrm{St}}M_B = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix}$. One can calculate that

$$ {}_BM_{\mathrm{St}} = \begin{pmatrix} 1/2 & -1 & -1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix}. $$

Thus, we conclude that

$$ \begin{aligned} {}_{\mathrm{St}}[T]_{\mathrm{St}} &= {}_{\mathrm{St}}M_{BB}[T]_{BB}M_{\mathrm{St}} \\ &= \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1/2 & -1 & -1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix} \\ &= \begin{pmatrix} 1/2 & 0 & -1/2 \\ 0 & 1 & 0 \\ -1/2 & 0 & 1/2 \end{pmatrix}. \end{aligned} $$

4. The determinant and its applications

4.1. **Quick recall: permutations.** We refer to the notes of the previous course MATH 251 for basic properties of the symmetric group S_n, the group of permutations of n elements. In particular, recall the following:

- Every permutation can be written as a product of cycles.
- Disjoint cycles commute.
- In fact, every permutation can be written as a product of disjoint cycles, unique up to their ordering.
- Every permutation is a product of transpositions.

4.2. **The sign of a permutation.**

Lemma 4.2.1. *Let $n \geq 2$. Let S_n be the group of permutations of $\{1, 2, \ldots, n\}$. There exists a surjective homomorphism of groups*

$$\operatorname{sgn} : S_n \to \{\pm 1\}$$

*(called the **sign**). It has the property that for every $i \neq j$,*

$$\operatorname{sgn}(\ (ij)\ ) = -1.$$

Proof. Consider the polynomial in n-variables[3]

$$p(x_1, \ldots, x_n) = \prod_{i<j}(x_i - x_j).$$

Given a permutation σ we may define a new polynomial

$$\prod_{i<j}(x_{\sigma(i)} - x_{\sigma(j)}).$$

Note that $\sigma(i) \neq \sigma(j)$ and for any pair $k < \ell$ we obtain in the new product either $(x_k - x_\ell)$ or $(x_\ell - x_k)$. Thus, for a suitable choice of sign $\operatorname{sgn}(\sigma) \in \{\pm 1\}$, we have[4]

$$\prod_{i<j}(x_{\sigma(i)} - x_{\sigma(j)}) = \operatorname{sgn}(\sigma)\prod_{i<j}(x_i - x_j).$$

We obtain a function

$$\operatorname{sgn} : S_n \to \{\pm 1\}.$$

This function satisfies $\operatorname{sgn}(\ (k\ell)\ ) = -1$ (for $k < \ell$): Let $\sigma = (k\ell)$ and consider the product

$$\prod_{i<j}(x_{\sigma(i)} - x_{\sigma(j)}) = (x_\ell - x_k) \prod_{\substack{i<j \\ i\neq k, j\neq \ell}} (x_{\sigma(i)} - x_{\sigma(j)}) \prod_{\substack{k<j \\ j\neq \ell}}(x_\ell - x_j) \prod_{\substack{i<\ell \\ i\neq k}}(x_i - x_k)$$

Counting the number of signs that change we find that

$$\prod_{i<j}(x_{\sigma(i)} - x_{\sigma(j)}) = (-1)(-1)^{\sharp\{j:k<j<\ell\}}(-1)^{\sharp\{i:k<i<\ell\}}\prod_{i<j}(x_i - x_j) = -\prod_{i<j}(x_i - x_j).$$

It remains to show that sgn is a group homomorphism. We first make the innocuous observation that for <u>any</u> variables $y_1, \ldots, y_n$ and for <u>any</u> permuation σ we have

$$\prod_{i<j}(y_{\sigma(i)} - y_{\sigma(j)}) = \operatorname{sgn}(\sigma)\prod_{i<j}(y_i - y_j).$$

---

[3] For $n = 2$ we get $x_1 - x_2$. For $n = 3$ we get $(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)$.

[4] For example, if $n = 3$ and σ is the cycle (123) we have

$$(x_{\sigma(1)} - x_{\sigma(2)})(x_{\sigma(1)} - x_{\sigma(3)})(x_{\sigma(2)} - x_{\sigma(3)}) = (x_2 - x_3)(x_2 - x_1)(x_3 - x_1) = (x_1 - x_2)(x_1 - x_3)(x_2 - x_3).$$

Hence, $\operatorname{sgn}(\ (1\ 2\ 3)\ ) = 1$.

Let τ be a permutation. We apply this observation for the variables $y_i := x_{\tau(i)}$. We get

$$\begin{aligned}\operatorname{sgn}(\tau\sigma)p(x_1,\dots,x_n) &= p(x_{\tau\sigma(1)},\dots,x_{\tau\sigma(n)})\\ &= p(y_{\sigma(1)},\dots,y_{\sigma(n)})\\ &= \operatorname{sgn}(\sigma)p(y_1,\dots,y_n)\\ &= \operatorname{sgn}(\sigma)p(x_{\tau(1)},\dots,x_{\tau(n)})\\ &= \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)p(x_1,\dots,x_n).\end{aligned}$$

This gives

$$\operatorname{sgn}(\tau\sigma) = \operatorname{sgn}(\tau)\operatorname{sgn}(\sigma).$$

□

4.2.1. *Calculating* sgn *in practice.* Recall that every permutation σ can be written as a product of disjoint cycles

$$\sigma = (a_1 \dots a_\ell)(b_1 \dots b_m)\dots(f_1 \dots f_n).$$

Lemma 4.2.2. $\operatorname{sgn}(a_1 \dots a_\ell) = (-1)^{\ell-1}$.

Corollary 4.2.3. $\operatorname{sgn}(\sigma) = (-1)^{\sharp \text{ even length cycles}}$.

Proof. We write

$$(a_1 \dots a_\ell) = \underbrace{(a_1 a_\ell)\dots(a_1 a_3)(a_1 a_2)}_{\ell-1 \text{ transpositions}}.$$

Since a transposition has sign -1 and sgn is a homomorphism, the claim follows. □

Example 4.2.4. Let $n = 11$ and

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 2 & 5 & 4 & 3 & 1 & 7 & 8 & 10 & 6 & 9 \end{pmatrix}.$$

Then

$$\sigma = (1\ 2\ 5)(3\ 4)(6\ 7\ 8\ 10\ 9).$$

Now,

$$\operatorname{sgn}(\ (1\ 2\ 5)\ ) = 1, \quad \operatorname{sgn}(\ (3\ 4)\ ) = -1, \quad \operatorname{sgn}(\ (6\ 7\ 8\ 10\ 9)\ ) = 1.$$

We conclude that $\operatorname{sgn}(\sigma) = -1$.

4.3. **Determinants.** Let $\mathbb{F}$ let a field. We will consider $n \times n$ matrices with entries in $\mathbb{F}$, namely elements of $M_n(\mathbb{F})$. We shall use the notation $(v_1 v_2 \dots v_n)$ to denote such a matrix whose columns are the vectors $v_1, v_2, \dots, v_n$ of $\mathbb{F}^n$.

Theorem 4.3.1. *Let $\mathbb{F}$ be a field. There is a function, called the* ***determinant****,*

$$\det : M_n(\mathbb{F}) \to \mathbb{F}$$

having the following properties:

(1) $\det(v_1 v_2 \dots \alpha v_i \dots v_n) = \alpha \cdot \det(v_1 v_2 \dots v_n)$.
(2) $\det(v_1 v_2 \dots v_i + v_i' \dots v_n) = \det(v_1 v_2 \dots v_i \dots v_n) + \det(v_1 v_2 \dots v_i' \dots v_n)$.
(3) $\det(v_1 v_2 \dots v_i \dots v_j \dots v_n) = 0$ *if* $v_i = v_j$ *for* $i < j$.
(4) $\det(e_1 e_2 \dots e_n) = 1$.

Corollary 4.3.2. *For every permutation $\tau \in S_n$ we have*

$$\det(v_1 v_2 \dots v_n) = \operatorname{sgn}(\tau) \cdot \det(v_{\tau(1)} v_{\tau(2)} \dots v_{\tau(n)}).$$

Proof. Suppose the formula holds for $\sigma, \tau \in S_n$ and any choice of vectors. Then

$$\det(w_1 w_2 \dots w_n) = \operatorname{sgn}(\tau) \cdot \det(w_{\tau(1)} w_{\tau(2)} \dots w_{\tau(n)}).$$

Let $w_i = v_{\sigma(i)}$, then

$$\det(v_{\sigma(1)} v_{\sigma(2)} \dots v_{\sigma(n)}) = \operatorname{sgn}(\tau) \cdot \det(v_{\sigma\tau(1)} v_{\sigma\tau(2)} \dots v_{\sigma\tau(n)}).$$

Therefore,

$$\operatorname{sgn}(\sigma) \cdot \det(v_1 v_2 \dots v_n) = \operatorname{sgn}(\tau) \cdot \det(v_{\sigma\tau(1)} v_{\sigma\tau(2)} \dots v_{\sigma\tau(n)}).$$

Rearranging, we get

$$\det(v_1 v_2 \dots v_n) = \operatorname{sgn}(\sigma\tau) \cdot \det(v_{\sigma\tau(1)} v_{\sigma\tau(2)} \dots v_{\sigma\tau(n)}).$$

Since transpositions generate S_n, it is therefore enough to prove the corollary for transpositions. We need to prove then that for $i < j$ we have

$$\det(v_1 v_2 \dots v_i \dots v_j \dots v_n) + \det(v_1 v_2 \dots v_j \dots v_i \dots v_n) = 0.$$

However,

$$\begin{aligned} 0 &= \det(v_1 v_2 \dots (v_i + v_j) \dots (v_i + v_j) \dots v_n) \\ &= \det(v_1 v_2 \dots v_i \dots v_i \dots v_n) + \det(v_1 v_2 \dots v_j \dots v_j \dots v_n) \\ &\qquad + \det(v_1 v_2 \dots v_i \dots v_j \dots v_n) + \det(v_1 v_2 \dots v_j \dots v_i \dots v_n) \\ &= \det(v_1 v_2 \dots v_i \dots v_j \dots v_n) + \det(v_1 v_2 \dots v_j \dots v_i \dots v_n). \end{aligned}$$

□

Proof. (Of Theorem) We define the function

$$\det : M_n(\mathbb{F}) \to \mathbb{F}, \qquad A = (a_{ij}) \mapsto \det A,$$

by the formula

$$\boxed{\det(a_{ij}) = \textstyle\sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)\,1} \cdots a_{\sigma(n)\,n}}$$

We verify that it satisfies properties (1) - (4).

(1) Let us write $(v_1 v_2 \dots v_i \dots v_n) = (a_{\ell j})$ and let

$$b_{\ell j} = \begin{cases} a_{\ell j} & j \neq i \\ \alpha a_{\ell i} & j = i. \end{cases}$$

Namely,

$$(v_1 v_2 \dots \alpha v_i \dots v_n) = (b_{\ell j}).$$

By definition,

$$\det(v_1 v_2 \dots \alpha v_i \dots v_n) = \det((b_{\ell j})) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) b_{\sigma(1)\,1} \cdots b_{\sigma(n)\,n}.$$

Since in every summand there is precisely one element of the form $b_{\ell i}$, namely, the element $b_{\sigma(i)\,i} = \alpha a_{\sigma(i)\,i}$, we have

$$\begin{aligned}\det(v_1 v_2 \dots \alpha v_i \dots v_n) &= \alpha \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)\,1}\cdots a_{\sigma(n)\,n}\\ &= \alpha \det((a_{ij}))\\ &= \alpha \det(v_1 v_2 \dots v_i \dots v_n).\end{aligned}$$

(2) We let

$$(v_1 v_2 \dots v_i \dots v_n) = (a_{\ell j}), \qquad (v_1 v_2 \dots v_i' \dots v_n) = (a'_{\ell j}).$$

We note that if $j \neq i$ then $a_{\ell j} = a'_{\ell j}$. Define now

$$b_{\ell j} = \begin{cases} a_{\ell j} & j \neq i \\ a_{\ell j} + a'_{\ell j} & j = i. \end{cases}$$

Then we have

$$(v_1 v_2 \dots v_i + v_i' \dots v_n) = (b_{\ell j}).$$

Now,

$$\begin{aligned}\det(v_1 v_2 \dots v_i + v_i' \dots v_n) &= \det(b_{\ell j})\\ &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) b_{\sigma(1)1}\cdots b_{\sigma(n)n}\\ &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1}\cdots (a_{\sigma(i)i} + a'_{\sigma(i)i})\cdots a_{\sigma(n)n}\\ &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1}\cdots a_{\sigma(i)i}\cdots a_{\sigma(n)n} + \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1}\cdots a'_{\sigma(i)i}\cdots a_{\sigma(n)n}\\ &= \det(v_1 v_2 \dots v_i \dots v_n) + \det(v_1 v_2 \dots v_i' \dots v_n).\end{aligned}$$

(3) Let $S \subset S_n$ be a set of representatives for the subgroup $\{1, (ij)\}$. Then $S_n = S \coprod S(ij)$. For $\sigma \in S$ let $\sigma' = \sigma(ij)$. We have

$$\begin{aligned}\det(v_1 v_2 \dots v_i \dots v_j \dots v_n) &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)\,1}\cdots a_{\sigma(n)\,n}\\ &= \sum_{\sigma\in S} \operatorname{sgn}(\sigma) a_{\sigma(1)\,1}\cdots a_{\sigma(n)\,n} + \sum_{\sigma\in S} \operatorname{sgn}(\sigma') a_{\sigma'(1)\,1}\cdots a_{\sigma'(n)\,n}\\ &= \sum_{\sigma\in S} \operatorname{sgn}(\sigma) a_{\sigma(1)\,1}\cdots a_{\sigma(n)\,n} - \sum_{\sigma\in S} \operatorname{sgn}(\sigma) a_{\sigma'(1)\,1}\cdots a_{\sigma'(n)\,n}.\end{aligned}$$

It is enough to show that for every σ we have

$$a_{\sigma(1)\,1}\cdots a_{\sigma(n)\,n} = a_{\sigma'(1)\,1}\cdots a_{\sigma'(n)\,n}.$$

If $\ell \notin \{i, j\}$ then $a_{\sigma'(\ell)\,\ell} = a_{\sigma\circ(ij)(\ell)\,\ell} = a_{\sigma(\ell)\,\ell}$. If $\ell = i$ then $a_{\sigma'(i)\,i} = a_{\sigma(j)\,i} = a_{\sigma(j)\,j}$ and if $\ell = j$ then $a_{\sigma'(j)\,j} = a_{\sigma(i)\,j} = a_{\sigma(i)\,i}$. We get $a_{\sigma(1)\,1}\cdots a_{\sigma(n)\,n} = a_{\sigma'(1)\,1}\cdots a_{\sigma'(n)\,n}$.

(4) Recall the definition of Kronecker's delta δ_{ij}. By definition $\delta_{ij} = 1$ if $i = j$ and zero otherwise.

For the identity matrix $I_n = (\delta_{ij})$ we have

$$\det I_n = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \delta_{\sigma(1)\,1}\dots\delta_{\sigma(n)\,n}.$$

If there's a single i such that $\sigma(i) \neq i$ then $\delta_{\sigma(i)\,i} = 0$ and so $\delta_{\sigma(1)\,1}\dots\delta_{\sigma(n)\,n} = 0$. Thus,

$$\det I_n = \operatorname{sgn}(1)\delta_{11}\cdots\delta_{nn} = 1.$$

□

Remark 4.3.3. In the assignments you are proving that

$$\det(A) = \det(A^t).$$

From the it follows immediately that the determinant has properties analogous to (1) - (4) of Theorem 4.3.1 relative to rows.

4.4. **Examples and geometric interpretation of the determinant.**

4.4.1. *Examples in low dimension.* Here we calculate the determinants of matrices of size $1, 2, 3$ from the definition.

(1) $n = 1$.
We have

$$\det(a) = a.$$

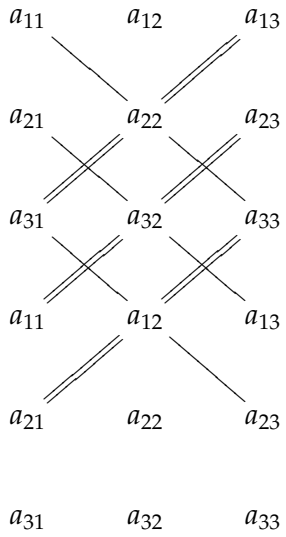
(2) $n = 2$.
$S_2 = \{1, (12)\} = \{1, \sigma\}$ and so

$$\begin{aligned}\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} &= \operatorname{sgn}(1) a_{11} a_{22} + \operatorname{sgn}(\sigma) a_{\sigma(1)\,1} a_{\sigma(2)\,2} \\ &= a_{11}a_{22} - a_{21}a_{12}.\end{aligned}$$

(3) $n = 3$.
$S_3 = \{1, (12), (13), (23), (123), (132)\}$. The sign of a transposition is -1 and of a 3-cycle is 1. We find the formula,

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{11}a_{23}a_{32} - a_{31}a_{22}a_{13} - a_{21}a_{12}a_{33}.$$

One way to remember this is to write the matrix twice and draw the diagonals.



You multiply elements on the diagonal with the signs depending on the direction of the diagonal.

4.4.2. *Geometric interpretation.* We notice in the one dimensional case that $|\det(a)| = |a|$ is the length of the segment from 0 to a. Thus, the determinant is a signed length function.

In the two dimensional case, it is easy to see that

$$\det \begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix}$$

is the area, up to a sign, of the parallelogram with sides $(0,0) - (a_1, 0)$ and $(0,0) - (a_{12}, a_{21})$. The sign depends on the orientation of the vectors. Similarly, for the matrix

$$\begin{pmatrix} 0 & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Using the decomposition,

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \det \begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix} + \det \begin{pmatrix} 0 & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

one sees that the signed area (the sign depending on the orientation of the vectors) of the parallelogram with sides $(0,0) - (a_{11}, a_{21})$ and $(0,0) - (a_{12}, a_{22})$ is the determinant of the matrix $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$.

More generally, for every n we may interpret the determinant as a signed **volume function**. It associates to an n-tuple of vectors $v_1, \ldots, v_n$ the volume of the parallelepiped whose edges are $v_1, v_2, \ldots, v_n$ by the formula $\det(v_1 v_2 \ldots v_n)$. It is scaled by the requirement (4) that $\det I_n = 1$, that is the volume of the unit cube is 1. The other properties of the determinant can be viewed as change of volume under stretching one of the vectors (property (1)) and writing the volume of a parallelepiped when we decompose it in two parallelepiped (property (2)). Property (3) can be viewed as saying that if two vectors lie on the same line then the volume is zero; this makes perfect sense as in this case the parallelepiped actually lives in a lower dimensional vector space spanned by the $n-1$ vectors $v_1, \ldots, v_i, \ldots \hat{v}_j, \ldots, v_n$. We shall soon see that, in the same spirit, if the vectors $v_1, \ldots, v_n$ are linearly dependent (so again the parallelepiped lives in a lower dimension space) then the determinant is zero.

4.4.3. *Realizing S_n as linear transformations.* Let $\mathbb{F}$ be any field. Let $\sigma \in S_n$. There is a unique linear transformation

$$T_\sigma : \mathbb{F}^n \to \mathbb{F}^n,$$

such that

$$T(e_i) = e_{\sigma(i)}, \quad i = 1, \ldots n,$$

where, as usual, $e_1, \ldots, e_n$ are the standard basis of $\mathbb{F}^n$. Note that

$$T_\sigma \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_{\sigma^{-1}(1)} \\ x_{\sigma^{-1}(2)} \\ \vdots \\ x_{\sigma^{-1}(n)} \end{pmatrix}.$$

(For example, because $T_\sigma x_1 e_1 = x_1 e_{\sigma(1)}$, the $\sigma(1)$ coordinate is x_1, namely, in the $\sigma(1)$ place we have the entry $x_{\sigma^{-1}(\sigma(1))}$.) Since for every i we have $T_\sigma T_\tau(e_i) = T_\sigma e_{\tau(i)} = e_{\sigma\tau(i)} = T_{\sigma\tau} e_i$, we have the relation

$$T_\sigma T_\tau = T_{\sigma\tau}.$$

The matrix representing T_σ is the matrix (a_{ij}) with $a_{ij} = 0$ unless $i = \sigma(j)$. For example, for $n = 4$ the matrices representing the permutations $(12)(34)$ and $(1\ 2\ 3\ 4)$ are, respectively

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Otherwise said,[5]

$$T_\sigma = \left(e_{\sigma(1)} \mid e_{\sigma(2)} \mid \ldots \mid e_{\sigma(n)}\right) = \begin{pmatrix} e_{\sigma^{-1}(1)} \\ \hline e_{\sigma^{-1}(2)} \\ \hline \vdots \\ \hline e_{\sigma^{-1}(n)} \end{pmatrix}.$$

It follows that

$$\begin{aligned} \operatorname{sgn}(\sigma)\det(T_\sigma) &= \operatorname{sgn}(\sigma)\det\left(e_{\sigma(1)} \mid e_{\sigma(2)} \mid \ldots \mid e_{\sigma(n)}\right) \\ &= \det\left(e_1 \mid e_2 \mid \ldots \mid e_n\right) \\ &= \det(I_n) \\ &= 1. \end{aligned}$$

Recall that $\operatorname{sgn}(\sigma) \in \{\pm 1\}$. We get

$$\det(T_\sigma) = \operatorname{sgn}(\sigma).$$

Corollary 4.4.1. *Let $\mathbb{F}$ be a field. Any finite group G is isomorphic to a subgroup of $\mathrm{GL}_n(\mathbb{F})$ for some n.*

Proof. By Cayley's theorem $G \hookrightarrow S_n$ for some n, and we have shown $S_n \hookrightarrow \mathrm{GL}_n(\mathbb{F})$. □

4.5. **Multiplicativity of the determinant.**

Theorem 4.5.1. *We have for any two matrices A, B in $M_n(\mathbb{F})$,*

$$\det(AB) = \det(A)\det(B).$$

Proof. We first introduce some notation: For vectors $r = (r_1, \ldots, r_n), s = (s_1, \ldots, s_n)$ we let

$$\langle r, s\rangle = \sum_{i=1}^{n} r_i s_i.$$

We allow s to be a column vector in this definition. We note the following properties:

$$\langle r, s\rangle = \langle s, r\rangle, \quad \langle r, s + s'\rangle = \langle r, s\rangle + \langle r, s'\rangle, \quad \langle r, \alpha s\rangle = \alpha\langle r, s\rangle, \qquad \text{for } r, s, s' \in \mathbb{F}^n, \alpha \in \mathbb{F}.$$

Let A be a $n \times n$ matrix with rows

$$A = \begin{pmatrix} u_1 \\ \hline \\ \vdots \\ \hline u_n \end{pmatrix},$$

and let B be a $n \times n$ matrix with columns

$$B = \left(v_1 | \cdots | v_n\right).$$

[5] This gives the interesting relation $T_{\sigma^{-1}} = T_\sigma^t$. Because $\sigma \mapsto T_\sigma$ is a group homomorphism we may conclude that $T_\sigma^{-1} = T_\sigma^t$. Of course for a general matrix this doesn't hold.

In this notation,

$$AB = \left(\langle u_i, v_j\rangle\right)_{i,j=1}^{n} = \begin{pmatrix} \langle u_1, v_1\rangle & \dots & \langle u_1, v_n\rangle \\ \vdots & & \vdots \\ \langle u_n, v_1\rangle & \dots & \langle u_n, v_n\rangle \end{pmatrix}.$$

Having set this notation, we begin the proof. Consider the function

$$h : M_n(\mathbb{F}) \to \mathbb{F}, \qquad h(B) = \det(AB).$$

We prove that h has properties (1) - (3) of Theorem 4.3.1.

(1) We have

$$h((v_1 \dots \alpha v_i \dots v_n)) = \det \begin{pmatrix} \langle u_1,v_1\rangle & \dots & \alpha\langle u_1,v_i\rangle & \dots & \langle u_1,v_n\rangle \\ \vdots & & & & \vdots \\ \langle u_n,v_1\rangle & \dots & \alpha\langle u_n,v_i\rangle & \dots & \langle u_n,v_n\rangle \end{pmatrix}.$$

This is equal to

$$\alpha \det \begin{pmatrix} \langle u_1,v_1\rangle & \dots & \langle u_1,v_n\rangle \\ \vdots & & \vdots \\ \langle u_n,v_1\rangle & \dots & \langle u_n,v_n\rangle \end{pmatrix} = \alpha \det(AB) = \alpha \cdot h(B).$$

(2) We have

$$h((v_1 \dots \alpha v_i + v_i' \dots v_n)) = \det \begin{pmatrix} \langle u_1,v_1\rangle & \dots & \langle u_1,v_i\rangle+\langle u_1,v_i'\rangle & \dots & \langle u_1,v_n\rangle \\ \vdots & & & & \vdots \\ \langle u_n,v_1\rangle & \dots & \langle u_n,v_i\rangle+\langle u_n,v_i'\rangle & \dots & \langle u_n,v_n\rangle \end{pmatrix},$$

which is equal to

$$\det \begin{pmatrix} \langle u_1,v_1\rangle & \dots & \langle u_1,v_i\rangle & \dots & \langle u_1,v_n\rangle \\ \vdots & & & & \vdots \\ \langle u_n,v_1\rangle & \dots & \langle u_n,v_i\rangle & \dots & \langle u_n,v_n\rangle \end{pmatrix} + \det \begin{pmatrix} \langle u_1,v_1\rangle & \dots & \langle u_1,v_i'\rangle & \dots & \langle u_1,v_n\rangle \\ \vdots & & & & \vdots \\ \langle u_n,v_1\rangle & \dots & \langle u_n,v_i'\rangle & \dots & \langle u_n,v_n\rangle \end{pmatrix}$$
$$= h((v_1 \dots v_i \dots v_n)) + h((v_1 \dots v_i' \dots v_n)).$$

(3) Suppose that $v_i = v_j$ for $i < j$. Then

$$h((v_1 \dots v_i \dots v_j \dots v_n)) = \det \begin{pmatrix} \langle u_1,v_1\rangle & \dots & \langle u_1,v_i\rangle & \dots & \langle u_1,v_j\rangle & \dots & \langle u_1,v_n\rangle \\ \vdots & & & & & & \vdots \\ \langle u_n,v_1\rangle & \dots & \langle u_n,v_i\rangle & \dots & \langle u_n,v_j\rangle & \dots & \langle u_n,v_n\rangle \end{pmatrix}.$$

This last matrix has its i-th column equal to its j-th column so the determinant vanishes and we get $h((v_1 \dots v_i \dots v_j \dots v_n)) = 0$.

Lemma 4.5.2. *Let $H : M_n(\mathbb{F}) \to \mathbb{F}$ be a function satisfying properties (1) - (3) of Theorem 4.3.1 then $H = \gamma \cdot \det$, where $\gamma = H(I_n)$.*

We note that with the lemma the proof of the theorem is complete: since h satisfies the assumptions of the lemma, we have

$$h(B) = h(I_n) \cdot \det(B).$$

Since $h(B) = \det(AB), h(I_n) = \det(A)$, we have

$$\det(AB) = \det(A) \cdot \det(B).$$

It remains to prove the lemma.

Proof of Lemma. Let us write $v_j = \sum_{i=1}^{n} a_{ij} e_i$, where e_i is the i-th element of the standard basis, written as a column vector (all the entries of e_i are zero except for the i-th entry which is 1). Then we have

$$(v_1 v_2 \dots v_n) = (a_{ij}).$$

Now, by the linearity properties of H, we have

$$\begin{aligned} H(v_1 \dots v_n) &= H(\sum_{i=1}^{n} a_{i1}e_i, \dots, \sum_{i=1}^{n} a_{in}e_i) \\ &= \sum_{(i_1,\dots,i_n)} a_{i_1 1} \cdots a_{i_n n} \cdot H(e_{i_1}, \dots, e_{i_n}), \end{aligned}$$

where in the summation each $1 \leq i_j \leq n$. If in this sum $i_\ell = i_k$ for some $\ell \neq k$ then $H(e_{i_1}, \dots, e_{i_n}) = 0$. We therefore have,

$$\begin{aligned} H(v_1 \dots v_n) &= \sum_{(i_1,\dots,i_n), i_j \text{ distinct}} a_{i_1 1} \cdots a_{i_n n} \cdot H(e_{i_1}, \dots, e_{i_n}) \\ &= \sum_{\sigma \in S_n} a_{\sigma(1)\,1} \cdots a_{\sigma(n)\,n} H(e_{\sigma(1)}, \dots, e_{\sigma(n)}). \end{aligned}$$

Now, inspection of the proof of Corollary 4.3.2 shows that it holds for H as well. We therefore get,

$$\begin{aligned} H(v_1 \dots v_n) &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)\,1} \cdots a_{\sigma(n)\,n} H(e_1, \dots, e_n) \\ &= \gamma \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)\,1} \cdots a_{\sigma(n)\,n} \\ &= \gamma \cdot \det(v_1 \dots v_n). \end{aligned}$$

□

4.6. **Laplace's theorem and the adjoint matrix.** Consider again the formula for the determinant of an $n \times n$ matrix $A = (a_{ij})$:

$$\det(a_{ij}) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)\,1} \cdots a_{\sigma(n)\,n}.$$

Choose an index j. We have then

$$\det(A) = \sum_{i=1}^{n} a_{ij} \sum_{\{\sigma : \sigma(j) = i\}} \operatorname{sgn}(\sigma) \prod_{\ell \neq j} a_{\sigma(\ell)\ell}. \tag{4}$$

Let $A_{i,j}$ be the ij-**minor** of A. This is the matrix obtained from A by deleting the i-th row and j-th column. Let A^{ij} be the ij-**cofactor** of A, which is defined as

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

Note that A_{ij} is an $(n-1) \times (n-1)$ matrix, while A^{ij} is a scalar.

Lemma 4.6.1. $\sum_{\{\sigma : \sigma(j) = i\}} \operatorname{sgn}(\sigma) \prod_{\ell \neq j} a_{\sigma(\ell)\ell} = A^{ij}$.

Proof. Let $f = (j \;\; j+1 \;\; \dots \;\; n), g = (i \;\; i+1 \;\; \dots \;\; n)$ be the cyclic permutations, considered as an elements of S_n. Define

$$b_{\alpha\beta} = a_{g(\alpha)\, f(\beta)}.$$

Note that

$$A_{ij} = (b_{\alpha\beta})_{1 \leq \alpha, \beta \leq n-1}$$

Then,

$$\sum_{\{\sigma:\sigma(j)=i\}} \operatorname{sgn}(\sigma) \prod_{\ell\neq j} a_{\sigma(\ell)\,\ell} = \sum_{\{\sigma:\sigma(j)=i\}} \operatorname{sgn}(\sigma) \prod_{\ell\neq j} b_{g^{-1}(\sigma(\ell))\,f^{-1}(\ell)}$$
$$= \sum_{\{\sigma:\sigma(j)=i\}} \operatorname{sgn}(\sigma) \prod_{t=1}^{n-1} b_{g^{-1}(\sigma(f(t)))\,t},$$

because $\{f^{-1}(\ell) : \ell \neq j\} = \{1, 2, \ldots, n-1\}$. Now, the permutations $\sigma' = g^{-1}\sigma f$, where $\sigma(j) = i$ are precisely the permutations in S_n fixing n and thus may be identified with S_{n-1}. Moreover, $\operatorname{sgn}(\sigma) = \operatorname{sgn}(f)\operatorname{sgn}(g)\operatorname{sgn}(g^{-1}\sigma f) = (-1)^{(n-i)+(n-j)}\operatorname{sgn}(g^{-1}\sigma f) = (-1)^{i+j}\operatorname{sgn}(g^{-1}\sigma f)$. Thus, we have,

$$\sum_{\{\sigma:\sigma(j)=i\}} \operatorname{sgn}(\sigma) \prod_{\ell\neq j} a_{\sigma(\ell)\,\ell} = (-1)^{i+j} \sum_{\sigma'\in S_{n-1}} \operatorname{sgn}(\sigma') \prod_{t=1}^{n-1} b_{\sigma'(t)\,t} = A^{ij}.$$

□

Theorem 4.6.2 (Laplace)**.** *For any i or j we have*

$$\det(A) = a_{i1}A^{i1} + a_{i2}A^{i2} + \cdots + a_{in}A^{in}, \qquad \textit{(developing by row)}$$

and

$$\det(A) = a_{1j}A^{1j} + a_{2j}A^{2j} + \cdots + a_{nj}A^{nj}, \qquad \textit{(developing by column).}$$

Also, if $\ell \neq j$ then (for any j)

$$\sum_{i=1}^{n} a_{ij}A^{i\ell} = 0,$$

and if $\ell \neq i$ then

$$\sum_{j=1}^{n} a_{ij}A^{\ell j} = 0.$$

Example 4.6.3. We introduce also the notation

$$|a_{ij}| = \det(a_{ij}).$$

We have then:

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = 1 \cdot \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - 2 \cdot \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + 3 \cdot \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix}$$
$$= 1 \cdot \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - 4 \cdot \begin{vmatrix} 2 & 3 \\ 8 & 9 \end{vmatrix} + 7 \cdot \begin{vmatrix} 2 & 3 \\ 5 & 6 \end{vmatrix}$$
$$= -4 \cdot \begin{vmatrix} 2 & 3 \\ 8 & 9 \end{vmatrix} + 5 \cdot \begin{vmatrix} 1 & 3 \\ 7 & 9 \end{vmatrix} - 6 \cdot \begin{vmatrix} 1 & 2 \\ 7 & 8 \end{vmatrix}.$$

Here we developed the determinant according to the first row, the first column and the second row, respectively. When we develop according to a certain row we sum the elements of the row, each multiplied by the determinant of the matrix obtained by erasing the row and column containing the element we are at, only that we also need to introduce signs. The signs are easy to remember by the following checkerboard picture:

$$\begin{pmatrix} + & - & + & - & \ldots \\ - & + & - & + & \ldots \\ + & - & + & - & \ldots \\ - & + & - & + & \ldots \\ \vdots & & & & \end{pmatrix}.$$

Proof. The formulas for the developing according to columns are immediate consequence of Equation (4) and Lemma 4.6.1. The formula for rows follows formally from the formula for columns using $\det A = \det A^t$.

The identity $\sum_{i=1}^n a_{ij}A^{i\ell} = 0$ can be obtained by replacing the ℓ-th column of A by its j-th column (keeping the j-column as it is). This doesn't affect the cofactors $A^{i\ell}$ and changes the elements $a_{i\,\ell}$ to a_{ij}. Thus, the expression $\sum_{i=1}^n a_{ij}A^{i\ell}$ is the determinant of the new matrix. But this matrix has two equal columns so its determinant is zero! A similar argument applies to the last equality. □

Definition 4.6.4. Let $A = (a_{ij})$ be an $n \times n$ matrix. Define the **adjoint** of A to be the matrix

$$\mathrm{Adj}(A) = (c_{ij}), \qquad c_{ij} = A^{ji}.$$

That is, the ij entry of $\mathrm{Adj}(A)$ is the ji-cofactor of A.

Theorem 4.6.5.

$$\mathrm{Adj}(A) \cdot A = A \cdot \mathrm{Adj}(A) = \det(A) \cdot I_n.$$

Proof. We prove one equality; the second is completely analogous. The proof is just by noting that the ij entry of the product $\mathrm{Adj}(A) \cdot A$ is

$$\sum_{\ell=1}^{n} \mathrm{Adj}(A)_{i\ell} \cdot a_{\ell j} = \sum_{\ell=1}^{n} a_{\ell j} \cdot A^{\ell i}.$$

According to Theorem 4.6.2 this is equal to $\det(A)$ if $i = j$ and equal to zero if $i \neq j$. □

Corollary 4.6.6. *The matrix A is invertible if and only if $\det(A) \neq 0$. If $\det(A) \neq 0$ then*

$$A^{-1} = \frac{1}{\det(A)} \cdot \mathrm{Adj}(A).$$

Proof. Suppose that A is invertible. There is then a matrix B such that $AB = I_n$. Then $\det(A)\det(B) = \det(AB) = \det(I_n) = 1$ and so $\det(A)$ is invertible (and, in fact, $\det(A^{-1}) = \det(B) = \det(A)^{-1}$).

Conversely, if $\det(A)$ is invertible then the formulas,

$$\mathrm{Adj}(A) \cdot A = A \cdot \mathrm{Adj}(A) = \det(A) \cdot I_n,$$

show that

$$A^{-1} = \det(A)^{-1} \mathrm{Adj}(A).$$

□

Corollary 4.6.7. *Let $B = \{v_1, \ldots, v_n\}$ be a set of n vectors in $\mathbb{F}^n$. Then B is a basis if and only if* $\det(v_1 v_2 \ldots v_n) \neq 0$.

Proof. If B is a basis then $(v_1 v_2 \ldots v_n) = {}_{St}M_B$. Since ${}_{St}M_{BB}M_{St} = {}_{B}M_{StSt}M_B = I_n$ then $(v_1 v_2 \ldots v_n)$ is invertible and so $\det(v_1 v_2 \ldots v_n) \neq 0$.

If B is not a basis then one of its vectors is a linear combination of the preceding vectors. By renumbering the vectors we may assume this vector is v_n. Then $v_n = \sum_{i=1}^{n-1} \alpha_i v_i$. We get

$$\begin{aligned}
\det(v_1 v_2 \ldots v_n) &= \det(v_1 v_2 \ldots v_{n-1}(\sum_{i=1}^{n-1} \alpha_i v_i)) \\
&= \sum_{i=1}^{n-1} \alpha_i \det(v_1 v_2 \ldots v_{n-1} v_i) \\
&= 0,
\end{aligned}$$

because in each determinant there are two columns that are the same. □

5. Systems of linear equations

Let $\mathbb{F}$ be a field. We have the following dictionary:

$$\boxed{\text{system of } m \text{ linear equations in } n \text{ variables}} \leftrightarrow \boxed{\text{matrix } A \text{ in } M_{m\times n}(\mathbb{F})} \leftrightarrow \boxed{\text{linear map } T:\mathbb{F}^n\to\mathbb{F}^m}$$

$$\begin{matrix} a_{11}x_1+\cdots+a_{1n}x_n \\ \vdots \\ a_{m1}x_1+\cdots+a_{mn}x_n \end{matrix} \leftrightarrow A=\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \leftrightarrow \begin{matrix} T:\mathbb{F}^n\longrightarrow\mathbb{F}^m \\ T\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}=A\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \end{matrix}$$

In particular: $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ solves the system

$$\begin{matrix} a_{11}x_1+\cdots+a_{1n}x_n=b_1 \\ \vdots \\ a_{m1}x_1+\cdots+a_{mn}x_n=b_m, \end{matrix}$$

if and only if

$$A\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}=\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix},$$

if and only if

$$T\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}=\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

A special case is $\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}=\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$. We see that $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ solves the **homogenous system of equations**

$$\begin{matrix} a_{11}x_1+\cdots+a_{1n}x_n=0 \\ \vdots \\ a_{m1}x_1+\cdots+a_{mn}x_n=0, \end{matrix}$$

if and only if

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\in \mathrm{Ker}(T).$$

We therefore draw the following corollary:

Corollary 5.0.1. *The solution set to a non homogenous system of equations*

$$\begin{matrix} a_{11}x_1+\cdots+a_{1n}x_n=b_1 \\ \vdots \\ a_{m1}x_1+\cdots+a_{mn}x_n=b_m, \end{matrix}$$

is either empty or has the form $\mathrm{Ker}(T)+\begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix}$, *where* $\begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix}$ *is (any) solution to the non homogenous system. In particular, if* $\mathrm{Ker}(T)=\{0\}$, *that is if the homogenous system has only the zero solution, then*

any non homogenous system

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m, \end{aligned}$$

has at most one solution.

We note also the following:

Corollary 5.0.2. *The non homogenous system*

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m, \end{aligned}$$

has a solution if and only if $\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$ *is in the image of* T*, if and only if*

$$\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \text{Span}\left\{ \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \ldots, \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} \right\}.$$

Corollary 5.0.3. *If $n > m$ there is a non-zero solution to the homogenous system of equations. That is, if the number of variables is greater than the number of equations there's always a non-trivial solution.*

Proof. We have $\dim(\text{Ker}(T)) = \dim(\mathbb{F}^n) - \dim(\text{Im}(T)) \geq n - m > 0$, therefore $\text{Ker}(T)$ has a non-zero vector. □

Definition 5.0.4. The dimension of $\text{Span}\left\{ \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \ldots, \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} \right\}$, i.e., the dimension of $\text{Im}(T)$, is called the **column rank** of A and is denoted $\text{rk}_c(A)$. We also call $\text{Im}(T)$ the **column space** of A.

Similarly, the **row space** of A is the subspace of $\mathbb{F}^n$ spanned by the rows of A. Its dimension is called the **row rank** of A and is denoted $\text{rk}_r(A)$.

Example 5.0.5. Consider the matrix

$$A = \begin{pmatrix} 1 & 2 & 3 & -1 \\ 3 & 4 & 7 & -3 \\ 0 & 1 & 1 & 0 \end{pmatrix}.$$

Its column rank is 2 since the third column is the sum of the first two and the fourth column is the second minus the third. The first two columns are independent (over any field). Its row rank is also two as the first and third rows are independent and the second row is 3×(the first row) - 2×(the third row). As we shall see later, this is no accident. It is always true that $\text{rk}_c(A) = \text{rk}_r(A)$, though the row space is a subspace of $\mathbb{F}^n$ and the column space is a subspace of $\mathbb{F}^m$!

We note the following identities:

$$\text{rk}_r(A) = \text{rk}_c(A^t), \qquad \text{rk}_c(A) = \text{rk}_r(A^t).$$

5.1. **Row reduction.** Let A be an $m \times n$ matrix with rows $R_1, \dots, R_m$. They span the row space of A. The row space can be understood as the space of linear conditions a solution to the homogenous system satisfies. Let $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ be a solution to the homogenous system

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= 0 \end{aligned}.$$

We can also express it by saying

$$\langle R_1, x \rangle = \cdots = \langle R_m, x \rangle = 0.$$

Then

$$\langle \sum \alpha_i R_i, x \rangle = \sum \alpha_i \langle R_i, x \rangle = 0.$$

This shows that $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ satisfies any linear condition in the row space.

Corollary 5.1.1. *Any homogenous system on m equations in n unknowns can be reduced to a system of m' equations in n unknowns where $m' \leq n$.*

Proof. Indeed, x solves the system

$$\langle R_1, x \rangle = \cdots = \langle R_m, x \rangle = 0$$

if and only if

$$\langle S_1, x \rangle = \cdots = \langle S_{m'}, x \rangle = 0,$$

where $S_1, \dots, S_{m'}$ are a basis of the row space. The row space is a subspace of $\mathbb{F}^n$ and so $m' \leq n$. □

Let again A be the matrix giving a system of linear equations and $R_1, \dots, R_m$ its rows. **Row reduction** is (the art of) repeatedly performing any of the following operations on the rows of A in succession:

$R_i \mapsto \lambda R_i, \quad \lambda \in \mathbb{F}^\times$ (multiplying a row by a non-zero scalar)
$R_i \leftrightarrow R_j$ (exchanging two rows)
$R_i \mapsto R_i + \lambda R_j, \quad i \neq j$ (adding any multiple of a row to another row)

Proposition 5.1.2. *Two linear systems of equations obtained from each other by row reduction have the same space of solutions to the homogenous systems of equations they define.*

Proof. This is clear since the row space stays the same (easy to verify!). □

Remark 5.1.3. Since row reduction operations are invertible, it is easy to check that row reduction defines an equivalence relation on $m \times n$ matrices.

5.2. **Matrices in reduced echelon form.**

Definition 5.2.1. A matrix is called in **reduced echelon form** if it has the shape

$$\begin{pmatrix} 0 & \dots & 0 & a_{1i_1} & & & & & \\ 0 & & & \dots & 0 & a_{2i_2} & & & \\ 0 & & & & & \dots & 0 & a_{3i_3} & \\ \vdots & & & & & & & & \vdots \\ 0 & & & & & & \dots & \dots 0 & a_{ri_r} & \dots \\ 0 & & & & & & & & & 0 \\ \vdots & & & & & & & & & \vdots \\ 0 & & & & & & & & & 0 \end{pmatrix},$$

where each $a_{ki_k} = 1$ and for every $\ell \neq k$ we have $a_{\ell i_k} = 0$. The columns $i_1, \dots, i_r$ are distinguished. Notice that they are just part of the standard basis – they are equal to $e_1, \dots, e_r$.

Example 5.2.2. The real matrix

$$\begin{pmatrix} 0 & 2 & 1 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

is in echelon form but not in reduced echelon form. By performing row operations (do $R_1 \mapsto R_1 - R_2$, then $R_1 \mapsto \frac{1}{2}R_1$) we can bring it to reduced echelon form:

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

Theorem 5.2.3. *Every matrix is equivalent by row reduction to a matrix in reduced echelon form.*

We shall not prove this theorem (it is not hard to prove, say by induction on the number of columns), but we shall make use of it. We illustrate the theorem by an example.

Example 5.2.4. $\begin{pmatrix} 3 & 2 & 0 \\ 1 & 1 & 1 \\ 2 & 1 & -1 \end{pmatrix} \to_{R_1 \leftrightarrow R_2} \begin{pmatrix} 1 & 1 & 1 \\ 3 & 2 & 0 \\ 2 & 1 & -1 \end{pmatrix} \to_{R_2 \mapsto R_2 - 3R_1, R_3 \mapsto R_3 - 2R_1} \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & -3 \\ 0 & -1 & -3 \end{pmatrix} \to_{R_3 \mapsto R_3 - R_2, R_2 \mapsto -R_2}$

Theorem 5.2.5. *Two $m \times n$ matrices in reduced echelon form having the same row space are equal.*

Before proving this theorem, let us draw some corollaries:

Corollary 5.2.6. *Every matrix is row equivalent to a unique matrix in reduced echelon form.*

Proof. Suppose A is row-equivalent to two matrices B, B' in reduced echelon form. Then A and B and A and B' have the same row space. Thus, B and B' have the same row space, hence are equal. □

Corollary 5.2.7. *Two matrices with the same row space are row equivalent.*

Proof. Let A, B be two matrices with the same row space. Then A is row equivalent to A' in reduced echelon form and B is row equivalent to B' in reduced echelon form. Since A', B' have the same row space they are equal and it follows that A is row equivalent to B. □

Proof. (Of Theorem 5.2.5) Write

$$A = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_\alpha \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad B = \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_\beta \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

where the R_i, S_j are the non-zero rows of the matrices in reduced echelon form. We have

$$R_i = (0, \dots 0, a_{ij_i} = 1, \dots)$$

and $a_{\ell j_i} = 0$ for $\ell \neq i$. We claim that $R_1, \dots, R_\alpha$ is a basis for the row space. Indeed, if $0 = \sum c_i R_i$ then since $\sum c_i R_i = (\dots c_1 \dots c_2 \dots c_\alpha \dots)$, where the places the c_i's appear are $j_1, j_2, \dots, j_\alpha$, we must have $c_i = 0$ for all i. An independent spanning set is a basis. It therefore follows also that $\alpha = \beta$, there is the same number of rows in A and B.

Let us also write

$$S_i = (0, \dots 0, b_{ik_i} = 1, \dots).$$

Suppose we know already that $R_{i+1} = S_{i+1}, \dots, R_\alpha = S_\alpha$ for some $i \leq \alpha$ and let us prove that for i. Suppose that $k_i > j_i$. We have

$$R_i = (0 \dots 0 \; a_{ij_i} \dots\dots a_{in})$$

and

$$S_i = (0 \dots 0 \dots 0 \; b_{ik_i} \dots b_{in})$$

with $a_{ij_i} = b_{ik_i} = 1$. Now, for some scalars t_a we have

$$S_i = t_1 R_1 + \dots + t_\alpha R_\alpha = (\dots t_1 \dots t_2 \dots t_i \dots t_\alpha \dots),$$

where t_a appears in the j_a place. Now, at those place $j_1, \dots, j_i$ the entry of S_i is zero (because also $j_i < k_i$). We conclude that $t_1 = \dots = t_i = 0$ and so

$$S_i = t_{i+1} R_{i+1} + \dots + t_\alpha R_\alpha = t_{i+1} S_{i+1} + \dots + t_\alpha S_\alpha,$$

which contradicts the independence of the vectors $\{S_1, \dots, S_\alpha\}$. By symmetry, $k_i < j_i$ is also not possible and so $k_i = j_i$.

Again, we write

$$S_i = t_1 R_1 + \dots + t_\alpha R_\alpha = (\dots t_1 \dots t_2 \dots t_i \dots t_\alpha \dots),$$

where t_a appears in the j_a place. The same reasoning tells us that $t_1 = \dots = t_{i-1} = 0$ and so

$$S_i = t_i R_i + t_{i+1} S_{i+1} + \dots + t_\alpha S_\alpha = (0 \dots 0 \; t_i \dots t_{i+1} \dots t_\alpha \dots),$$

where t_a appears in the k_a place. However, at each k_a place where $a > i$ the coordinate of S_i is zero and at the k_i coordinate it is one. It follows that $t_i = 1, t_{i+1} = \dots = t_\alpha = 0$ and so $S_i = R_i$. □

5.3. **Row rank and column rank.**

Theorem 5.3.1. *Let $A \in M_{m \times n}(\mathbb{F})$. Then*

$$\mathrm{rk}_r(A) = \mathrm{rk}_c(A).$$

Proof. Let $T : \mathbb{F}^n \to \mathbb{F}^m$ be the associated linear map. Then

$$\mathrm{rk}_c(A) = \dim(\mathrm{Im}(T)) = n - \dim(\mathrm{Ker}(T)).$$

Let $\widetilde{A}$ be the matrix in reduced echelon form which is row equivalent to A. Since $\mathrm{Ker}(T)$ are the solutions to the homogenous system of equations defined by A, it is also the solutions to the homogenous system of equations defined by $\widetilde{A}$. If we let $\widetilde{T}$ be the linear transformation associated to $\widetilde{A}$ then

$$\mathrm{Ker}(T) = \mathrm{Ker}(\widetilde{T}).$$

We therefore obtain

$$\mathrm{rk}_c(A) = n - \dim(\mathrm{Ker}(\widetilde{T})) = \dim(\mathrm{Im}(\widetilde{T})).$$

(We should remark at this point that this is not a priori obvious as the column space of A and $\widetilde{A}$ are completely different!)

Now $\dim(\mathrm{Im}(\widetilde{T})) = \mathrm{rk}_c(\widetilde{A})$ is equal to the number of non-zero rows in $\widetilde{A}$. Indeed, if $\widetilde{A}$ has k non-zero rows than clearly we can get at most k non-zero entries in every vector in the column space of $\widetilde{A}$. On the other hand, the distinguished columns of $\widetilde{A}$ (where the steps occur) give us the vectors $e_1, \ldots, e_k$ and so we see that the dimension is precisely k. However, the number of non-zero rows is precisely the basis for the row space that is provided by those non zero rows. That is,

$$\dim(\mathrm{Im}(\widetilde{T})) = \mathrm{rk}_r(\widetilde{A}) = \mathrm{rk}_r(A),$$

because A and $\widetilde{A}$ have the same row space. □

Corollary 5.3.2. *The dimension of the space of solutions to the homogenous system of equations is* $n - \mathrm{rk}_r(A)$*, namely, the codimension of the space of linear conditions* $\quad$ *row-space*(A).

Proof. Indeed, this is $\dim(\mathrm{Ker}(T)) = n - \dim(\mathrm{Im}(T)) = n - \mathrm{rk}_c(A) = n - \mathrm{rk}_r(A)$. □

5.4. **Cramer's rule.** Consider a non homogenous system of n equations in n unknowns:

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{n1}x_1 + \cdots + a_{nn}x_n &= b_n. \end{aligned} \tag{5}$$

Introduce the notation A for the coefficients and write A as n-columns vectors in $\mathbb{F}^n$:

$$A = \left(v_1|v_2|\cdots|v_n\right).$$

Let

$$b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

Theorem 5.4.1. *Assume that* $\det(A) \neq 0$*. Then there is a unique solution* $(x_1, x_2, \ldots, x_n)$ *to the non homogenous system (5). Let* A_i *be the matrix obtained by replacing the i-th column of A by b:*

$$A_i = \left(v_1|\ldots|v_{i-1}|\, b\, |v_{i+1}|\cdots|v_n\right).$$

Then,

$$x_i = \frac{\det(A_i)}{\det(A)}.$$

Proof. Let T be the associated linear map. First, since $\mathrm{Ker}(T) = \{0\}$ and the solutions are a coset of $\mathrm{Ker}(T)$, there is at most one solution. Secondly, since $\dim(\mathrm{Im}(T)) = \dim(\mathbb{F}^n) - \dim(\mathrm{Ker}(T)) = n$, we have $\mathrm{Im}(T) = \mathbb{F}^n$ and thus for any vector b there is a solution to the system (5).

Now,

$$\begin{aligned}\det(A_i) &= \det\Big(v_1|\dots|v_{i-1}|x_1v_1+\dots+x_nv_n|v_{i+1}|\cdots|v_n\Big)\\ &= \sum_{j=1}^{n} x_j \det\Big(v_1|\dots|v_{i-1}|v_j|v_{i+1}|\cdots|v_n\Big)\\ &= x_i \det\Big(v_1|\dots|v_{i-1}|v_i|v_{i+1}|\cdots|v_n\Big)\\ &= x_i \det(A).\end{aligned}$$

(In any determinant in the sum there are two vectors that are equal, except when we deal with the i-th summand.) □

5.5. **About solving equations in practice and calculating the inverse matrix.**

Definition 5.5.1. An **elementary matrix** is a square matrix having one of the following shapes:

(1) A diagonal matrix $\mathrm{diag}[1,\dots,1,\lambda,1,\dots,1]$ for some $\lambda \in \mathbb{F}^\times$.
(2) The image of a transposition. That is, a matrix D such that for some $i < j$ has entries $d_{kk} = 1$ for $k \notin \{i,j\}$, $d_{ij} = d_{ji} = 1$ and all its other entries are zero.
(3) A matrix D whose diagonal elements are all 1, has an entry $d_{ij} = \lambda$ for some $i \neq j$ and the rest of its entries are zero. Here λ can be any element of $\mathbb{F}$.

Let E be an elementary $m \times m$ matrix and A a matrix with m rows, $R_1, \dots, R_m$. Consider the product EA.

(1) If E if of type (1) then the rows of EA are
$$R_1, \dots, R_{i-1}, \lambda R_i, R_{i+1}, \dots, R_m.$$
(2) If E if of type (2) then the rows of EA are
$$R_1, \dots, R_{i-1}, R_j, R_{i+1}, \dots, R_{j-1}, R_i, R_{j+1}, \dots, R_m.$$
(3) If E is of type (3) then EA has rows
$$R_1, \dots, R_{i-1}, R_i + \lambda R_j, R_{i+1}, \dots, R_m$$

.

It is easy to check that any elementary matrix E is invertible. Any iteration of row reduction operations (such as reducing to the reduced echelon form) can be viewed as

$$A \rightsquigarrow EA,$$

where E is product of elementary matrices and in particular invertible. Therefore:

$$A\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \quad \Leftrightarrow \quad (EA)\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = E\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

This, of course, just means that the following. If we perform on the vector of conditions $\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$ exactly the same operations we perform when row reducing A then a solution to the reduced

system

$$(EA)\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = E\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

is a solution to the original system and vice-versa.

This reduction can be done simultaneously for several conditions, namely, we can attempt to solve

$$A\begin{pmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_n & y_n \end{pmatrix} = \begin{pmatrix} b_1 & c_1 \\ \vdots & \vdots \\ b_m & c_m \end{pmatrix}.$$

Again,

$$A\begin{pmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_n & y_n \end{pmatrix} = \begin{pmatrix} b_1 & c_1 \\ \vdots & \vdots \\ b_m & c_m \end{pmatrix} \quad \Leftrightarrow \quad (EA)\begin{pmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_n & y_n \end{pmatrix} = E\begin{pmatrix} b_1 & c_1 \\ \vdots & \vdots \\ b_m & c_m \end{pmatrix}.$$

We can of course do this process for any number of condition vectors $\begin{pmatrix} x_1 & y_1 & z_1 & \cdots \\ \vdots & \vdots & \vdots & \\ x_n & y_n & z_n & \cdots \end{pmatrix}$. We note that if A is a square matrix, to find A^{-1} is to solve the particular system:

$$A\begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nn} \end{pmatrix} = I_n.$$

If A is invertible then the matrix in reduced echelon form corresponding to A must be the identity matrix, because this is the only matrix in reduced echelon form having rank n. Therefore, there is a product E of elementary matrices such that $EA = I_n$. We conclude the following:

Corollary 5.5.2. *Let A be an $n \times n$ matrix. Perform row operations on A, say $A \rightsquigarrow EA$ so that EA is in reduced echelon form and at the same time perform the same operations on I_n, $I_n \rightsquigarrow EI_n$. A is invertible if and only if $EA = I_n$; in that case E is the inverse of A and we get A^{-1} by applying to the identity matrix the same row operations we apply to A.*

Example 5.5.3. Let us find out if $A = \begin{pmatrix} 7 & 11 & 0 \\ 0 & 0 & 1 \\ 5 & 8 & 0 \end{pmatrix}$ is invertible and what is its inverse. First, the determinant of A is $7 \cdot 0 \cdot 0 + 0 \cdot 8 \cdot 0 + 5 \cdot 11 \cdot 1 - 0 \cdot 0 \cdot 5 - 1 \cdot 8 \cdot 7 - 0 \cdot 11 \cdot 0 = -1$. Thus, A is

invertible. To find the inverse we do

$$\begin{pmatrix} 7 & 11 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 5 & 8 & 0 & 0 & 0 & 1 \end{pmatrix} \to_{R_1 \mapsto 5R_1, R_3 \mapsto 5R_3}$$

$$\begin{pmatrix} 35 & 55 & 0 & 5 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 35 & 56 & 0 & 0 & 0 & 7 \end{pmatrix} \to_{R_3 \mapsto R_3 - R_1}$$

$$\begin{pmatrix} 35 & 55 & 0 & 5 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -5 & 0 & 7 \end{pmatrix} \to_{R_1 \mapsto R_1 - 55R_3}$$

$$\begin{pmatrix} 35 & 0 & 0 & 280 & 0 & -385 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -5 & 0 & 7 \end{pmatrix} \to_{R_1 \mapsto \frac{1}{35} R_1, R_2 \leftrightarrow R_3}$$

$$\begin{pmatrix} 1 & 0 & 0 & 8 & 0 & -11 \\ 0 & 1 & 0 & -5 & 0 & 7 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

Thus, the inverse of A is

$$\begin{pmatrix} 8 & 0 & -11 \\ -5 & 0 & 7 \\ 0 & 1 & 0 \end{pmatrix}.$$

6. The dual space

The **dual vector space** is the space of linear functions on a vector space. As such, it is a natural object to consider and arises in many situations. It is also perhaps the first example of **duality** you will learn. The concept of duality is a key concept in mathematics.

6.1. Definition and first properties and examples.

Definition 6.1.1. Let $\mathbb{F}$ be a field and V a finite dimensional vector space over $\mathbb{F}$. We let

$$V^* = \mathrm{Hom}(V, \mathbb{F}).$$

Since $\mathbb{F}$ is a vector space over $\mathbb{F}$, we know by a general result, proven in the assignments, that V^* is a vector space, called the **dual space**, under the operations

$$(S+T)(v) = S(v) + T(v), \quad (\alpha S)(v) = \alpha S(v).$$

The elements of V^* are often called **linear functionals**.

Recall the general formula

$$\dim \mathrm{Hom}(V, W) = \dim(V) \cdot \dim(W),$$

proved in Corollary 3.8.2. This implies that $\dim V^* = \dim V$. This also follows from the following proposition.

Proposition 6.1.2. *Let V be a finite dimensional vector space. Let $B = \{b_1, \ldots, b_n\}$ be a basis for V. There is then a unique basis $B^* = \{f_1, \ldots, f_n\}$ of V^* such that*

$$f_i(b_j) = \delta_{ij}.$$

*The basis B^* is called the **dual basis**.*

Proof. Given an index i, there is a unique linear map,

$$f_i : V \to \mathbb{F},$$

such that,

$$f_i(b_j) = \delta_{ij}.$$

This is a special case of a general result proven in the assignments. We therefore get functions

$$f_1, \ldots, f_n : V \to \mathbb{F}.$$

We claim that they form a basis for V^*. Firstly, $\{f_1, \ldots, f_n\}$ are linearly independent. Suppose that $\sum \alpha_i f_i = 0$, where 0 stands for the constant map with value $0_{\mathbb{F}}$. Then, for every j, we have $0 = (\sum \alpha_i f_i)(b_j) = \sum_i \alpha_i \delta_{ij} = \alpha_j$. Furthermore, $\{f_1, \ldots, f_n\}$ are a maximal independent set. Indeed, let f be any linear functional and let $\alpha_i = f(b_i)$. Consider the linear functional $f' = \sum_i \alpha_i f_i$. We have for every j, $f'(b_j) = (\sum \alpha_i f_i)(b_j) = \sum_i \alpha_i \delta_{ij} = \alpha_j = f(b_j)$. Since the two linear functions, f and f', agree on a basis, they are equal (by the same result in the assignments). □

Example 6.1.3. Consider the space $\mathbb{F}^n$ together with its standard basis $\mathrm{St} = \{e_1, \ldots, e_n\}$. Let f_i be the function

$$(x_1, \ldots, x_n) \overset{f_i}{\mapsto} x_i.$$

Then,

$$\mathrm{St}^* = \{f_1, \ldots, f_n\}.$$

To see that we simply need to verify that f_i is a linear function, which is clear, and that $f_i(e_j) = \delta_{ij}$, which is also clear.

Therefore, the form of the general element of $\mathbb{F}^{n*}$ is a function $\sum a_i f_i$ given by

$$(x_1, \ldots, x_n) \mapsto a_1 x_1 + \cdots + a_n x_n,$$

for some fixed $a_i \in \mathbb{F}$. We see that *we can identify* $\mathbb{F}^{n*}$ *with* $\mathbb{F}^n$*, where the vector* $(a_1, \ldots, a_n)$ *is identified with the linear functional* $(x_1, \ldots, x_n) \mapsto a_1x_1 + \cdots + a_nx_n$.

Example 6.1.4. Let $V = \mathbb{F}[t]_{n+1}$ be the space of polynomials of degree at most n. Consider the basis

$$B = \{1, t, t^2, \ldots, t^n\}.$$

The dual basis is

$$B^* = \{f_0, \ldots, f_n\},$$

where,

$$f_j(\sum_{i=0}^{n} \alpha_i t^i) = \alpha_j.$$

In general that's it, but if the field $\mathbb{F}$ contains the field of rational numbers we can say more. One checks that

$$f_j(f) = \frac{1}{j!}\frac{d^j f}{dt^j}(0).$$

(For $j = 0$, $\frac{d^j f}{dt^j}$ is interpreted as f.) Thus, elements of the dual space, which are just linear combinations of $\{f_0, \ldots, f_n\}$, can be viewed as linear differential operators.

Now, quite generally, if $B = \{v_1, \ldots, v_n\}$ is a basis for a vector space V and $B^* = \{f_1, \ldots, f_n\}$ is the dual basis, then any vector in V satisfies:

$$v = \sum_{i=1}^{n} f_i(v)v_i.$$

(This holds because $v = \sum_{i=1}^{n} a_i v_i$ for some a_i and now apply f_i to both sides to get $f_i(v) = a_i$.)

Applying these general considerations to our example above for real polynomials (say) we find that

$$f = \sum_{i=0}^{n} \frac{1}{i!}\frac{d^i f}{dt^i}(0) \cdot t^i,$$

which is none else than the Taylor expansion of f around 0!

6.2. **Duality.**

Proposition 6.2.1. *There is a natural isomorphism*

$$V \cong V^{**}.$$

Proof. We first define a map $V \to V^{**}$. Let $v \in V$. Define,

$$\phi_v : V^* \to \mathbb{F}$$

by

$$\phi_v(f) = f(v).$$

We claim that ϕ_v is a linear map $V^* \to \mathbb{F}$. Indeed,

$$\phi_v(f + g) = (f + g)(v) = f(v) + g(v) = \phi_v(f) + \phi_v(g),$$

and

$$\phi_v(\alpha g) = (\alpha g)(v) = \alpha g(v) = \alpha\phi_v(g).$$

We therefore get a map

$$V \to V^{**}, \qquad v \mapsto \phi_v.$$

This map is linear:

$$\phi_{v+w}(f) = f(v + w) = f(v) + f(w) = (\phi_v + \phi_w)(f),$$

and

$$\phi_{\alpha v}(f) = f(\alpha v) = \alpha\phi_v(f) = (\alpha\phi_v)(f).$$

Next, we claim that $v \mapsto \phi_v$ is injective. Since this is a linear map, we only need to show that its kernel is zero. Suppose that $\phi_v = 0$. Then, for every $f \in V^*$ we have $\phi_v(f) = f(v) = 0$. If $v \neq 0$ then let $v = v_1$ and complete it to a basis for V, say $B = \{v_1, \ldots, v_n\}$. Let $B^* = \{f_1, \ldots, f_n\}$ be the dual basis. Then $f_1(v_1) = 0$, which is a contradiction. Thus, $v = 0$.

We have found an injective linear map

$$V \to V^{**}, \qquad v \mapsto \phi_v.$$

Since $\dim(V) = \dim(V^*) = \dim(V^{**})$ the map $V \to V^{**}$ is an isomorphism. □

Remark 6.2.2. It is easy to verify that if B is a basis for V and B^* its dual basis, then B is the dual basis for B^* when we interpret V as V^{**}.

Definition 6.2.3. Let V be a finite dimensional vector space. Let $U \subseteq V$ be a subspace. Let

$$U^\perp := \{f \in V^* : f(u) = 0 \ \forall u \in U\}.$$

$U^\perp$ (read: U perp) is called the annihilator of U.

Lemma 6.2.4. *The following hold:*

(1) $U^\perp$ is a subspace.
(2) If $U \subseteq U_1$ then $U^\perp \supseteq U_1^\perp$.
(3) $U^\perp$ is a subspace of dimension $\dim(V) - \dim(U)$.
(4) We have $U^{\perp\perp} = U$.

Proof. It is easy to check that $U^\perp$ is a subspace. The second claim is obvious from the definitions.

Let $v_1, \ldots, v_a$ be a basis for U and complete to a basis B of V, $B = \{v_1, \ldots, v_n\}$. Let $B^* = \{f_1, \ldots, f_n\}$ be the dual basis. Suppose that $\sum_{i=1}^n \alpha_i f_i \in U^\perp$ then for every $j = 1, \ldots, a$ we have

$$0 = (\sum_{i=1}^n \alpha_i f_i)(v_j) = \alpha_j.$$

Thus, $U^\perp \subseteq \mathrm{Span}(f_{a+1}, \ldots, f_n)$. Conversely, it is easy to check that each $f_i, i = a+1, \ldots, n$, is in $U^\perp$ and so $U^\perp \supseteq \mathrm{Span}(f_{a+1}, \ldots, f_n)$. The third claim follows.

Note that this proof, applied now to $U^\perp$ gives that $U^{\perp\perp} = U$. □

Proposition 6.2.5. *Let U_1, U_2 be subspaces of V. Then*

$$(U_1 + U_2)^\perp = U_1^\perp \cap U_2^\perp, \qquad (U_1 \cap U_2)^\perp = U_1^\perp + U_2^\perp.$$

Proof. Let $f \in (U_1 + U_2)^\perp$. Since $U_i \subset U_1 + U_2$ we have $f \in U_i^\perp$ and so $f \in U_1^\perp \cap U_2^\perp$. Conversely, if $f \in U_1^\perp \cap U_2^\perp$ then for $v \in U_1 + U_2$, say $v = u_1 + u_2$, we have

$$f(v) = f(u_1 + u_2) = f(u_1) + f(u_2) = 0 + 0 = 0,$$

and we get the opposite inclusion.

The second claim follows formally. Note that $U_1 \cap U_2 = U_1^{\perp\perp} \cap U_2^{\perp\perp} = (U_1^\perp + U_2^\perp)^\perp$. Taking $\perp$ on both sides we get $(U_1 \cap U_2)^\perp = U_1^\perp + U_2^\perp$. □

Proposition 6.2.6. *Let U be a subspace of V then there is a natural isomorphism*

$$U^* \cong V^*/U^\perp.$$

Proof. Consider the map

$$S : V^* \to U^*, \qquad f \mapsto f|_U.$$

It is clearly a linear map. The kernel of S is by definition $U^\perp$. We therefore get a well defined injective linear map

$$S' : V^*/U^\perp \to U^*.$$

Now, $\dim(V^*/U^\perp) = \dim(V) - \dim(U^\perp) = \dim(U) = \dim(U^*)$. Thus, S' is an isomorphism. □

Corollary 6.2.7. *We have* $(V/U)^* \cong U^\perp$.

Proof. This follows formally from the above: think of V as $(V^*)^*$ and $U = (U^\perp)^\perp$. We already know that $(V^*)^*/(U^\perp)^\perp \cong (U^\perp)^*$. That is, $V/U \cong (U^\perp)^*$. Then, $(V/U)^* \cong (U^\perp)^{**} \cong U^\perp$.

Of course, one can also argue directly. Any element of $U^\perp$ is a linear functional $V \to \mathbb{F}$ that vanishes on U and so, by the first isomorphism theorem, induces a linear functional $V/U \to \mathbb{F}$. One shows that this provides a linear map $U^\perp \to (V/U)^*$. One can next show it's surjective and calculate the dimensions of both sides. □

Given a linear map

$$T : V \to W,$$

we get a function

$$T^* : W^* \to V^*, \qquad (T^*(f))(v) := f(Tv).$$

We leave the following lemma as an exercise:

Lemma 6.2.8. *(1) T^* is a linear map. It is called the **dual map** to T.*

(2) Let B, C be bases to V, W, respectively. Let $A = {}_C[T]_B$ be the $m \times n$ matrix representing T, where $n = \dim(V), m = \dim(W)$. Then the matrix representing T^* with respect to the bases B^*, C^* is the transpose of A:

$${}_{B^*}[T^*]_{C^*} = {}_C[T]_B{}^t.$$

(3) If T is injective then T^* is surjective.

(4) If T is surjective then T^* is injective.

Proposition 6.2.9. *Let $T : V \to W$ be a linear map with kernel U. Then* $\mathrm{Im}(T^*)$ *is $U^\perp$.*

Proof. Let $u_1, \ldots, u_a$ be a basis for U and $B = \{u_1, \ldots, u_a, u_{a+1}, \ldots, u_n\}$ an extension to a basis of V. Let $B^* = \{f_1, \ldots, f_n\}$ be the dual basis. We know that $U^\perp = \mathrm{Span}(\{f_{a+1}, \ldots, f_n\})$. We also know that $\{w_1, \ldots, w_{n-a}\}$, $w_i = T(u_{a+i})$, is a linearly independent set in W (cf. the proof of Theorem 3.2.1). Complete it to a basis $C = \{w_1, \ldots, w_m\}$ of W and let $C^* = \{g_1, \ldots, g_m\}$ be the dual basis. Let us calculate $T^*(g_i)$.

We have $T^*(g_i)(u_j) = g_i(T(u_j)) = 0$ if $j = 1, \ldots, a$ because $T(u_j)$ is then 0. We also have $T^*(g_i)(u_{a+j}) = g_i(T(u_{a+j})) = g_i(w_j) = \delta_{ij}$, for $j = 1, \ldots, n-a$. It follows that if $i > n-a$ then $T^*(g_i)$ is zero on every basis element of V and so must be the zero linear functional; it also follows that for $i \leq n-a$, $T^*(g_i)$ agrees with f_{a+i} on the basis B and so $T^*(g_i) = f_{a+i}, i = 1, \ldots, n-a$. We conclude that $\mathrm{Im}(T^*)$, being equal to $\mathrm{Span}(\{T^*(g_1), \ldots, T^*(g_m)\})$ is precisely $\mathrm{Span}(\{f_{a+1}, \ldots, f_n\}) = U^\perp$. □

6.3. **An application.** We provide another proof of Theorem 5.3.1.

Theorem 6.3.1. *Let A be an $m \times n$ matrix. Then* $\mathrm{rk}_r(A) = \mathrm{rk}_c(A)$.

Proof. Let T be the linear map associated to A, then A^t is the linear map associated to T^*. Let $U = \mathrm{Ker}(T)$. We have $\mathrm{rk}_c(A) = \dim(\mathrm{Im}(T)) = n - \dim(U)$. We also have $\mathrm{rk}_r(A) = \mathrm{rk}_c(A^t) = \mathrm{rk}_c(T^*) = \dim(\mathrm{Im}(T^*)) = \dim(U^\perp) = n - \dim(U)$. □

7. Inner product spaces

In contrast to the previous sections, the field $\mathbb{F}$ over which the vector spaces in this section are defined is very special: we always assume $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. We shall denote complex conjugation by $r \mapsto \bar{r}$. We shall use this notation even if $\mathbb{F} = \mathbb{R}$, where complex conjugation is trivial, simply to have uniform notation.

7.1. Definition and first examples of inner products.

Definition 7.1.1. An **inner product** on a vector space V over $\mathbb{F}$ is a function:

$$\langle \cdot, \cdot \rangle : V \times V \longrightarrow \mathbb{F},$$

satisfying the following:

(1) $\langle v_1 + v_2, w \rangle = \langle v_1, w \rangle + \langle v_2, w \rangle$ for $v_1, v_2, w \in V$;
(2) $\langle \alpha v, w \rangle = \alpha \cdot \langle v, w \rangle$ for $\alpha \in \mathbb{F}$, $v, w \in V$.
(3) $\langle v, w \rangle = \overline{\langle w, v \rangle}$ for $v, w \in V$.
(4) $\langle v, v \rangle \geq 0$ with equality if and only if $v = 0$.

Remark 7.1.2. First note that $\langle v, v \rangle = \overline{\langle v, v \rangle}$ by axiom (3), so $\langle v, v \rangle \in \mathbb{R}$ and axiom (4) makes sense! We also remark that it follows easily from the axioms that:

- $\langle w, v_1 + v_2 \rangle = \langle w, v_1 \rangle + \langle w, v_2 \rangle$;
- $\langle v, \alpha w \rangle = \overline{\alpha} \cdot \langle v, w \rangle$.
- $\langle v, 0 \rangle = \langle 0, v \rangle = 0$.

Definition 7.1.3. We define the **norm** of $v \in V$ by

$$\|v\| = \langle v, v \rangle^{1/2},$$

and the distance between v and w by

$$d(v, w) := \|v - w\|.$$

Example 7.1.4. The most basic example is $\mathbb{F}^n$ with the inner product:

$$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle = \sum_{i=1}^{n} x_i \overline{y_i}.$$

Theorem 7.1.5 (Cauchy-Schwartz inequality)**.** *Let V be an inner product space. For every $u, v \in V$ we have*

$$|\langle u, v \rangle| \leq \|u\| \cdot \|v\|.$$

Proof. If $\|v\| = 0$ then $v = 0$ and the inequality holds trivially. Else, let $\alpha = \frac{\langle u, v \rangle}{\|v\|^2}$. We have:

$$\begin{aligned} 0 &\leq \|u - \alpha v\|^2 \\ &= \langle u - \alpha v, u - \alpha v \rangle \\ &= \|u\|^2 + \alpha\overline{\alpha}\|v\|^2 - \alpha\overline{\langle u, v \rangle} - \overline{\alpha}\langle u, v \rangle \\ &= \|u\|^2 - \frac{|\langle u, v \rangle|^2}{\|v\|^2}. \end{aligned}$$

The theorem follows by rearranging and taking square roots. □

Proposition 7.1.6. *The norm function is indeed a **norm**. That is, the function $v \mapsto \|v\|$ satisfies:*

(1) $\|v\| \geq 0$ with equality if and only if $v = 0$;
(2) $\|\alpha v\| = |\alpha| \cdot \|v\|$;
*(3) (**Triangle inequality**) $\|u + v\| \leq \|u\| + \|v\|$.*

*The distance function is indeed a **distance function**. Namely, it satisfies:*

(1) $d(v,w) \geq 0$ with equality if and only if $v = w$;
(2) $d(v,w) = d(w,v)$;
*(3) (**Triangle inequality**) $d(v,w) \leq d(v,u) + d(u,w)$.*

Proof. The first axiom of a norm holds because $\langle v, v\rangle \geq 0$ with equality if and only if $v = 0$. The second is just

$$\|\alpha v\| = \sqrt{\langle \alpha v, \alpha v\rangle} = \sqrt{\alpha\overline{\alpha} \cdot \langle v, v\rangle} = |\alpha| \cdot \|v\|.$$

The third axiom is less trivial. We have:

$$\begin{aligned}
\|u+v\|^2 &= \langle u+v, u+v\rangle \\
&= \|u\|^2 + \|v\|^2 + \langle u, v\rangle + \langle v, u\rangle \\
&= \|u\|^2 + \|v\|^2 + 2\Re\langle u, v\rangle \\
&\leq \|u\|^2 + \|v\|^2 + 2|\langle u, v\rangle| \\
&\leq \|u\|^2 + \|v\|^2 + 2\|u\| \cdot \|v\| \\
&= (\|u\| + \|v\|)^2.
\end{aligned}$$

(In these inequalities we first used that for a complex number z the real part of z, $\Re z$, is less or equal to $|z|$, and then we used the Cauchy-Schwartz inequality.)

The axioms for the distance function follow immediately from those for the norm function and we leave the verification to you. □

Example 7.1.7. (**Parallelogram law**). We have

$$\|u+v\|^2 + \|u-v\|^2 = 2\|u\|^2 + 2\|v\|^2.$$

This is easy to check from the definitions.

Suppose now, for simplicity, that V is a vector space over $\mathbb{R}$. Note that we also have

$$\langle u, v\rangle = \frac{1}{2}\left(\|u+v\|^2 - \|u\|^2 - \|v\|^2\right).$$

Suppose that we are given any continuous norm function $\|\cdot\| : V \to \mathbb{R}$. Namely, a continuous function satisfying the axioms of a norm function but not necessarily arising from an inner product. One can prove that

$$\langle u, v\rangle = \frac{1}{2}\left(\|u+v\|^2 - \|u\|^2 - \|v\|^2\right)$$

defines an inner product if and only if the parallelogram law holds.

Example 7.1.8. Let $M \in M_n(\mathbb{F})$. Define

$$M^* = \overline{M}^t.$$

That is, if $M = (m_{ij})$ then $M^* = (\overline{m_{ji}})$. A matrix $M \in M_n(\mathbb{F})$ is called **Hermitian** if

$$M = M^*.$$

Note that if $\mathbb{F} = \mathbb{R}$ then $M^* = M^t$ and a Hermitian matrix is simply a symmetric matrix.

Now let $M \in M_n(\mathbb{F})$ be a Hermitian matrix such that for every vector $(x_1, \ldots, x_n) \neq 0$ we have

$$(x_1, \ldots, x_n) M \begin{pmatrix} \overline{x_1} \\ \vdots \\ \overline{x_n} \end{pmatrix} > 0,$$

(one calls M **positive definite** in that case) and define

$$\langle (x_1,\dots,x_n),(y_1,\dots,y_n)\rangle = (x_1,\dots,x_n)\ M \begin{pmatrix} \overline{y_1} \\ \vdots \\ \overline{y_n} \end{pmatrix} = \sum_{i,j} m_{ij} x_i \overline{y_j}.$$

This is an inner product. The case of $M = I_n$ gives back our first example $\langle (x_1,\dots,x_n),(y_1,\dots,y_n)\rangle = \sum_{i=1}^n x_i\overline{y_i}$. It is not hard to prove that any inner product on $\mathbb{F}^n$ arises this way from a positive definite Hermitian matrix (Exercise).

Deciding whether $M = M^*$ is trivial. Deciding whether M is positive definite is much harder, though there are good criterions for that. For 2×2 matrices M, we have that M is Hermitian if and only if

$$M = \begin{pmatrix} a & b \\ \bar{b} & d \end{pmatrix}.$$

Such M is positive definite if and only if a and d are positive real numbers and $ad - b\bar{b} > 0$ (Exercise).

For such an inner product on $\mathbb{F}^n$, the Cauchy-Schwartz inequality says the following:

$$\left|\sum_{i,j} m_{ij} x_i \overline{y_j}\right| \leq \sqrt{\sum_{i,j} m_{ij} x_i \overline{x_j}} \cdot \sqrt{\sum_{i,j} m_{ij} y_i \overline{y_j}}.$$

In the simplest case, of $\mathbb{R}^n$ and $M = I_n$, we get a well known inequality:

$$\left|\sum_{i,j} x_i y_j\right| \leq \sqrt{\sum_i x_i^2} \cdot \sqrt{\sum_i y_i^2}.$$

Example 7.1.9. Let V be the space of continuous real functions $f : [a,b] \to \mathbb{R}$. Define an inner product by

$$\langle f, g\rangle = \int_a^b f(x)g(x)dx.$$

The fact that this is an inner product uses some standard results in analysis (including the fact that the integral of a non-zero non-negative continuous function is positive). The Cauchy-Schwartz inequality now says:

$$\left|\int_a^b f(x)g(x)dx\right| \leq \left(\int_a^b f(x)^2 dx\right)^{1/2} \cdot \left(\int_a^b g(x)^2 dx\right)^{1/2}.$$

7.2. **Orthogonality and the Gram-Schmidt process.** Let $V/\mathbb{F}$ be an inner product space.

Definition 7.2.1. We say that $u, v \in V$ are **orthogonal** if

$$\langle u, v\rangle = 0.$$

We use the notation $u \perp v$. We also say u is **perpendicular** to v.

Example 7.2.2. Let $V = \mathbb{F}^n$ with the standard inner product. Then $e_i \perp e_j$. However, if we take $n = 2$, say, and the inner product defined by the matrix $\begin{pmatrix} 1 & 1+i \\ 1-i & 5 \end{pmatrix}$ then e_1 is not perpendicular to v. Indeed,

$$(1,0)\begin{pmatrix} 1 & 1+i \\ 1-i & 5 \end{pmatrix}\begin{pmatrix} 0 \\ \bar{1} \end{pmatrix} = 1+i \neq 0.$$

So, as you may have suspected, orthogonality is not an absolute notion, it depends on the inner product.

Definition 7.2.3. Let V be a finite dimensional inner product space. A basis $\{v_1, \ldots, v_n\}$ for V is called **orthonormal** if:

(1) For $i \neq j$ we have $v_i \perp v_j$;
(2) $\|v_i\| = 1$ for all i.

Theorem 7.2.4 (The Gram-Schmidt process). *Let $\{s_1, \ldots, s_n\}$ be any basis for V. There is an orthonormal basis $\{v_1, \ldots, v_n\}$ for V, such that for every i,*

$$\mathrm{Span}(\{v_1, \ldots, v_i\}) = \mathrm{Span}(\{s_1, \ldots, s_i\}).$$

Proof. We construct $v_1, \ldots, v_n$ inductively on i, such that $\mathrm{Span}(\{v_1, \ldots, v_i\}) = \mathrm{Span}(\{s_1, \ldots, s_i\})$. Note that this implies that $\dim \mathrm{Span}(\{v_1, \ldots, v_i\}) = i$ and so that $\{v_1, \ldots, v_i\}$ are linearly independent. In particular, $\{v_1, \ldots, v_n\}$ is a basis.

We let

$$v_1 = \frac{s_1}{\|s_1\|}.$$

Then $\|v_1\| = 1$ and $\mathrm{Span}(\{v_1\}) = \mathrm{Span}(\{s_1\})$.

Assume we have defined already $v_1, \ldots, v_k$ such that for all $i \leq k$ we have $\mathrm{Span}(\{v_1, \ldots, v_i\}) = \mathrm{Span}(\{s_1, \ldots, s_i\})$. Let

$$s'_{k+1} = s_{k+1} - \sum_{i=1}^{k} \langle s_{k+1}, v_i \rangle \cdot v_i, \qquad v_{k+1} = \frac{s'_{k+1}}{\|s'_{k+1}\|}.$$

First, note that s'_{k+1} cannot be zero since $\{s_1, \ldots, s_{k+1}\}$ are independent and $\mathrm{Span}(\{v_1, \ldots, v_k\}) = \mathrm{Span}(\{s_1, \ldots, s_k\})$. Thus, v_{k+1} is well defined and $\|v_{k+1}\| = 1$. It is also clear from the definitions and induction that $\mathrm{Span}(\{v_1, \ldots, v_{k+1}\}) = \mathrm{Span}(\{s_1, \ldots, s_k, v_{k+1}\}) = \mathrm{Span}(\{s_1, \ldots, s_k, s'_{k+1}\}) = \mathrm{Span}(\{s_1, \ldots, s_k, s_{k+1}, s'_{k+1}\}) = \mathrm{Span}(\{s_1, \ldots, s_k, s_{k+1}\})$. Finally, for $j \leq k$,

$$\begin{aligned}
\langle v_{k+1}, v_j \rangle &= \frac{\langle s'_{k+1}, v_j \rangle}{\|s'_{k+1}\|} \\
&= \frac{1}{\|s'_{k+1}\|} \cdot \langle s_{k+1} - \sum_{i=1}^{k} \langle s_{k+1}, v_i \rangle \cdot v_i, v_j \rangle \\
&= \frac{1}{\|s'_{k+1}\|} \cdot \left(\langle s_{k+1}, v_j \rangle - \sum_{i=1}^{k} \langle s_{k+1}, v_i \rangle \cdot \langle v_i, v_j \rangle \right) \\
&= \frac{1}{\|s'_{k+1}\|} \cdot \left(\langle s_{k+1}, v_j \rangle - \sum_{i=1}^{k} \langle s_{k+1}, v_i \rangle \cdot \delta_{ij} \right) \\
&= 0.
\end{aligned}$$

Thus, $\{v_1, \ldots, v_{k+1}\}$ is an orthonormal set. □

Here are some reasons an orthonormal basis is useful. Let V be an inner product space and $B = \{v_1, \ldots, v_n\}$ an orthonormal basis for V. Let $v, w \in V$ and say $[v]_B = (\alpha_1, \ldots, \alpha_n), [w]_B =$

$(\beta_1, \ldots, \beta_n)$. Then

$$\begin{aligned}\langle u, v\rangle &= \langle \sum \alpha_i v_i, \sum \beta_i v_i\rangle \\ &= \sum_{i,j=1}^{n} \alpha_i \overline{\beta_j} \langle v_i, v_j\rangle \\ &= \sum_{i,j=1}^{n} \alpha_i \overline{\beta_j} \delta_{ij} \\ &= \sum_{i=1}^{n} \alpha_i \overline{\beta_i}.\end{aligned}$$

That is, switching to the coordinate system supplied by the orthonormal basis, the inner product looks like the standard one and, in particular, the formulas are much easier to write down and work with.

A further property is the following: Let $v \in V$ and write $v = \sum_{i=1}^{n} \alpha_i v_i$. Then $\langle v, v_j\rangle = \sum_{i=1}^{n} \alpha_i \langle v_i, v_j\rangle = \alpha_j$. That is, in any orthonormal basis $\{v_1, \ldots, v_n\}$ we have

$$v = \sum_{i=1}^{n} \langle v, v_i\rangle \cdot v_i. \tag{6}$$

Example 7.2.5. Let $W \subseteq \mathbb{R}^3$ be the plane given by

$$W = \mathrm{Span}(\{(1,0,1), (0,1,1)\}).$$

Let us find an orthonormal basis for W and a line perpendicular to it.

One way to do it is the following. Complete $\{(1,0,1), (0,1,1)\}$ to a basis of $\mathbb{R}^3$. For example, $\{s_1, s_2, s_3\} = \{(1,0,1), (0,1,1), (0,0,1)\}$. Note that

$$\det \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} = 1,$$

and so this is indeed a basis. We now perform the Gram-Schmidt process on this basis.

We have $v_1 = \frac{1}{\sqrt{2}} \cdot (1,0,1)$. Then

$$\begin{aligned} s_2' &= s_2 - \langle s_2, v_1\rangle \cdot v_1 \\ &= (0,1,1) - \frac{1}{2}(1,0,1) \\ &= (-1/2, 1, 1/2),\end{aligned}$$

and

$$v_2 = \frac{1}{\sqrt{6}}(-1,2,1).$$

Therefore,

$$\{v_1, v_2\} = \left\{ \frac{1}{\sqrt{2}} \cdot (1,0,1), \frac{1}{\sqrt{6}}(-1,2,1) \right\}$$

is an orthonormal basis for W. Next,

$$\begin{aligned} s_3' &= s_3 - \langle s_3, v_1\rangle v_1 - \langle s_3, v_2\rangle v_2 \\ &= (0,0,1) - \langle (0,0,1), \frac{1}{\sqrt{2}} \cdot (1,0,1)\rangle \cdot \frac{1}{\sqrt{2}}(1,0,1) - \langle (0,0,1), \frac{1}{\sqrt{6}}(-1,2,1)\rangle \cdot \frac{1}{\sqrt{6}}(-1,2,1) \\ &= (0,0,1) - \frac{1}{2}(1,0,1) - \frac{1}{6}(-1,2,1) \\ &= (-1/3, -1/3, 1/3).\end{aligned}$$

If we want an orthonormal basis for $\mathbb{R}^3$ we can take $v_3 = \frac{1}{\sqrt{3}}(-1/3, -1/3, 1/3)$ but, to find a line orthogonal to W, we can just take the line through s_3' which is $\text{Span}((-1,-1,1))$.

Definition 7.2.6. Let $S \subset V$ be a subset and let

$$S^\perp = \{v \in V : \langle s, v\rangle = 0, \forall s \in S\}.$$

It is easy to see that $S^\perp$ is a subspace and in fact, if we let $U = \text{Span}(S)$ then

$$S^\perp = U^\perp.$$

Proposition 7.2.7. *Let U be a subspace of V then*

$$V = U \oplus U^\perp,$$

and

$$U^{\perp\perp} = U.$$

*The subspace $U^\perp$ is called the **orthogonal complement** of U in V.*

Proof. Find a basis $\{s_1, \dots, s_a\}$ to U and complete it to any basis of V, say $\{s_1, \dots, s_n\}$. Apply the Gram-Schmidt process and obtain an orthonormal basis $\{v_1, \dots, v_n\}$ then

$$U := \text{Span}(\{v_1, \dots, v_a\}).$$

We claim that $U^\perp = \text{Span}(\{v_{a+1}, \dots, v_n\})$. Indeed, let $v = \sum_{i=1}^n \alpha_i v_i$. Then $v \in U^\perp$ if and only if $\langle v, v_i\rangle = 0$ for $1 \le i \le a$. But, $\langle v, v_i\rangle = \alpha_i$ so $v \in U^\perp$ if and only if $v = \sum_{i=a+1}^n \alpha_i v_i$, which is equivalent to $v \in \text{Span}(\{v_{a+1}, \dots, v_n\})$.

It is clear then that $V = U \oplus U^\perp$, and $U^{\perp\perp} = U$. □

Let us consider linear equations again: Suppose that we have m linear equations over $\mathbb{R}$ in n variables:

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= 0 \end{aligned}$$

If we let $S = \{(a_{11}, \dots, a_{1n}), \dots, (a_{m1}, \dots, a_{mn})\}$ then the space of solutions is precisely $U^\perp$. Conversely, given a subspace $W \subseteq \mathbb{R}^n$ to find a set of equations defining W is to find $W^\perp$. The Gram-Schmidt process gives a way to do that: Find a basis $\{s_1, \dots, s_a\}$ to W and complete it to any basis of V, say $\{s_1, \dots, s_n\}$. Apply the Gram-Schmidt process and obtain an orthonormal basis $\{v_1, \dots, v_n\}$ then, as we have seen,

$$W^\perp := \text{Span}(\{v_{a+1}, \dots, v_n\}).$$

7.3. **Applications.**

7.3.1. *Orthogonal projections.* Let V be an inner product space of finite dimension and $U \subseteq V$ a subspace. Then $V = U \oplus U^\perp$ and we let

$$T : V \to U,$$

be the projection on U along $U^\perp$. We also call T the **orthogonal projection** on U.

Theorem 7.3.1. *Let $\{v_1, \dots, v_r\}$ be an orthonormal basis for U. Then:*

(1) $T(v) = \sum_{i=1}^r \langle v, v_i\rangle \cdot v_i$;
(2) $(v - T(v)) \perp T(v)$;
(3) $T(v)$ is the vector in U that is closest to v.

Proof. Clearly the function

$$v \mapsto T'(v) := \sum_{i=1}^{r} \langle v, v_i \rangle \cdot v_i$$

is a linear map from V into U. If $v \in U^\perp$ then $T'(v) = 0$, while if $v \in U$ we have $v = \sum_{i=1}^{r} \alpha_i v_i$ and as we have noted before (see 6) $\alpha_i = \langle v, v_i \rangle$. That is, if $v \in U$ then $T'(v) = v$. It follows, see Theorem 3.7.2, that T' is the projection on U along $U^\perp$ and so $T' = T$.

We have seen that if T is the projection on a subspace U along W then $v - T(v) \in W, T(v) \in U$; apply that to $W = U^\perp$ to get $v - T(v) \in U^\perp$ and, in particular, $(v - T(v)) \perp T(v)$.

We now come to the last part. We wish to show that

$$\|v - \sum_{i=1}^{r} \alpha_i v_i\|$$

is minimal (equivalently, $\|v - \sum_{i=1}^{r} \alpha_i v_i\|^2$ is minimal) when $\alpha_i = \langle v, v_i \rangle$ for all $i = 1, \ldots, r$.

Complete $\{v_1, \ldots, v_r\}$ to an orthonormal basis of V, say $\{v_1, \ldots, v_n\}$ (this is possible because we first complete to any basis and then apply Gram-Schmidt, which will not change $\{v_1, \ldots, v_r\}$ as is easy to check). Then

$$v = \sum_{i=1}^{n} \beta_i v_i, \qquad \beta_i = \langle v, v_i \rangle.$$

Then,

$$\begin{aligned} \|v - \sum_{i=1}^{r} \alpha_i v_i\|^2 &= \|(\beta_1 - \alpha_1)v_1 + \cdots + (\beta_r - \alpha_r)v_r + \beta_{r+1}v_{r+1} + \cdots + \beta_n v_n\|^2 \\ &= \sum_{i=1}^{r} |\beta_i - \alpha_i|^2 + \sum_{i=r+1}^{n} |\beta_i|^2. \end{aligned}$$

Clearly this is minimized when $\alpha_i = \beta_i$ for $i = 1, \ldots, r$. That is, when $\alpha_i = \langle v, v_i \rangle$. □

Remark 7.3.2 (Gram-Schmidt revisited). Recall the process. We have an initial basis $\{s_1, \ldots, s_n\}$, which we wish to transform into an orthonormal basis $\{v_1, \ldots, v_n\}$. Suppose we have already constructed $\{v_1, \ldots, v_k\}$. They form an orthonormal basis for $U = \text{Span}(\{v_1, \ldots, v_k\})$. The next step in the process is to construct:

$$s'_{k+1} = s_{k+1} - \sum_{i=1}^{k} \langle s_{k+1}, v_i \rangle \cdot v_i.$$

We now recognize $\sum_{i=1}^{k} \langle s_{k+1}, v_i \rangle \cdot v_i$ as the orthogonal projection of s_{k+1} on U (by part (1) of the Theorem). s_{k+1} is then decomposed into its orthogonal projection on U and s'_{k+1} which lies in $U^\perp$ (by part (2) of the Theorem). It only remains to normalize it and we indeed have let $v_{k+1} = \frac{s'_{k+1}}{\|s'_{k+1}\|}$.

7.3.2. *Least squares approximation.* (In assignments).

8. Eigenvalues, eigenvectors and diagonalization

We come now to a subject which has many important applications. The notions we shall discuss in this section will allow us: (i) to provide a criterion for a matrix to be positive definite and that is relevant to the study of inner products and extrema of functions of several variables; (ii) to compute efficiently high powers of a matrix and that is relevant to study of recurrence sequences and Markov processes, and many other applications; (iii) to give structure theorems for linear transformations.

8.1. Eigenvalues, eigenspaces and the characteristic polynomial.

Let V be a vector space over a field $\mathbb{F}$.

Definition 8.1.1. Let $T : V \to V$ be a linear map. A scalar $\lambda \in \mathbb{F}$ is called an **eigenvalue** of T if there is a non-zero vector $v \in V$ such that

$$T(v) = \lambda v.$$

Any vector v like that is called an **eigenvector** of T. The definition applies for $n \times n$ matrices, viewed as linear maps $\mathbb{F}^n \to \mathbb{F}^n$.

Remark 8.1.2. λ is an eigenvalue of T if and only if λ is an eigenvalue of the matrix ${}_B[T]_B$, with respect to one (any) basis B. Indeed, we have

$$Tv = \lambda v \Leftrightarrow {}_B[T]_B[v]_B = \lambda[v]_B.$$

Note that if we think about a matrix A as a linear transformation then this remark show that λ is an eigenvalue of A if and only if λ is an eigenvalue of $M^{-1}AM$ for one (any) invertible matrix M. This is no mystery... you can check that $M^{-1}v$ is the corresponding eigenvector.

Example 8.1.3. $\lambda = 1, 2$ are eigenvalues of the matrix $A = \begin{pmatrix} -1 & 6 \\ -1 & 4 \end{pmatrix}$. Indeed,

$$\begin{pmatrix} -1 & 6 \\ -1 & 4 \end{pmatrix} \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \qquad \begin{pmatrix} -1 & 6 \\ -1 & 4 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 2 \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Definition 8.1.4. Let V be a finite dimensional vector space over $\mathbb{F}$ and $T : V \to V$ a linear map. The **characteristic polynomial** Δ_T of T is defined as follows: Let B be a basis for V and $A = {}_B[T]_B$ the matrix representing T in the basis B. Let

$$\Delta_T = \det(t \cdot I_n - A),$$

where t is a free variable and $n = \dim(A)$.

Example 8.1.5. Consider $T = A = \begin{pmatrix} -1 & 6 \\ -1 & 4 \end{pmatrix}$. Then

$$\Delta_T = \Delta_A = \det\left(\begin{pmatrix} t & 0 \\ 0 & t \end{pmatrix} - \begin{pmatrix} -1 & 6 \\ -1 & 4 \end{pmatrix}\right) = \det\begin{pmatrix} t+1 & -6 \\ 1 & t-4 \end{pmatrix} = t^2 - 3t + 2.$$

With respect to the basis $B = \left\{ \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}$, T is diagonal.

$${}_B[T]_B = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix},$$

and

$$\Delta_T = \det \begin{pmatrix} t-1 & 0 \\ 0 & t-2 \end{pmatrix} = (t-1)(t-2) = t^2 - 3t + 2.$$

Proposition 8.1.6. *The polynomial Δ_T has the following properties:*

(1) Δ_T is independent of the choice of basis used to compute it. In particular, if A is a matrix and M an invertible matrix then $\Delta_A = \Delta_{M^{-1}AM}$.

(2) Suppose that $\dim(V) = n$ and $A = {}_B[T]_B = (a_{ij})$. Let

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii}.$$

Then,

$$\Delta_T = t^n - \mathrm{Tr}(A)t^{n-1} + \cdots + (-1)^n \det(A).$$

In particular, $\mathrm{Tr}(A)$ and $\det(A)$ do not depend on the basis B and we let $\mathrm{Tr}(T) = \mathrm{Tr}(A), \det(T) = \det(A)$.

Proof. Let B, C be two bases for V. Let $A = {}_B[T]_B, D = {}_C[T]_C, M = {}_C M_B$. Then,

$$\begin{aligned} \det(t \cdot I_n - A) &= \det(t \cdot I_n - M^{-1}DM) \\ &= \det(M^{-1}(t \cdot I_n - D)M) \\ &= \det(M^{-1}) \det(t \cdot I_n - D) \det(M) \\ &= \det(t \cdot I_n - D). \end{aligned}$$

This proves the first assertion.

Put $A = (a_{ij})$ and let us calculate Δ_T. We have

$$\Delta_T = \det(t \cdot I_n - A) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{n\sigma(n)},$$

where $(b_{ij}) = t \cdot I_n - A$. Each b_{ij} contains at most a single power of t and so clearly Δ_T is a polynomial of degree at most n. The monomial t^n arises only from the summand $b_{11}b_{22}\cdots b_{nn} = (t-a_{11})(t-a_{22})\cdots(t-a_{nn})$ and so appears with coefficient 1 in Δ_T. Also the monomial t^{n-1} comes only from this summand, because if there is an i such that $\sigma(i) \neq i$ then there is another index j such that $\sigma(j) \neq j$ and then in $b_{1\sigma(1)}b_{2\sigma(2)}\cdots b_{n\sigma(n)}$ the power of t is at most $n-2$. We see therefore that the coefficient of t^{n-1} comes from expanding $(t-a_{11})(t-a_{22})\cdots(t-a_{nn})$ and is $-a_{11} - a_{22} - \cdots - a_{nn} = -\mathrm{Tr}(A)$.

Finally, the constant coefficient is $\Delta_T(0) = (\det(t \cdot I_n - A))(0) = \det(-A) = (-1)^n \det(A)$. □

Example 8.1.7. We have

$$\Delta_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} = t^2 - (a+d)t + (ad - bc).$$

Example 8.1.8. For the matrix

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \\ \vdots & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$$

we have characteristic polynomial

$$\prod_{i=1}^{n} (t - \lambda_i).$$

Theorem 8.1.9. *The following are equivalent:*

(1) λ is an eigenvalue of A;
(2) The linear map $\lambda I - A$ is singular (i.e., has a kernel, not invertible);
(3) $\Delta_A(\lambda) = 0$, where Δ_A is the characteristic polynomial of A.

Proof. Indeed, λ is an eigenvalue of A if and only if there's a vector $v \neq 0$ such that $Av = \lambda v$. That is, if and only if there's a vector $v \neq 0$ such that $(\lambda I - A)v = 0$, which is equivalent to $\lambda I - A$ being singular. Thus, (1) and (2) are equivalent.

Now, a square matrix B is singular if and only if B is not invertible, if and only if $\det(B) = 0$. Therefore, $\lambda I - A$ is singular if and only if $\det(\lambda I - A) = 0$, if and only if $[\det(tI - A)](\lambda) = 0$. Thus, (2) is equivalent to (3).

□

Corollary 8.1.10. *Let A be an $n \times n$ matrix then A has at most n distinct eigenvalues, i.e., the roots of its characteristic polynomial.*

Definition 8.1.11. Let $T : V \to V$ be a linear map, V a vector space of dimension n. Let

$$E_\lambda = \{v \in V : Tv = \lambda v\}.$$

We call E_λ the **eigenspace** of λ. The definition applies to matrices (thought of as linear transformations). Namely, let A be an $n \times n$ matrix then

$$E_\lambda = \{v \in \mathbb{F}^n : Av = \lambda v\}.$$

If A is the matrix representing T with respect to some basis the definitions agree.

Example 8.1.12. $A = \begin{pmatrix} -1 & 6 \\ -1 & 4 \end{pmatrix}$, $\Delta_A(t) = (t-1)(t-2)$. The eigenvalues are $1, 2$. The eigenspaces are

$$E_1 = \operatorname{Ker}\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} -1 & 6 \\ -1 & 4 \end{pmatrix}\right) = \operatorname{Ker}\left(\begin{pmatrix} 2 & -6 \\ 1 & -3 \end{pmatrix}\right) = \operatorname{Span}\left\{\begin{pmatrix} 3 \\ 1 \end{pmatrix}\right\},$$

and

$$E_2 = \operatorname{Ker}\left(\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} - \begin{pmatrix} -1 & 6 \\ -1 & 4 \end{pmatrix}\right) = \operatorname{Ker}\left(\begin{pmatrix} 3 & -6 \\ 1 & -2 \end{pmatrix}\right) = \operatorname{Span}\left\{\begin{pmatrix} 2 \\ 1 \end{pmatrix}\right\}.$$

Example 8.1.13. $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, $\Delta_A(t) = t^2 - t - 1$. The eigenvalues are

$$\lambda_1 = \frac{1+\sqrt{5}}{2}, \qquad \lambda_2 = \frac{1-\sqrt{5}}{2}.$$

The eigenspaces are

$$E_{\lambda_1} = \operatorname{Ker}\left(\begin{pmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1+\sqrt{5}}{2} \end{pmatrix} - \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}\right) = \operatorname{Ker}\left(\begin{pmatrix} \frac{1+\sqrt{5}}{2} & -1 \\ -1 & -\frac{1-\sqrt{5}}{2} \end{pmatrix}\right) = \operatorname{Span}\left\{\begin{pmatrix} 1 \\ \frac{1+\sqrt{5}}{2} \end{pmatrix}\right\},$$

and

$$E_{\lambda_2} = \operatorname{Ker}\left(\begin{pmatrix} \frac{1-\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{pmatrix} - \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}\right) = \operatorname{Ker}\left(\begin{pmatrix} \frac{1-\sqrt{5}}{2} & -1 \\ -1 & -\frac{1+\sqrt{5}}{2} \end{pmatrix}\right) = \operatorname{Span}\left\{\begin{pmatrix} 1 \\ \frac{1-\sqrt{5}}{2} \end{pmatrix}\right\}.$$

Definition 8.1.14. Let λ be an eigenvalue of a linear map T. Let

$$m_g(\lambda) = \dim(E_\lambda);$$

$m_g(\lambda)$ is called the **geometric multiplicity** of λ. Let us also write, using unique factorization,

$$\Delta_T(t) = (t-\lambda)^{m_a(\lambda)} g(t), \qquad g(\lambda) \neq 0;$$

$m_a(\lambda)$ is called the **algebraic multiplicity** of λ.

Proposition 8.1.15. *Let λ be an eigenvalue of $T : V \to V$, $\dim(V) = n$. The following inequalities hold:*

$$1 \leq m_g(\lambda) \leq m_a(\lambda) \leq n.$$

Proof. Since λ is an eigenvalue we have $\dim(E_\lambda) > 0$ and so we get the first inequality. The inequality $m_a(\lambda) \leq n$ is clear since $\deg(\Delta_T(t)) = \dim(V) = n$. Thus, it only remains to prove that $m_g(\lambda) \leq m_a(\lambda)$.

Choose a basis $\{v_1, \ldots, v_m\}$ ($m = m_g(\lambda)$) to E_λ and complete it to a basis $\{v_1, \ldots, v_n\}$ of V. With respect to this basis T is represented by a matrix of the form

$$[T] = \begin{pmatrix} \lambda I_m & B \\ 0 & C \end{pmatrix},$$

where B is an $m \times (n-m)$ matrix, $C = (n-m) \times (n-m)$ matrix and 0 here stands for the $(n-m) \times m$ matrix of zeros. Therefore,

$$\begin{aligned} \Delta_T(t) &= \det \begin{pmatrix} (t-\lambda) I_m & -B \\ 0 & tI_{n-m} - C \end{pmatrix} \\ &= \det((t-\lambda) I_m) \cdot \det(tI_{n-m} - C) \\ &= (t-\lambda)^m \cdot \det(tI_{n-m} - C). \end{aligned} \tag{7}$$

This shows, $m = m_g(\lambda) \leq m_a(\lambda)$. □

Example 8.1.16. Let A be the matrix $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. We have $\Delta_A(t) = (t-1)^2$. Thus, $m_a(1) = 2$. On the other hand $m_g(1) = 1$. To see that, by pure thought, note that $1 \leq m_g(1) \leq 2$. However, if $m_g(1) = 2$ then $E_1 = \mathbb{F}^2$ and so $Av = v$ for every $v \in \mathbb{F}^2$. This implies that $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and that's a contradiction.

8.2. **Diagonalization.** Let V be a finite dimensional vector space over a field $\mathbb{F}$, $\dim(V) = n$. We denote a diagonal matrix with entries $\lambda_1, \ldots, \lambda_n$ by

$$\operatorname{diag}(\lambda_1, \ldots, \lambda_n).$$

Definition 8.2.1. A linear map T (resp., a matrix A) is called **diagonalizable** if there is a basis B (resp., an invertible matrix M) such that

$${}_B[T]_B = \operatorname{diag}(\lambda_1, \ldots, \lambda_n),$$

with $\lambda_i \in \mathbb{F}$, not necessarily distinct (resp.

$$M^{-1} A M = \operatorname{diag}(\lambda_1, \ldots, \lambda_n).)$$

Remark 8.2.2. Note that in this case the characteristic polynomial is $\prod_{i=1}^n (t - \lambda_i)$ and so the λ_i are the eigenvalues.

Lemma 8.2.3. *T is diagonalizable if and only if there is a basis of V consisting of eigenvectors of V.*

Proof. If T is diagonalizable and in the basis $B = \{v_1, \ldots, v_n\}$ is given by a diagonal matrix $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ then $[Tv_i]_B = [T]_B[v_i]_B = \lambda_i e_i = \lambda_i [v_i]_B$ so $Tv_i = \lambda_i v_i$. It follows that each v_i is an eigenvector of V.

Conversely, suppose that $B = \{v_1, \ldots, v_n\}$ is a basis of V consisting of eigenvectors of V. Say, $Tv_i = \lambda_i v_i$. Then, by definition of $[T]_B$, we have $[T]_B = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. □

Theorem 8.2.4. *Let V be a finite dimensional vector space over a field $\mathbb{F}$, $T : V \to V$ a linear map. T is diagonalizable if and only if $m_g(\lambda) = m_a(\lambda)$ for any eigenvalue λ and the characteristic polynomial of T factors into linear factors over $\mathbb{F}$.*

Proof. Suppose first that T is diagonalizable and with respect to some basis $B = \{v_1, \ldots, v_n\}$ we have

$$[T]_B = \mathrm{diag}(\lambda_1, \ldots, \lambda_n).$$

By renumbering the vectors in B we may assume that in fact

$$[T]_B = \mathrm{diag}(\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_2, \ldots, \lambda_k, \ldots, \lambda_k),$$

where the λ_i are distinct and λ_i appears n_i times. We have then

$$\Delta_T(t) = \prod_{i=1}^{k} (t - \lambda_i)^{n_i}.$$

We see that the characteristic polynomial factors into linear factors and that the algebraic multiplicity of λ_i is n_i. Since we know that $m_g(\lambda_i) \leq m_a(\lambda_i) = n_i$, it is enough to prove that there are at least n_i independent eigenvectors for the eigenvalue λ_i. Without loss of generality, $i = 1$. Then $\{v_1, \ldots, v_{n_1}\}$ are independent and $Tv_j = \lambda_1 v_j$, $j = 1, \ldots, n_1$.

Conversely, suppose that the characteristic polynomial factors as

$$\Delta_T(t) = \prod_{i=1}^{k} (t - \lambda_i)^{n_i},$$

and that for every i we have $m_g(\lambda_i) = m_a(\lambda_i)$. Then, (for each $i = 1, \ldots, k$) we may find vectors $\{v_1^i, \ldots, v_{n_i}^i\}$, which form a basis for E_{λ_i}.

Lemma 8.2.5. *Let $\mu_1, \ldots, \mu_r$ be distinct eigenvalues of a linear map S and let $w_i \in E_{\mu_i}$ be non-zero vectors. If $\sum_{i=1}^{r} \alpha_i w_i = 0$ then each $\alpha_i = 0$.*

Proof of lemma. Suppose not. Then $\{w_1, \ldots, w_r\}$ are linearly dependent and so there is a first vector which is a linear combination of the preceding vectors, say w_{a+1}. Say,

$$w_{a+1} = \sum_{i=1}^{a} \beta_i \cdot w_i.$$

Apply S to get

$$\mu_{a+1} w_{a+1} = \sum_{i=1}^{a} \mu_i \beta_i \cdot w_i,$$

and also,

$$\mu_{a+1} w_{a+1} = \sum_{i=1}^{a} \mu_{a+1} \beta_i \cdot w_i.$$

Subtract to get

$$0 = \sum_{i=1}^{a} (\mu_{a+1} - \mu_i) \beta_i \cdot w_i.$$

Note that if $\beta_i \neq 0$ the coefficient $(\mu_{a+1} - \mu_i)\beta_i \neq 0$. Let $1 \leq j \leq a$ be the maximal index so that $\beta_j \neq 0$ then we get by rearranging

$$w_j = \sum_{i=1}^{j-1} -((\mu_{a+1} - \mu_j)\beta_j)^{-1})(\mu_{a+1} - \mu_i)\beta_i \cdot w_i.$$

This shows that a is not minimal and we got a contradiction. □

Coming back now to the proof of the Theorem, we shall prove that

$$\{v_j^i : i = 1, \ldots, k,\ j = 1, \ldots, n_i\}$$

is a linearly independent set. Since its cardinality is $\sum_{i=1}^k n_i = n$, it is a basis. Thus T has a basis consisting of eigenvectors, hence diagonalizable.

Suppose that we have a linear relation $\sum_{i=1}^k \sum_{j=1}^{n_i} \alpha_j^i v_j^i = 0$. Let $w_i = \sum_{j=1}^{n_i} \alpha_j^i v_j^i$ then $w_i \in E_{\lambda_i}$ and we have $w_1 + \cdots + w_k = 0$. Using the lemma, it follows that each w_i must be zero. Fixing an i, we find that $\sum_{j=1}^{n_i} \alpha_j^i v_j^i = 0$. But $\{v_j^i : j = 1, \ldots, n_i\}$ is a basis for E_{λ_i} so each $\alpha_j^i = 0$. □

The problem of diagonalization will occupy us for quite for a while. We shall study when we can diagonalize a matrix and how to do that, but for now, we just notice an important application of diagonalization using "primitive methods".

Lemma 8.2.6. *Let A be an $n \times n$ matrix and M an invertible $n \times n$ matrix such that $A = M\text{diag}(\lambda_1, \ldots, \lambda_n)M^{-1}$ then for $N \geq 1$,*

$$\boxed{A^N = M\text{diag}(\lambda_1^N, \ldots, \lambda_n^N)M^{-1}} \tag{8}$$

Proof. This is easily proved by induction. The case $N = 1$ is clear. Suppose that $A^N = M\text{diag}(\lambda_1^N, \ldots, \lambda_n^N)M^{-1}$ then

$$\begin{aligned} A^{N+1} &= A^N A = (M\text{diag}(\lambda_1^N, \ldots, \lambda_n^N)M^{-1})(MAM^{-1}) \\ &= M\text{diag}(\lambda_1^N, \ldots, \lambda_n^N)\text{diag}(\lambda_1, \ldots, \lambda_n)M^{-1} \\ &= M\text{diag}(\lambda_1^{N+1}, \ldots, \lambda_n^{N+1})M^{-1}. \end{aligned}$$

□

If we think about A as a linear transformation $T : \mathbb{F}^n \to \mathbb{F}^n$ then $A = [T]_{St}$ and B is the basis of eigenvectors so that $[T]_B = \text{diag}(\lambda_1^N, \ldots, \lambda_n^N)$, then

$$\boxed{M = {}_{St}M_B}$$

and, as we have seen, is simply the matrix whose columns are the elements of the basis B.

8.2.1. Here is a classical application. Consider the Fibonacci sequence :

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \ldots$$

It is defined recursively by

$$a_0 = 0, a_1 = 1, \qquad a_{n+2} = a_{n+1} + a_n,\ n \geq 0.$$

Let $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. Then

$$A \begin{pmatrix} a_n \\ a_{n+1} \end{pmatrix} = \begin{pmatrix} a_{n+1} \\ a_{n+2} \end{pmatrix}.$$

Therefore,

$$A^N \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} a_N \\ a_{N+1} \end{pmatrix}.$$

If we find a formula for A^N we then get a formula for a_N. We shall make use of Equation (8).

We saw in Example 8.1.13 that A has a basis of eigenvectors $B = \{v_1, v_2\}$ where $v_1 = \begin{pmatrix} 1 \\ \lambda_1 \end{pmatrix}$ corresponds to the eigenvalue $\lambda_1 = \frac{1+\sqrt{5}}{2}$ and $v_2 = \begin{pmatrix} 1 \\ \lambda_2 \end{pmatrix}$ corresponds to the eigenvalue $\lambda_2 = \frac{1-\sqrt{5}}{2}$. Let

$$M = \begin{pmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix} = {}_{St}M_B, \qquad M^{-1} = \frac{1}{\lambda_2 - \lambda_1} \begin{pmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{pmatrix} = {}_{B}M_{St}.$$

Then

$$A = \frac{1}{\lambda_2 - \lambda_1} \begin{pmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{pmatrix},$$

and so

$$\begin{aligned} A^N &= \frac{1}{\lambda_2 - \lambda_1} \begin{pmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix} \begin{pmatrix} \lambda_1^N & 0 \\ 0 & \lambda_2^N \end{pmatrix} \begin{pmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{pmatrix} \\ &= \frac{1}{\lambda_2 - \lambda_1} \begin{pmatrix} \lambda_1^N & \lambda_2^N \\ \lambda_1^{N+1} & \lambda_2^{N+1} \end{pmatrix} \begin{pmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{pmatrix} \\ &= \frac{1}{\lambda_2 - \lambda_1} \begin{pmatrix} \lambda_1^N \lambda_2 - \lambda_2^N \lambda_1 & \lambda_2^N - \lambda_1^N \\ \lambda_1^{N+1} \lambda_2 - \lambda_2^{N+1} \lambda_1 & \lambda_2^{N+1} - \lambda_1^{N+1} \end{pmatrix}. \end{aligned}$$

Therefore,

$$\begin{pmatrix} a_N \\ a_{N+1} \end{pmatrix} = A^N \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{\lambda_2^N - \lambda_1^N}{\lambda_2 - \lambda_1} \\ * \end{pmatrix}.$$

We conclude that

$$a_N = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^N - \left(\frac{1-\sqrt{5}}{2} \right)^N \right).$$

8.2.2. *Diagonalization Algorithm I.* We can summarize our discussion of diagonalization thus far as follows.

Given: $T : V \to V$ over a field $\mathbb{F}$.

(1) Calculate $\Delta_T(t)$.
(2) If $\Delta_T(t)$ does not factor into linear terms, stop. (Non-diagonalizable). Else:
(3) Calculate for each eigenvalue λ, E_λ and $m_g(\lambda)$. If for some λ, $m_g(\lambda) \neq m_a(\lambda)$, stop. (Non-diagonalizable). Else:
(4) For every λ find a basis $B^\lambda = \{v_1^\lambda, \ldots, v_{n(\lambda)}^\lambda\}$ for E_λ. Then $B = \cup_\lambda B^\lambda = \{v_1, \ldots, v_n\}$ is a basis for V. If $Tv_i = \lambda_i v_i$ then $[T]_B = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

We note that this is not really an algorithm, since there is no method to determine if a polynomial p factors into linear terms over an arbitrary field. This is possible, though, for a finite field $\mathbb{F}$ with q elements (the first step is calculating $d(t) = \gcd(p(t), t^q - t)$ and repeating that for $p(t)/d(t)$, and so on) and for the field of rational numbers since the numerators and denominators of rational roots are bounded. It is also possible to do over $\mathbb{R}$ (and trivial over $\mathbb{C}$). Thus, the issue of factorization into linear terms is not so crucial in the most important cases. The real problem is that there is no algorithm for calculating the roots in general. Again, for finite fields one may proceed by brute force, and over the rationals this is possible as we have mentioned but, for example, there is no algorithm for finding the real or complex roots in a closed forms (i.e., in radicals).

Now, to actually diagonalize a linear map T, there is no choice but finding the roots. However, it will turn out that it is algorithmically possible to decide whether T is diagonalizable or not without finding the roots. This is very useful because, once we know T is diagonalizable, then for many applications it is enough to approximate the roots and this is certainly possible over $\mathbb{R}$ and $\mathbb{C}$.

8.3. **The minimal polynomial and the theorem of Cayley-Hamilton.**
Let $g(t) = a_m t^m + \cdots + a_1 t + a_0$ be a polynomial in $\mathbb{F}[t]$ and let $A \in M_n(\mathbb{F})$. Then, by $g(A)$ we mean $a_m A^m + \cdots + a_1 A + a_0 \cdot I_n$. It is again a matrix in $M_n(\mathbb{F})$. We note that $(f+g)(A) = f(A) + g(A), (fg)(A) = f(A)g(A)$.

We begin with a lemma:

Lemma 8.3.1. *Let $A \in M_n(\mathbb{F})$, $f(t) \in \mathbb{F}[t]$ a monic polynomial. One can solve the equation*

$$(tI - A)(B_a t^a + B_{a-1} t^{a-1} + \cdots + B_0) = f(t) \cdot I_n,$$

with matrices $B_i \in M_n(\mathbb{F})$ if and only if $f(A) = 0$. (If $B = (b_{ij})$ is a matrix then by Bt^a, or $t^a B$, we mean the matrix $(b_{ij} t^a)$.)

Proof. Suppose we can solve the equation and w.l.o.g. $B_a \neq 0$. It then follows that $f(t)$ has degree $a+1$. Write

$$f(t) = t^{a+1} + b_a t^a + \cdots + b_0.$$

Equating coefficients we get

$$\begin{aligned} B_a &= I \\ B_{a-1} - AB_a &= b_a I \\ B_{a-2} - AB_{a-1} &= b_{a-1} I \\ &\vdots \\ B_1 - AB_2 &= b_2 I \\ B_0 - AB_1 &= b_1 I \\ -AB_0 &= b_0 I. \end{aligned}$$

Multiply from the left the first equation by A^{a+1}, the second by A^a, etc. and sum (the last equation is multiplied by the identity). We get

$$0 = A^{a+1} + b_a A^a + \cdots + b_1 A + b_0 I = f(A).$$

Conversely, suppose that $f(A) = 0$. Define, using the equations above,

$$\begin{aligned} B_a &= I \\ B_{a-1} &= b_a I + AB_a \\ B_{a-2} &= b_{a-1} I + AB_{a-1} \\ &\vdots \\ B_1 &= b_2 I + AB_2 \\ B_0 &= b_1 I + AB_1. \end{aligned}$$

We then have equality

$$(tI - A)(B_a t^a + B_{a-1} t^{a-1} + \cdots + B_0) = f(t) \cdot I_n,$$

if and only if $-AB_0 = b_0 I$. But,

$$\begin{aligned} AB_0 &= AB_0 - A^2 B_1 + A^2 B_1 - A^3 B_2 + \cdots + A^a B_{a-1} - A^{a+1} B_a + A^{a+1} B_a \\ &= A(B_0 - AB_1) + A^2(B_1 - AB_2) + \cdots + A^a(B_{a-1} - AB_a) + A^{a+1} B_a \\ &= Ab_1 + A^2 b_2 + \cdots + A^a b_a + A^{a+1} \\ &= f(A) - b_0 I \\ &= -b_0 I. \end{aligned}$$

□

Theorem 8.3.2 (Cayley-Hamilton)**.** *Let A be a matrix in $M_n(\mathbb{F})$. Then*

$$\Delta_A(A) = 0.$$

Namely, A solves its own characteristic polynomial.

Proof. We have

$$(tI - A)\operatorname{Adj}(tI - A) = \det(tI - A) \cdot I_n = \Delta_A(t) \cdot I_n.$$

Note that $\operatorname{Adj}(tI - A)$ is a matrix in $M_n(\mathbb{F}[t])$ and so can be written as $B_a t^a + B_{a-1} t^{a-1} + \cdots + B_0$ with $B_i \in M_n(\mathbb{F})$ (and in fact $a = n - 1$). It follows from Lemma 8.3.1 that $\Delta_A(A) = 0$. □

Proposition 8.3.3. *Let $A \in M_n(\mathbb{F})$. Let $m_A(t)$ be a monic polynomial of minimal degree among all monic polynomials f in $\mathbb{F}[t]$ such that $f(A) = 0$. Then, if $f(A) = 0$ then $m_A(t) | f(t)$ and in particular, $\deg(m_A(t)) \leq \deg(f(t))$. The polynomial $m_A(t)$ is called the **minimal polynomial** of A.*

Proof. First note that since $\Delta_A(t) = 0$, it makes sense to take a monic polynomial of minimal degree vanishing on A. Suppose that $f(A) = 0$. Let $h(t) = \gcd(m_A(t), f(t)) = a(t)m_A(t) + b(t)f(t)$. Then, by definition, h is monic and $h(A) = a(A)m_A(A) + b(A)f(A) = 0$. Since $h(t) | m_A(t)$ we have $\deg(h(t)) \leq \deg(m_A(t))$ and so, by definition of the minimal polynomial, we have $\deg(h(t)) = \deg(m_A(t))$. Since h is monic, divides $m_A(t)$, and of the same degree, we must have $h(t) = m_A(t)$ and thus $m_A(t) | f(t)$.

In particular, if there are two monic polynomials $m_A(t), m_A(t)'$ of that minimal degree that vanish on A they must be equal. Thus, the definition of $m_A(t)$ really depends on A alone. □

Theorem 8.3.4. *The polynomials m_A and Δ_A have the same irreducible factors. More precisely,*

$$m_A(t) | \Delta_A(t) | m_A(t)^n.$$

Proof. By Cayley-Hamilton $\Delta_A(A) = 0$ and so $m_A | \Delta_A$. On the other hand, because $m_A(A) = 0$ we can solve the equation in matrices:

$$(tI - A)(B_a t^a + B_{a-1} t^{a-1} + \cdots + B_0) = m_A(t) \cdot I_n,$$

which we write more compactly as

$$(tI - A)G(t) = m_A(t) \cdot I_n,$$

where $G(t) \in M_n(\mathbb{F}[t])$. Taking determinants, we get

$$\Delta_A(t) \cdot \det(G(t)) = m_A(t)^n,$$

and so $\Delta_A | m_A^n$. □

Corollary 8.3.5. *If $\Delta_A(t)$ is a product of distinct irreducible factors then $m_A = \Delta_A$.*

Example 8.3.6. The matrix $A = \left(\begin{smallmatrix} -1 & 6 \\ -1 & 4 \end{smallmatrix}\right)$ has $\Delta_A(t) = (t-1)(t-2)$ and so $m_A(t) = \Delta_A(t)$.

The matrix $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ has $\Delta_A(t) = (t-1)^2$ so we know $m_A(t)$ is either $t-1$ or $(t-1)^2$. Since $A - I = 0$, we conclude that $m_A(t) = t - 1$.

On the other hand, the matrix $A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ also has $\Delta_A(t) = (t-1)^2$ so again we know that $m_A(t)$ is either $t-1$ or $(t-1)^2$. Since $A - I = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right) \neq 0$, we conclude that $m_A(t) = (t-1)^2$.

8.4. **The Primary Decomposition Theorem.** Recall the definition of internal direct sum from § 3.5. One version is to say that a vector space V is the internal direct sum of subspaces $W_1, \ldots, W_r$ if every vector $v \in V$ has a unique expression

$$v = w_1 + \cdots + w_r, \qquad w_i \in W_i.$$

We write then

$$V = W_1 \oplus \cdots \oplus W_r. \tag{9}$$

Definition 8.4.1. Let $T : V \to V$ be a linear map. We say that a subspace $W \subseteq V$ is **T-invariant** if $T(W) \subseteq W$. In this case, we can consider $T|_W : W \to W$, the restriction of T to W,

$$T|_W(w) = T(w), \qquad w \in W.$$

Suppose that in the decomposition (9) each W_i is T-invariant. Denote $T|_{W_i}$ by T_i. We then write

$$T = T_1 \oplus \cdots \oplus T_r,$$

meaning

$$T(v) = T_1(w_1) + \cdots + T_r(w_r),$$

which indeed holds true! Conversely, given any linear maps $T_i : W_i \to W_i$ we can define a linear map $T : V \to V$ by

$$T(v) = T_1(w_1) + \cdots + T_r(w_r).$$

We have then $T|_{W_i} = T_i$.

In the setting of (9) we have $\dim(V) = \dim(W_1) + \cdots + \dim(W_r)$. More precisely, we have the following lemma.

Lemma 8.4.2. *Suppose that*

$$B_i = \{w_1^i, \ldots, w_{n(i)}^i\}$$

is a basis for W_i. Then

$$B = \cup_{i=1}^r B_i = \{w_1^1, \ldots, w_{n(1)}^1, \ldots, w_1^r, \ldots, w_{n(r)}^r\}$$

is a basis of V.

Proof. Indeed, B is a spanning set, because $v = w_1 + \cdots + w_r$ and we can write w_i using the basis B_i, $w_i = \sum_{j=1}^{n(j)} \alpha_j^i w_j^i$, and so

$$v = \sum_{i=1}^{r} \sum_{j=1}^{n(i)} \alpha_j^i w_j^i.$$

B is also independent. Suppose that

$$0 = \sum_{i=1}^{r} \sum_{j=1}^{n(i)} \alpha_j^i w_j^i = \sum_{i=1}^{r} (\sum_{j=1}^{n(i)} \alpha_j^i w_j^i).$$

Let $w_i = \sum_{j=1}^{n(i)} \alpha_j^i w_j^i$. Then $w_i \in W_i$ and, since the subspaces W_i give a direct sum, each $w_i = 0$. Since B_i is a basis for W_i, each $\alpha_j^i = 0$. □

We conclude that if T_i is represented on W_i by the matrix A_i, relative to the basis B_i, that is,

$$[T_i]_{B_i} = A_i,$$

then T is given in block diagonal form in the basis B,

$$[T]_B = \begin{pmatrix} \boxed{A_1} & & & \\ & \boxed{A_2} & & \\ & & \ddots & \\ & & & \boxed{A_r} \end{pmatrix}.$$

Example 8.4.3. Here is an example of a T-invariant subspace. Take any vector $v \in V$ and let $W = \text{Span}(\{v, Tv, T^2v, T^3v \ldots, \})$. This is called a **cyclic subspace**. In fact, any T-invariant space is a sum (not necessarily direct) of cyclic subspaces.

Another example is the following. Let $g(t) \in \mathbb{F}[t]$ and let $W = \text{Ker}(g(T))$. Namely, if $g(t) = a_m t^m + \cdots + a_1 t + a_0$, then W is the kernel of the linear map $a_m T^m + \cdots + a_1 T + a_0 \text{Id}$. Since $g(T)T = Tg(T)$, it is immediate that W is T-invariant.

Theorem 8.4.4 (Primary Decomposition Theorem)**.** *Let $T : V \to V$ be a linear operator with*

$$m_T(t) = f_1(t)^{n_1} \cdots f_r(t)^{n_r},$$

where the f_i are the irreducible factors of m_T. Let

$$W_i = \text{Ker}(f_i(T)^{n_i}).$$

Then,

$$V = W_1 \oplus \cdots \oplus W_r$$

and $f_i(t)^{n_i}$ is the minimal polynomial of $T_i := T|_{W_i}$.

Lemma 8.4.5. *Suppose $T : V \to V$ is a linear map, $f(t) \in \mathbb{F}[t]$ satisfies $f(T) = 0$ and $f(t) = g(t)h(t)$ with $\gcd(g(t), h(t)) = 1$. Then*

$$V = \text{Ker}(g(t)) \oplus \text{Ker}(h(t)).$$

Proof. For suitable polynomials $a(t), b(t) \in \mathbb{F}[t]$ we have

$$1 = g(t)a(t) + h(t)b(t),$$

and so,

$$\text{Id} = g(T) \circ a(T) + h(T) \circ b(T).$$

For every v we now have

$$v = g(T) \circ a(T)v + h(T) \circ b(T)v.$$

We note that $g(T) \circ a(T)v \in \text{Ker}(h(T)), h(T) \circ b(T)v \in \text{Ker}(g(T))$. Therefore,

$$V = \text{Ker}(h(T)) + \text{Ker}(g(T)).$$

Suppose that $v \in \text{Ker}(g(T)) \cap \text{Ker}(h(T))$. Then, $v = g(T) \circ a(T)v + h(T) \circ b(T)v = a(T) \circ g(T)v + b(T) \circ h(T)v = 0 + 0 = 0$. Thus,

$$V = \text{Ker}(g(T)) \oplus \text{Ker}(h(T)).$$

□

Corollary 8.4.6. *We have*

$$\text{Ker}(h(T)) = g(T)V, \qquad \text{Ker}(g(T)) = h(T)V.$$

Proof. We have seen $V = g(T) \circ a(T)V + h(T) \circ b(T)V$, which implies that $V = g(T)V + h(T)V$. Thus, $\dim(g(T)V) + \dim(h(T)V) \geq \dim(V)$. On the other hand $g(T)V \subseteq \text{Ker}(h(T))$ and $h(T)V \subseteq \text{Ker}(g(T))$ and so $\dim(g(T)V) + \dim(h(T)V) \leq \dim \text{Ker}(h(T)) + \dim \text{Ker}(g(T)) = \dim(V)$. We conclude that $\dim(g(T)V) + \dim(h(T)V) = \dim(V)$ and so that $\dim(g(T)V) = \dim \text{Ker}(h(T)), \dim(h(T)V) = \dim \text{Ker}(g(T))$. Therefore, $\text{Ker}(h(T)) = g(T)V$ and $\text{Ker}(g(T)) = h(T)V$. □

Lemma 8.4.7. *In the situation of the previous Lemma, let $W_1 = \text{Ker}(g(T)), W_2 = \text{Ker}(h(T))$. Assume that f is the minimal polynomial of T then $g(t)$ is the minimal polynomial of $T_1 := T|_{W_1}$ and $h(t)$ is the minimal polynomial of $T_2 := T|_{W_2}$.*

Proof. Let $m_i(t)$ be the minimal polynomial of T_i. Clearly $g(T_1) = 0$ and $h(T_2) = 0$. Thus, $m_1|g, m_2|h$ and $m_1 m_2|gh = f$, which is the minimal polynomial. But it is clear that $(m_1 m_2)(T) = m_1(T)m_2(T)$ is zero because it is a linear transformation whose restriction to W_i is $m_1(T_i)m_2(T_i) = 0$. Therefore $f|m_1 m_2$ and it follows that $m_1 = g, m_2 = h$. □

With these preparations we are ready to prove the Primary Decomposition Theorem.

Proof. We argue by induction on r. The case $r = 1$ is trivial. Assume the theorem holds true for some $r \geq 1$. We prove it for $r + 1$.

Write

$$\begin{aligned} m_T(t) &= (f_1(t)^{n_1} \cdots f_r(t)^{n_r}) \cdot f_{r+1}(t)^{n_{r+1}} \\ &= g(t) \cdot h(t). \end{aligned} \tag{10}$$

Applying the two lemmas, we conclude that

$$V = W_1' \oplus W_2',$$

where the W_i' are the T-invariant subspaces $W_1' = \text{Ker}(g(T)), W_2' = \text{Ker}(h(T))$ and, furthermore, $g(t)$ is the minimal polynomial of $T_1 := T|_{W_1'}$, $h(t)$ of $T_2 := T|_{W_2'}$.

We let $W_{r+1} = W_2'$. Using induction applied to W_1', since

$$g(t) = f_1(t)^{n_1} \cdots f_r(t)^{n_r},$$

we get

$$W_1' = W_1 \oplus \cdots \oplus W_r,$$

with $W_i = \text{Ker}(f_i(T)^{n_i} : W_1' \to W_1')$. It only remains to show that $W_i = \text{Ker}(f_i(T)^{n_i} : V \to V)$ and the inclusion $\subseteq$ is clear. Now, if $v \in V$ and $f_i(T)^{n_i}(v) = 0$ then $g(t)v = 0$ and so $v \in W_1'$. Thus we get the opposite inclusion $\supseteq$. □

We can now deduce one of the most important results in the theory of linear maps.

Corollary 8.4.8. *$T : V \to V$ is diagonalizable if and only if the minimal polynomial of T factors into distinct linear terms over $\mathbb{F}$,*

$$m_T(t) = (t - \lambda_1) \cdots (t - \lambda_r), \qquad \lambda_i \in \mathbb{F},\ \lambda_i \neq \lambda_j \text{ for } i \neq j.$$

Proof. Suppose that T is diagonalizable. Thus, in some basis B of V we have

$$A = [T]_B = \mathrm{diag}(\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_2, \ldots, \lambda_r, \ldots, \lambda_r).$$

Then, the minimal polynomial divides $\prod_{i=1}^{r}(t-\lambda_i)$ and equal to it if this polynomial has A as a root. Note that

$$\begin{aligned}\prod_{i=1}^{r}(A-\lambda_i I_n) &= \mathrm{diag}(0,\ldots,0,\lambda_2-\lambda_1,\ldots,\lambda_2-\lambda_1,\ldots,\lambda_r-\lambda_1,\ldots,\lambda_r-\lambda_1)\\ &\times \mathrm{diag}(\lambda_1-\lambda_2,\ldots,\lambda_1-\lambda_2,0,\ldots,0,\ldots,\lambda_r-\lambda_2,\ldots,\lambda_r-\lambda_2)\\ &\times \mathrm{diag}(\lambda_1-\lambda_r,\ldots,\lambda_1-\lambda_r,\lambda_2-\lambda_r,\ldots,\lambda_2-\lambda_r,\ldots,0,\ldots,0) = 0.\end{aligned}$$

Now suppose that

$$m_T(t) = (t-\lambda_1)\cdots(t-\lambda_r), \qquad \lambda_i \in \mathbb{F},\ \lambda_i \neq \lambda_j \text{ for } i \neq j.$$

Consider the Primary Decomposition. We get that

$$T = T_1 \oplus \cdots \oplus T_r,$$

where T_i is $T|_{W_i}$ and $W_i = \mathrm{Ker}(T-\lambda_i) = E_{\lambda_i}$. And so, $T_i|_{W_i}$ is represented by a diagonal matrix $\mathrm{diag}(\lambda_i, \ldots, \lambda_i)$. Since $V = W_1 \oplus \cdots \oplus W_r$, we have that if we take as a basis for V the set $B = \cup_{i=1}^{r} B_i$, where B_i is a basis for W_i then T is represented in the basis B by $\mathrm{diag}(\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_2, \ldots, \lambda_r, \ldots, \lambda_r)$. □

Corollary 8.4.9. *Let $T : V \to V$ be a diagonalizable linear map and let $W \subseteq V$ be a T-invariant subspace. Then $T_1 := T|_W$ is diagonalizable.*

Proof. We know that m_T is a product of distinct linear factors over the field $\mathbb{F}$. Clearly $m_T(T_1) = 0$ (this just says that $m_T(T)$ is zero on W, which is clear since $m_T(T) = 0$). It follows that $m_{T_1} | m_T$ and so is also a product of distinct linear terms over the field $\mathbb{F}$. Thus, T_1 is diagonalizable. □

Here is another very useful corollary of our results; the proof is left as an exercise.

Corollary 8.4.10. *Let $S, T : V \to V$ be commuting and diagonalizable linear maps ($ST = TS$). Then there is a basis B of V in which both S and T are diagonal. ("commuting matrices can be simultaneously diagonalized".)*

Example 8.4.11. For some numerical examples see the files ExampleA, ExampleB, ExampleB1 on the course webpage.

8.5. **More on finding the minimal polynomial.** In the assignments we explain how to find the minimal polynomial without factoring.

8.5.1. *Diagonalization Algorithm II.*

Given: $T : V \to V$ over a field $\mathbb{F}$.

(1) Calculate $m_T(t)$, for example using the method of cyclic subspaces (see assignments).
(2) If $\gcd(m_T(t), m_T(t)') \neq 1$, stop. (Non-diagonalizable). Else:
(3) If $m_T(t)$ does not factor into linear terms, stop. (Non-diagonalizable). Else:
(4) The map T is diagonalizable. For every λ find a basis $B^\lambda = \{v_1^\lambda, \ldots, v_{n(\lambda)}^\lambda\}$ for the eignespace E_λ. Then $B = \cup_\lambda B^\lambda = \{v_1, \ldots, v_n\}$ is a basis for V. If $Tv_i = \lambda_i v_i$ then $[T]_B = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

We make some remarks on the advantage of this algorithm. First, the minimal polynomial can be calculated without factoring the characteristic polynomial (which we don't even need to calculate for this algorithm). If $\gcd(m_T(t), m_T(t)') = 1$ then $m_T(t)$ has no repeated roots. The calculation of $\gcd(m_T(t), m_T(t)')$ doesn't require factoring. It is done using the Euclidean algorithm and is very fast. Thus, we can efficiently and quickly decide if T is diagonalizable or not. Of course, the actual diagonalization requires finding the eigenvalues and hence the roots of $m_T(t)$. There is no algorithm for that. There are other ways to simplify the study of a linear map which do not require factoring (and in particular do not bring T into diagonal form). This is the rational canonical form, studied in MATH 371.

9. The Jordan canonical form

Let $T : V \to V$ be a linear map on a finite dimensional vector space V. In this section we assume that the minimal polynomial of T factors into linear terms:

$$\boxed{m_T = (t-\lambda_1)^{m_1} \cdots (t-\lambda_r)^{m_r}, \qquad \lambda_i \in \mathbb{F}}$$

We therefore get by the Primary Decomposition Theorem (PDT)

$$V = W_1 \oplus \cdots \oplus W_r,$$

where $W_i = \mathrm{Ker}((T - \lambda \cdot \mathrm{Id})^{m_i})$ and the minimal polynomial of $T_i = T|_{W_i}$ on W_i is $(t-\lambda_i)^{m_i}$. If we use the notation $\Delta_T(t) = (t-\lambda_1)^{n_1} \cdots (t-\lambda_r)^{n_r}$ then, since the characteristic polynomial of T_i is a power of $(t-\lambda_i)$, we must have that the characteristic polynomial of T_i is precisely $(t-\lambda_i)^{n_i}$ and in particular $\dim(W_i) = n_i$.

9.1. **Preparations.** The Jordan canonical form theory picks up where PDT is signing off. Using PDT, we restrict our attention to linear transformations $T : V \to V$ whose minimal polynomial is of the form $(t-\lambda)^m$ and, say, $\dim(V) = n$. Thus,

$$m_T(t) = (t-\lambda)^m, \qquad \Delta_T(t) = (t-\lambda)^n.$$

We write

$$T = \lambda \cdot \mathrm{Id} + U \;\Rightarrow\; U = T - \lambda \cdot \mathrm{Id},$$

then U is nilpotent. In fact,

$$m_U(t) = t^m, \qquad \Delta_U(t) = t^n.$$

The integer m is also called the **index of nilpotence** of U. Let us assume for now the following fact.

Proposition 9.1.1. *A nilpotent operator U is represented in a suitable basis by a block diagonal matrix*

$$\begin{pmatrix} \boxed{N_1} & & & \\ & \boxed{N_2} & & \\ & & \ddots & \\ & & & \boxed{N_d} \end{pmatrix},$$

*such that each N_i is a **standard nilpotent matrix** of size k_i, i.e., of the form:*

$$N = \begin{pmatrix} 0 & 1 & 0 & \cdots & & 0 \\ & 0 & 1 & 0 & \cdots & 0 \\ & & \ddots & \ddots & & \\ & & & & & \\ & & & & 0 & 1 \\ & & & & & 0 \end{pmatrix}.$$

Relating this back to the transformation T, it follows that in the same basis T is given by

$$\begin{pmatrix} \boxed{\lambda \cdot I_{k_1} + N_1} & & & \\ & \boxed{\lambda \cdot I_{k_2} + N_2} & & \\ & & \ddots & \\ & & & \boxed{\lambda \cdot I_{k_d} + N_d} \end{pmatrix}.$$

The blocks have the shape

$$\lambda \cdot I + N = \begin{pmatrix} \lambda & 1 & 0 & \cdots & & 0 \\ & \lambda & 1 & 0 & \cdots & 0 \\ & & \ddots & \ddots & & \\ & & & & & \\ & & & & \lambda & 1 \\ & & & & & \lambda \end{pmatrix}.$$

Such blocks are called Jordan canonical blocks.

Suppose that the size of N is k. If the corresponding basis vectors are $\{b_1, \ldots, b_k\}$ then N has the effect

$$b_k \to b_{k-1} \to \ \ldots \ \to b_2 \to b_1 \to 0.$$

From that, or by actually performing the multiplication, it is easy to see that

$$N^a = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 & & & \\ & & & & 0 & 1 & & \\ & & & & & \ddots & & \\ & & & & & & 0 & 1 \\ & & & & & & & 0 \\ & & & 0 & & & & \end{pmatrix},$$

where the first row begins with a zeros. In particular, if N has size k, the minimal polynomial of N is t^k. We conclude that m, the index of nilpotence of U, is the maximum of the index of nilpotence of the matrices $N_1, \ldots, N_d$. That is,

$$m = \max\{k_i : i = 1, \ldots, d\}.$$

We introduce the following notation: Let S be a linear map, the **nullity of** S is

$$\text{null}(S) = \dim(\text{Ker}(S)).$$

Then $\text{null}(N) = 1$ and, more generally,

$$\text{null}(N^a) = \begin{cases} a & a \leq k \\ k & a \geq k. \end{cases}$$

We illustrate it for matrices of small size:

N	N^2	N^3	N^4
$\begin{pmatrix} 0 \end{pmatrix}$	$\begin{pmatrix} 0 \end{pmatrix}$	$\begin{pmatrix} 0 \end{pmatrix}$	$\begin{pmatrix} 0 \end{pmatrix}$
$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$
$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$
$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

We have

$$\text{null}(U) = \sharp\ 1\text{-blocks} + \sharp\ 2\text{-blocks} + \sharp\ 3\text{-blocks} + \sharp\ 4\text{-blocks} + \sharp\ 5\text{-blocks} + \dots$$

and so

$$\text{null}(U) = \sharp\text{ -blocks.}$$

Similarly,

$$\text{null}(U^2) = \sharp\ 1\text{-blocks} + 2 \cdot \sharp\ 2\text{-blocks} + 2 \cdot \sharp\ 3\text{-blocks} + 2 \cdot \sharp\ 4\text{-blocks} + 2 \cdot \sharp\ 5\text{-blocks} + \dots,$$

$$\text{null}(U^3) = \sharp\ 1\text{-blocks} + 2 \cdot \sharp\ 2\text{-blocks} + 3 \cdot \sharp\ 3\text{-blocks} + 3 \cdot \sharp\ 4\text{-blocks} + 3 \cdot \sharp\ 5\text{-blocks} + \dots$$

$$\text{null}(U^4) = \sharp\ 1\text{-blocks} + 2 \cdot \sharp\ 2\text{-blocks} + 3 \cdot \sharp\ 3\text{-blocks} + 4 \cdot \sharp\ 4\text{-blocks} + 4 \cdot \sharp\ 5\text{-blocks} + \dots$$

To simplify notation, let $U^0 = \text{Id}$. Then we conclude that

$$\sharp\ 1\text{-blocks} = 2 \cdot \text{null}(U) - (\text{null}(U^2) + \text{null}(U^0)),$$

$$\sharp\ 2\text{-blocks} = 2 \cdot \text{null}(U^2) - (\text{null}(U^3) + \text{null}(U)),$$

$$\sharp\ 3\text{-blocks} = 2 \cdot \text{null}(U^3) - (\text{null}(U^4) + \text{null}(U^2)),$$

and so on. We summarize our discussion as follows.

Proposition 9.1.2. *The number of blocks is* $\text{null}(U)$ *and the size of the largest block is* m*, the index of nilpotence of* U*, where* $m_U(t) = t^m$*. The number of blocks of size* b*,* $b \geq 1$*, is given by the following formula:*

$$\sharp\ b\text{-blocks} = 2 \cdot \text{null}(U^b) - (\text{null}(U^{b+1}) + \text{null}(U^{b-1}))$$

9.2. **The Jordan canonical form.** Let $\lambda \in \mathbb{F}$. A **Jordan block** $J_a(\lambda)$ is an $a \times a$ matrix with λ on the diagonal and 1's above the diagonal (the rest of the entries being zero):

$$J_a(\lambda) = \begin{pmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \lambda & 1 \\ & & & \lambda \end{pmatrix}.$$

Thus, $J_a(\lambda) - \lambda I_a$ is a standard nilpotent matrix. Here are how Jordan blocks of size at most 4 look like:

$$\begin{pmatrix} \lambda \end{pmatrix}, \quad \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}, \quad \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}, \quad \begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{pmatrix}.$$

We can now state the main theorem of this section.

Theorem 9.2.1 (Jordan Canonical Form)**.** *Let $T : V \to V$ be a linear map whose characteristic and minimal polynomials are*

$$\Delta_T(t) = (t-\lambda_1)^{n_1} \cdots (t-\lambda_r)^{n_r}, \qquad m_T(t) = (t-\lambda_1)^{m_1} \cdots (t-\lambda_r)^{m_r}.$$

*Then, in a suitable basis, T has a block diagonal matrix representation J (called a **Jordan canonical form**) where the blocks are Jordan blocks of the form $J_{a(i,1)}(\lambda_i), J_{a(i,2)}(\lambda_i), \dots$ with $a(i,1) \geq a(i,2) \geq \dots$.*

The following holds:

1. *For every λ_i, $a(i,1) = m_i$. ("The maximal block of λ_i is of size equal to the power of $(t-\lambda_i)$ in the minimal polynomial.")*
2. *For every λ_i, $\sum_{j\geq 1} a(i,j) = n_i$. ("The total size of the blocks of λ_i is the algebraic multiplicity of λ_i.")*
3. *For every λ_i, the number of blocks of the form $J_{a(i,k)}(\lambda_i)$ is $m_g(\lambda_i)$. ("The number of blocks corresponding to λ_i is the geometric multiplicity of λ_i.")*
4. *For every λ_i, the number of blocks $J_{a(i,k)}(\lambda_i)$ of size b is $2 \cdot \mathrm{null}(U_i^b) - (\mathrm{null}(U_i^{b+1}) + \mathrm{null}(U_i^{b-1}))$, where $U_i = T_i - \lambda_i \mathrm{Id}_{W_i}$. One may also take in this formula $U_i = T - \lambda_i \mathrm{Id}$.*

The theorem follows immediately from our discussion in the previous section. Perhaps the only remark one should still make that is that $\mathrm{null}(T - \lambda_i \cdot \mathrm{Id}) = \mathrm{null}((T - \lambda_i \cdot \mathrm{Id})|_{W_i}) = \mathrm{null}(T_i - \lambda_i \mathrm{Id}_{W_i})$, where $W_i = \mathrm{Ker}((T - \lambda_i \mathrm{Id})^{n_i})$, simply because the kernel of $(T - \lambda_i \cdot \mathrm{Id})^a$ for any a is contained in W_i.

It remains to explain Proposition 9.1.1 and how one finds such basis in practice. This is our next subject.

9.3. **Standard form for nilpotent operators.** Let $U : V \to V$ be a nilpotent operator,

$$m_U(t) = t^m, \qquad \Delta_U(t) = t^n.$$

We now explain how to find a basis for V with respect to which U is represented by a block diagonal form, where the blocks are standard nilpotent matrices, as asserted in Proposition 9.1.1.

Write

$$\mathrm{Ker}(U^m) = \mathrm{Ker}(U^{m-1}) \oplus C^m,$$

for some subspace C^m ("C" is for Complementary). We note the following:

$$UC^m \subseteq \mathrm{Ker}(U^{m-1}), \qquad UC^m \cap \mathrm{Ker}(U^{m-2}) = \{0\}.$$

Find $C^{m-1} \supseteq UC^m$ such that

$$\mathrm{Ker}(U^{m-1}) = \mathrm{Ker}(U^{m-2}) \oplus C^{m-1}.$$

Suppose we have already found a decomposition of V as

$$V = \underbrace{\underbrace{\mathrm{Ker}(U^{i-1}) \oplus C^i}_{Ker(U^i)} \oplus C^{i+1} \oplus \cdots \oplus C^m}_{Ker(U^{i+1})},$$

such that $UC^j \subseteq C^{j-1}$. We note that

$$UC^i \subseteq \mathrm{Ker}(U^{i-1}), \qquad UC^i \cap \mathrm{Ker}(U^{i-2}) = \{0\}.$$

We may find therefore a subspace C^{i-1} such that $C^{i-1} \supseteq UC^i$ and $\mathrm{Ker}(U^{i-1}) = \mathrm{Ker}(U^{i-2}) \oplus C^{i-1}$, and so on. We conclude that

$$V = \underbrace{\underbrace{\underbrace{C^1}_{\mathrm{Ker}(U)} \oplus C^2}_{Ker(U^2)} \oplus C^3 \oplus \cdots \oplus C^m}_{Ker(U^3)}$$

and

$$UC^i \subseteq C^{i-1}, \qquad 1 < i \leq m.$$

We also note that the map $U : C^i \to C^{i-1}$ $(1 < i \leq m)$ is injective. We therefore conclude the following procedure for finding a basis for V:

- Find a basis $\{v_1^m, \ldots, v_{n(m)}^m\}$ for C^m, where $V = \ker(U^m) = \ker(U^{m-1}) \oplus C^m$.
- $\{Uv_1^m, \ldots, Uv_{n(m)}^m\}$ are linearly independent and so can be completed to a basis for C^{m-1} by vectors $\{v_1^{m-1}, \ldots, v_{n(m-1)}^{m-1}\}$, where C^{m-1} is such that $\ker(U^{m-1}) = \ker(U^{m-2}) \oplus C^{m-1}$.
- Now the vectors $\{U^2v_1^m, \ldots, U^2v_{n(m)}^m, Uv_1^{m-1}, \ldots, Uv_{n(m-1)}^{m-1}\}$ are linearly independent and so can be completed to a basis for C^{m-2} by vectors $\{v_1^{m-2}, \ldots, v_{n(m-2)}^{m-2}\}$, where C^{m-2} is such that $\ker(U^{m-2}) = \ker(U^{m-3}) \oplus C^{m-2}$.
- etc.

We get a basis of V,

$$\{U^a v_i^j : 1 \leq j \leq m, 1 \leq i \leq n(j), 0 \leq a \leq j-1\}.$$

The basis now is ordered in the following way (first the first row, then the second row, etc.):

$$
\begin{array}{cccccc}
U^{m-1}v_1^m & U^{m-2}v_1^m & \dots & \dots & Uv_1^m & v_1^m \\
 & \vdots & & & & \\
U^{m-1}v_{n(m)}^m & U^{m-2}v_{n(m)}^m & \dots & \dots & Uv_{n(m)}^m & v_{n(m)}^m \\
 & U^{m-2}v_1^{m-1} & U^{m-3}v_1^{m-1} & \dots & Uv_1^{m-1} & v_1^{m-1} \\
 & \vdots & & & & \\
 & U^{m-2}v_{n(m-1)}^{m-1} & U^{m-3}v_{n(m-1)}^{m-1} & \dots & Uv_{n(m-1)}^{m-1} & v_{n(m-1)}^{m-1} \\
 & \vdots & & & & \\
 & & & \vdots & & \\
 & & & & Uv_1^2 & v_1^2 \\
 & & & & \vdots & \\
 & & & & Uv_{n(2)}^2 & v_{n(2)}^2 \\
 & & & & & v_1^1 \\
 & & & & & \vdots \\
 & & & & & v_{n(1)}^1
\end{array}
$$

A row of length k contributes a standard nilpoten matrix of size k. In particular, the last rows are the zero matrices, the rows before them that have length 2 give blocks of the form $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and so on.

Example 9.3.1. Here is a toy example. More complicated examples appear as ExampleC on the course webpage. Consider the matrix

$$U = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix},$$

which is nilpotent. The kernel of U is $\mathrm{Span}(e_1)$. The kernel of U^3 is the whole space. The kernel of $U^2 = \begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ is $\mathrm{Span}(e_1, e_2)$. Therefore, we may take $C^3 = \mathrm{Span}(e_3)$ and let $v_1^3 = e_3$. Then $Uv_1^3 = (1, 2, 0)$. We note that $\ker(U^2) = \ker(U) \oplus \mathrm{Span}((1, 2, 0))$. It follows that the basis we want is just $\{U^2v_1^3, Uv_1^3, v_1^3\}$, equal to $\{(2, 0, 0), (1, 2, 0), (0, 0, 1)\}$. In this basis the transformation is represented by a standard nilpotent matrix of size 3.

9.3.1. *An application of the Jordan canonical form.* One problem that arises often is the calculation of a high power of a matrix. Solving this problem was one of our motivations for discussing diagonalization. Even if a matrix is not diagonalizable, the Jordan canonical form can be used to great effect to calculate high powers of a matrix.

Let A therefore be a square matrix and J its Jordan canonical form. There is an invertible matrix M such that $A = MJM^{-1}$ and so $A^N = MJ^NM^{-1}$. Now, if $J = \mathrm{diag}(J_1, \dots, J_r)$, where the

J_i are the Jordan blocks then

$$J^N = \text{diag}(J_1^N, \ldots, J_r^N).$$

We therefore focus on calculating J^N, assuming J is a Jordan block $J(\lambda)$ of size N. Write

$$J(\lambda) = \lambda \cdot I_n + U.$$

Since $\lambda \cdot I_n$ and U commute, the binomial formula gives

$$J(\lambda)^N = \sum_{i=0}^{N} \binom{N}{i} \lambda^{N-i} U^i.$$

Notice that if $i \geq n$ then $U^i = 0$. We therefore get a convenient formula. We illustrate the formula for 2×2 and 3×3 matrices:

- For a 2×2 matrix $A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$ we have

$$A^N = \begin{pmatrix} \lambda^N & n\lambda^{N-1} \\ 0 & \lambda^N \end{pmatrix}.$$

- For a 3×3 matrix $A = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$ we have for $N \geq 2$

$$A^N = \begin{pmatrix} \lambda^N & N\lambda^{N-1} & \frac{N(N-1)}{2}\lambda^{N-2} \\ 0 & \lambda^N & N\lambda^{N-1} \\ 0 & 0 & \lambda^N \end{pmatrix}.$$

10. Diagonalization of symmetric, self-adjoint and normal operators

In this section, we marry the theory of inner products and the theory of diagonalization, to consider the diagonalization of special type of operators. Since we are dealing with inner product spaces, we assume that the field $\mathbb{F}$ over which all vector spaces, linear maps and matrices in this section are defined is either $\mathbb{R}$ or $\mathbb{C}$.

10.1. **The adjoint operator.** Let V be an inner product space, of finite dimension.

Proposition 10.1.1. *Let $T : V \to V$ be a linear operator. There exists a unique linear operator $T^* : V \to V$ such that*

$$\langle Tu, v\rangle = \langle u, T^*v\rangle, \qquad \forall u, v \in V.$$

*The linear operator T^* is called the **adjoint** of T.* [6] *Furthermore, if B is an orthonormal basis then*

$$[T^*]_B = [T]_B^* \ (:= \overline{[T]_B^t}).$$

Proof. We first show uniqueness. Suppose that we had two linear maps S_1, S_2 such that

$$\langle Tu, v\rangle = \langle u, S_i v\rangle, \qquad \forall u, v \in V.$$

Then, for all $u, v \in V$ we have

$$\langle u, (S_1 - S_2)v\rangle = 0.$$

In particular, this equation holds for all v with the vector $u = (S_1 - S_2)v$. This gives us that for all v, $\langle (S_1 - S_2)v, (S_1 - S_2)v\rangle = 0$, which in turn implies that for all v $(S_1 - S_2)v = 0$. That is, $S_1 = S_2$.

We now show T^* exists. Let B be an orthonormal basis for V. We have,

$$\begin{aligned}\langle Tu, v\rangle &= ([T]_B[u]_B)^t \cdot \overline{[v]_B} \\ &= [u]_B^t \cdot [T]_B^t \overline{[v]_B} \\ &= [u]_B^t \cdot \overline{\overline{[T]_B^t}[v]_B}.\end{aligned}$$

Let T^* be the linear map represented in the basis B by $\overline{[T]_B^t}$, i.e., by $[T]_B^*$. □

Lemma 10.1.2. *The following identities hold:*

(1) $(T_1 + T_2)^* = T_1^* + T_2^*$;
(2) $(T_1 \circ T_2)^* = T_2^* \circ T_1^*$;
(3) $(\alpha T)^* = \bar{\alpha} T^*$;
(4) $(T^*)^* = T$.

Proof. This all follows easily from the corresponding identities of matrices (using that $[T_1]_B = C, [T_2]_B = D$ implies $[T_1 + T_2]_B = C + D$ etc.):

(1) $(C + D)^* = C^* + D^*$;
(2) $(CD)^* = D^*C^*$; (Use that $(CD)^t = D^tC^t$.)
(3) $(\alpha C)^* = \bar{\alpha} C^*$;
(4) $(C^*)^* = C$.

□

[6] Caution: If A is the matrix representing T, the matrix representing T^* has nothing to do with the adjoint $\mathrm{Adj}(A)$ of the matrix A that was used in the section about determinants.

10.2. **Self-adjoint operators.** We keep the notation of the previous section.

Definition 10.2.1. T is called a **self-adjoint** operator if $T = T^*$. This equivalent to T being represented in an orthonormal basis by a matrix A satisfying $A = A^*$. Such a matrix was also called Hermitian.

Theorem 10.2.2. *Let T be a self-adjoint operator. Then:*

(1) Every eigenvalue of T is a real number.

(2) Let $\lambda \neq \mu$ be two eigenvalues of T then

$$E_\lambda \perp E_\mu.$$

Proof. We begin with the first claim. Suppose that $Tv = \lambda v$ for some vector $v \neq 0$. Then $\langle Tv, v\rangle = \langle \lambda v, v\rangle = \lambda \|v\|^2$. On the other hand, $\langle Tv, v\rangle = \langle v, T^*v\rangle = \langle v, Tv\rangle = \langle v, \lambda v\rangle = \bar{\lambda}\|v\|^2$. It follows that $\lambda = \bar{\lambda}$.

Now for the second part. Let $v \in E_\lambda, w \in E_\mu$. We need to show $v \perp w$. We have $\langle Tv, w\rangle = \lambda\langle v, w\rangle$ and also $\langle Tv, w\rangle = \langle v, T^*w\rangle = \langle v, Tw\rangle = \langle v, \mu w\rangle = \mu\langle v, w\rangle$ (we have already established that μ is real). It follows that $(\lambda - \mu)\langle v, w\rangle = 0$ and so that $\langle v, w\rangle = 0$. □

Theorem 10.2.3. *Let T be a self-adjoint operator. There exists an orthonormal basis B such that $[T]_B$ is diagonal.*

Proof. The proof is by induction on $\dim(V)$. The cases $\dim(V) = 0, 1$ are obvious. Assume that $\dim(V) > 1$.

Let λ_1 be an eigenvalue of T. By definition, there is a corresponding non-zero vector v_1 such that $Tv_1 = \lambda_1 v_1$ and we may assume that $\|v_1\| = 1$. Let $W = \mathrm{Span}(v_1)$.

Lemma 10.2.4. *Both W and $W^\perp$ are T-invariant.*

Proof of lemma. This is clear for W since v_1 is an eigenvector. Suppose that $w \in W^\perp$. Then $\langle Tw, \alpha v_1\rangle = \langle w, T^*\alpha v_1\rangle = \langle w, T\alpha v_1\rangle = \lambda_1\langle w, \alpha v_1\rangle = 0$. This shows that Tw is orthogonal to W and so is in $W^\perp$ □

We can therefore decompose V and T accordingly:

$$V = W \oplus W^\perp, \qquad T = T_1 \oplus T_2.$$

Lemma 10.2.5. *Both T_1 and T_2 are self adjoint.*

Proof. T_1 is just multiplication by the real scalar λ_1, hence is self-adjoint. Let w_1, w_2 be in $W^\perp$. Then: $\langle Tw_1, w_2\rangle = \langle w_1, Tw_2\rangle$. Since $W^\perp$ is T-invariant, Tw_i is just $T_2 w_i$ and we get $\langle T_2 w_1, w_2\rangle = \langle w_1, T_2 w_2\rangle$, showing T_2 is self-adjoint. □

We may therefore apply the induction hypothesis to T_2; there is an orthonormal basis B_2 of $W^\perp$, say $B_2 = \{v_2, \ldots, v_n\}$, such that $[T]_{B_2}$ is diagonal, say $\mathrm{diag}(\lambda_2, \ldots, \lambda_n)$. Let

$$B = \{v_1\} \cup B_2.$$

Then B is an orthonormal basis for V and $[T]_B = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. □

Corollary 10.2.6. *Let $T : V \to V$ be a self adjoint operator whose distinct eigenvalues are $\lambda_1, \ldots, \lambda_r$. Then*

$$V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_r}.$$

Choose an orthonormal basis B^i for E_{λ_i}. Then

$$B = B^1 \cup B^2 \cup \cdots \cup B^r,$$

is an orthonormal basis for V and in this basis

$$[T]_B = \mathrm{diag}(\lambda_1, \ldots, \lambda_1, \ldots, \lambda_r, \ldots, \lambda_r).$$

Proof. Such a decomposition exists because T is diagonalizable. For any bases B^i we get that $[T]_B$ is diagonal of the form given above. If the B^i are orthonormal then so is B, because $E_{\lambda_i} \perp E_{\lambda_j}$ for $i < j$. □

Definition 10.2.7. A complex matrix M is called **unitary** if $M^* = M^{-1}$. If M is real and unitary we call it **orthogonal**.

To say M is unitary is the same as saying that its columns form an orthonormal basis. We therefore rephrase Corollary 10.2.6. (Note that the point is also that we may write M^* instead of M^{-1}. The usefulness of that will become clearer when we deal with bilinear forms.)

Corollary 10.2.8. *Let A be a self-adjoint matrix, $A = A^*$. Then, there is a unitary matrix M such that ${}^t\overline{M}AM$ is diagonal.*

It is worth noting a special case.

Corollary 10.2.9. *Let A be a real symmetric matrix. Then there is an orthogonal matrix P such that ${}^tPAP$ is diagonal.*

Example 10.2.10. Consider the symmetric matrix

$$A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix},$$

whose characteristic polynomial is $(t-4)(t-1)^2$. One finds

$$E_4 = \mathrm{Span}((1,1,1)),$$

with orthonormal basis given by

$$\frac{1}{\sqrt{3}}(1,1,1),$$

and

$$E_1 = \mathrm{Span}(\{(1,-1,0),(0,1,-1)\}).$$

We get an orthonormal basis for E_1 by applying Gram-Schmidt: $v_1 = \frac{1}{\sqrt{2}}(1,-1,0)$, $s_2' = (0,1,-1) - \langle (0,1,-1), \frac{1}{\sqrt{2}}(1,-1,0)\rangle \cdot \frac{1}{\sqrt{2}}(1,-1,0) = (0,1,-1) + \frac{1}{2}(1,-1,0) = \frac{1}{2}(1,1,-2)$. An orthonormal basis for E_1 is given by

$$\left\{ \frac{1}{\sqrt{2}}(1,-1,0), \frac{1}{\sqrt{6}}(1,1,-2) \right\}.$$

The matrix

$$P = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \end{pmatrix}$$

is orthogonal, that is ${}^tP \cdot P = I_3$, and ${}^tPAP = \mathrm{diag}(4,1,1)$.

10.3. **Application to symmetric bilinear forms.** To simplify the exposition, we assume that V is an n-dimensional vector space over the real numbers $\mathbb{R}$, even though the basic definitions and results hold in general.

Definition 10.3.1. A **bilinear form** on V is a bilinear symmetric function

$$[\cdot,\cdot] : V \times V \to \mathbb{R}.$$

That is, for all $v_1, v_2, v, w \in V$ and scalars a_1, a_2:

(1) $[a_1v_2 + a_2v_2, w] = a_1[v_1, w] + a_2[v_2, w]$;
(2) $[v, w] = [w, v]$.

Note that there is no requirement of positivity such as $[v, v] > 0$ for $v \neq 0$ (and that would also make no sense if we wish to extend the definition to a general field). However, the same arguments as in the case of inner products allow one to conclude that the following example is really the most general case.

Example 10.3.2. Let C be any basis for V and let $A \in M_n(\mathbb{R})$ be a real symmetric matrix. Then

$$[v, w] = {}^t[v]_C A[w]_C,$$

is a bilinear form.

Suppose that we change a basis. Let B be another basis and let $M = {}_CM_B$ be the transition matrix. Then, in the basis B the bilinear form is represented by

$${}^tMAM,$$

because ${}^t[v]_B{}^tMAM[w]_B = {}^t(M[v]_B)AM[w]_B = {}^t[v]_CA[w]_C$. Since we can find an orthogonal matrix M such that ${}^tMAM$ is diagonal, we conclude the following:

Proposition 10.3.3 (Principal Axis Theorem)**.** *Let A be a real symmetric matrix representing a symmetric bilinear form $[\cdot,\cdot]$ in some basis of V. Then, there is an orthonormal basis B with respect to which the bilinear form is given by*

$${}^t[v]_B \operatorname{diag}(\lambda_1, \ldots, \lambda_n)\ [w]_B = \sum_{i=1}^{n} \lambda_i v_i w_i,$$

where

$$[v]_B = (v_1, \ldots, v_n), \quad [w]_B = (w_1, \ldots, w_n).$$

Moreover, the diagonal elements $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, each appearing with the same multiplicity as in the characteristic polynomial of A.

Example 10.3.4. Consider the bilinear form given by the matrix

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

This is the function

$$[(x_1, x_2, x_3), (y_1, y_2, y_3)] = 2x_1y_1 + 2x_2y_2 + 2x_3y_3 + x_1y_2 + x_2y_1 + x_1y_3 + x_3y_1 + x_2y_3 + x_3y_2.$$

There is another orthogonal coordinate system (see Example 10.2.10), given by the columns of

$$P = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \end{pmatrix}$$

in which the bilinear form is given by

$$\begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Namely, in this coordinate system the bilinear form is just the function

$$[(x_1, x_2, x_3), (y_1, y_2, y_3)] = 4x_1y_1 + x_2y_2 + x_3y_3;$$

a more palatable formula than the original one!

10.4. **Application to inner products.** Recall that a matrix $M \in M_n(\mathbb{F})$, $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, defines a function

$$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle = (x_1, \ldots, x_n)\ M\ \overline{(y_1, \ldots, y_n)}^t,$$

which is an inner product if and only if M is Hermitian, that is $M = M^*$ (which we also call now "self-adjoint") and for every non-zero vector $(x_1, \ldots, x_n)$, $\langle (x_1, \ldots, x_n), (x_1, \ldots, x_n) \rangle$ is a positive real number. We called the last property "positive-definite".

Theorem 10.4.1. *Let M be a Hermitian matrix then M is positive-definite if and only if every eigenvalue of M is positive.*

Proof. We claim that M is positive definite if and only if AMA^* is positive definite for some (any) invertible matrix A. Note the formula $(A^*)^{-1} = (A^{-1})^*$ obtained by taking $*$ of both sides of $AA^{-1} = I_n$. Using it one sees that it is enough to show one direction. Suppose M is positive definite. Given a vector v we write $v = {}^tA^{-1}w$ then

$${}^tvAMA^*\bar{v} = {}^t({}^tAv)M\overline{{}^tAv} = {}^twM\overline{w}.$$

If $v \neq 0$ then $w \neq 0$ and then ${}^twM\overline{w} > 0$. This shows AMA^* is positive definite.

We now choose A to a unitary matrix such that AMA^* is diagonal, say $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, and the λ_i are the eigenvalues of M (because $AMA^* = AMA^{-1}$). This is possible by Corollary 10.2.8. It is enough to prove then that a diagonal real matrix is positive definite if and only if all the diagonal entries are positive. The necessity is clear, since ${}^te_i\ \mathrm{diag}(\lambda_1, \ldots, \lambda_n)\ e_i = \lambda_i$ should be positive. Conversely, if each λ_i is positive, for any non-zero vector $(x_1, \ldots, x_n)$ we have

$$(x_1, \ldots, x_n)\ \mathrm{diag}(\lambda_1, \ldots, \lambda_n)\ \overline{(x_1, \ldots, x_n)}^t = \sum_{i=1}^{n} \lambda_i |x_i|^2 > 0.$$

□

Example 10.4.2. Consider the case of a two-by-two matrix Hermitian matrix

$$M = \begin{pmatrix} a & b \\ \bar{b} & d \end{pmatrix}.$$

The characteristic polynomial is $t^2 - (a + d)t + (ad - b\bar{b})$.

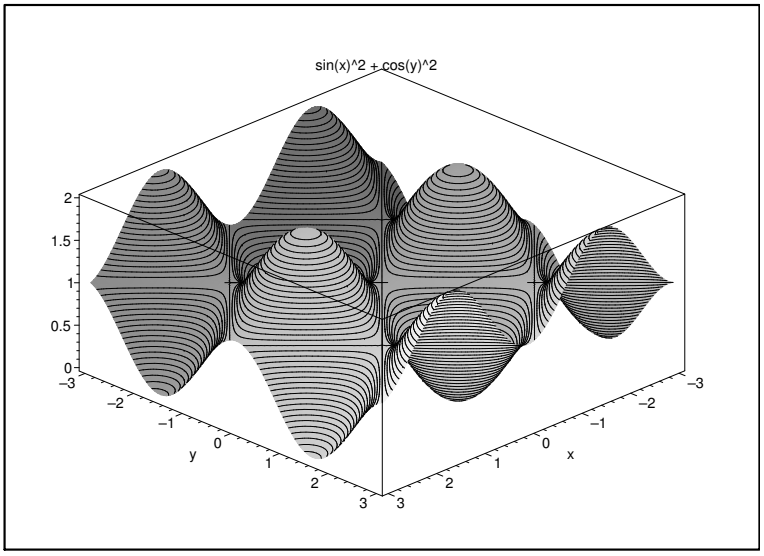
Now, two real numbers α, β are positive if and only if $\alpha + \beta$ and $\alpha\beta$ are positive. Namely, the eigenvalues of M are positive if and only if $\mathrm{Tr}(M) > 0, \det(M) > 0$. That is, if and only if $a + d > 0, ad - b\bar{b} > 0$, which is equivalent to $a > 0, d > 0$ and $ad - b\bar{b} > 0$.

10.4.1. *Extremum of functions of several variables.* Symmetric bilinear forms are of great importance everywhere in mathematics. For instance, given a twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, the local extremum points of f are points $\alpha = (\alpha_1, \ldots, \alpha_n)$ where the **gradient** $\left(\frac{\partial f}{\partial x_1}(\alpha), \ldots, \frac{\partial f}{\partial x_n}(\alpha)\right)$ vanishes. At these points one construct the **Hessian matrix**

$$\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\alpha)\right) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\alpha) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\alpha) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\alpha) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\alpha) \end{pmatrix},$$

which is symmetric by a fundamental result about functions in several variables. The function has a local minimum (maximum) iff and only if the Hessian is a positive definite (resp. minus the Hessian is positive definite). See also the assignments. We illustrate that for one pretty function.

Consider the function $f(x, y) = \sin(x)^2 + \cos(y)^2$. The gradient vanishes at points where



$\sin(x) = 0$ or $\cos(x) = 0$ and also $\sin(y) = 0$ or $\cos(y) = 0$. Namely, at points of the form $\{(x, y) : x, y \in \frac{\pi}{2}\mathbb{Z}\}$. The Hessian is

$$\begin{pmatrix} 2(1 - 2\sin(x)^2) & 0 \\ 0 & 2(1 - 2\cos(y)^2) \end{pmatrix}.$$

This matrix is positive definite at an extremum point $\{(x, y) : x, y \in \frac{\pi}{2}\mathbb{Z}\}$ iff $x \in \pi\mathbb{Z}$ and $y \in \pi(\mathbb{Z} + 1/2)$; those are the minima of the function. Similarly, we get maxima at the points $x \in \pi(\mathbb{Z} + 1/2)$ and $y \in \pi\mathbb{Z}$. The rest of the points are saddle points.

10.4.2. *Classification of quadrics.* Consider an equation of the form

$$ax^2 + bxy + cy^2 = N,$$

where N is some positive constant. What is the shape of the curve in $\mathbb{R}^2$, called a **quadric**, consisting of the solutions for this equation?

We can view this equation in the following form:

$$(x, y) \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = N.$$

Let

$$A = \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}.$$

We assume for simplicity that A is non-singular, i.e., $\det(A) \neq 0$. We can pass to another orthonormal coordinate system (the principal axis) such that the equation is written as

$$\lambda_1 x^2 + \lambda_2 y^2 = N$$

in the new coordinates. Here λ_i are the eigenvalues of A. Clearly, if both eigenvalues are positive we get an ellipse, if both are negative we get the empty set, and if one is negative and the other is positive then we get a hyperbole. We have

$$\lambda_1 + \lambda_2 = \mathrm{Tr}(A), \qquad \lambda_1 \lambda_2 = \det(A).$$

The case where λ_1, λ_2 are positive (negative) corresponds to $\mathrm{Tr}(A) > 0, \det(A) > 0$ (resp. $\mathrm{Tr}(A) < 0, \det(A) > 0$). The case of mixed signs is when $\det(A) < 0$.

Proposition 10.4.3. *Let*

$$A = \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}.$$

The curve defined by

$$ax^2 + bxy + cy^2 = N,$$

is an:

- *ellipse, if* $\mathrm{Tr}(A) > 0, \det(A) > 0$*;*
- *hyperbole, if* $\det(A) < 0$*;*
- *empty, if* $\mathrm{Tr}(A) < 0, \det(A) > 0$*.*

10.5. **Normal operators.** The normal operators are a much larger class than the self-adjoint operators. We shall see that we have a good structure theorem for this wider class of operators.

Definition 10.5.1. A linear map $T : V \to V$ is called **normal** if

$$TT^* = T^*T.$$

Example 10.5.2. Here are two classes of normal operators:

- The self adjoint operators. Those are the transformations T such that $T = T^*$. In this case $TT^* = T^2 = T^*T$.
- The unitary operators. Those are the transformations T such that $T^* = T^{-1}$. In this case, $TT^* = \mathrm{Id} = T^*T$.

Suppose that S is self-adjoint and U is unitary and, moreover, $SU = US$. Let $T = SU$. Then T is normal since $TT^* = SUU^*S^* = SS^* = S^*S$ and $T^*T = U^*S^*SU = S^*U^*US = S^*S$, where we have also used that if U and S commute so do U^* and S^*.

In fact, one can prove that any normal operator T can be written as SU, where S is self-adjoint, U is unitary, and $SU = US$.

Our goal is to prove orthonormal diagonalization for normal operators. We first proves some lemmas needed for the proof.

Lemma 10.5.3. *Let T be a linear operator and $U \subseteq V$ a T-invariant subspace. Then $U^\perp$ is T^*-invariant.*

Proof. Indeed, $v \in U^\perp$ iff $\langle u, v\rangle = 0, \forall u \in U$. Now, for every $u \in U$ and $v \in U^\perp$ we have

$$\langle u, T^*v\rangle = \langle Tu, v\rangle = 0,$$

because $Tu \in U$ as well. That shows $T^*v \in U^\perp$. □

Lemma 10.5.4. *Let T be a normal operator. Let v be an eigenvector for T with eigenvalue λ. Then v is also an eigenvector for T^* with eigenvalue $\bar{\lambda}$.*

Proof. We have

$$\begin{aligned}\langle T^*v - \bar{\lambda}v, T^*v - \bar{\lambda}v\rangle &= \langle (T-\lambda\cdot\mathrm{Id})^*v, (T-\lambda\cdot\mathrm{Id})^*v\rangle \\ &= \langle v, (T-\lambda\cdot\mathrm{Id})(T-\lambda\cdot\mathrm{Id})^*v\rangle \\ &= \langle v, (T-\lambda\cdot\mathrm{Id})^*\underbrace{(T-\lambda\cdot\mathrm{Id})v}_{0}\rangle \\ &= 0.\end{aligned} \tag{11}$$

(We have used the identity $(T-\lambda\cdot\mathrm{Id})(T-\lambda\cdot\mathrm{Id})^* = (T-\lambda\cdot\mathrm{Id})^*(T-\lambda\cdot\mathrm{Id})$, which is easily verified by expanding both sides and using $TT^* = T^*T$.) It follows that $T^*v - \bar{\lambda}v = 0$ and the lemma follows. □

Theorem 10.5.5. *Let $T : V \to V$ be a normal operator. Then there is an orthonormal basis B for V such that*

$$[T]_B = \mathrm{diag}(\lambda_1, \ldots, \lambda_n).$$

Proof. We prove that by induction on $n = \dim(V)$; the proof is very similar to the proof of Theorem 10.2.3. The theorem is clearly true for $n = 1$. Consider the case $n > 1$. Let v be a non-zero eigenvector, of norm 1, $U = \mathrm{Span}(v)$. Then U is T invariant, but also T^* invariant, because v is also a T^*-eigenvector by Lemma 10.5.4. Therefore, $U^\perp$ is T-invariant and T^*-invariant by Lemma 10.5.3 and clearly $T|_{U^\perp}$ is normal. By induction there is an orthonormal basis B' for $U^\perp$ in which T is diagonal. Then $B = \{v\} \cup B'$ is an orthonormal basis for V in which T is diagonal. □

Theorem 10.5.6 (The Spectral Theorem)**.** *Let $T : V \to V$ be a normal operator. Then*

$$T = \lambda_1\epsilon_1 + \cdots + \lambda_r\epsilon_r,$$

where $\lambda_1, \ldots, \lambda_r$ are the eigenvalues of T and the ϵ_i are orthogonal projections[7] such that

$$\epsilon_i \perp \epsilon_j,\ i \neq j, \qquad \mathrm{Id} = \epsilon_1 + \cdots + \epsilon_r.$$

Proof. We first prove the following lemma.

Lemma 10.5.7. *Let $\lambda \neq \mu$ be eigenvalues of T. Then*

$$E_\lambda \perp E_\mu.$$

Proof. Let $v \in E_\lambda, w \in E_\mu$. On the one hand,

$$\langle Tv, w\rangle = \langle \lambda v, w\rangle = \lambda\langle v, w\rangle.$$

On the other hand,

$$\langle Tv, w\rangle = \langle v, T^*w\rangle = \langle v, \bar{\mu}w\rangle = \mu\langle v, w\rangle.$$

Since $\lambda \neq \mu$ it follows that $\langle v, w\rangle = 0$. □

[7] If R is a ring and $\epsilon_1, \ldots, \epsilon_r$ are elements such that $\epsilon_i\epsilon_j = \delta_{ij}\epsilon_i$ and $1 = \epsilon_1 + \cdots + \epsilon_r$, we call them **orthogonal idempotents**. This is the situation we have in the theorem for $R = \mathrm{End}(V)$.

Now, by Theorem 10.5.5, T is diagonalizable. Thus,

$$V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_r}.$$

Let

$$\epsilon_i : V \to E_{\lambda_i}$$

be the projection. The Lemma says that the eigenspaces are orthogonal to each other and that implies that $\epsilon_i \epsilon_j = 0$ for $i \neq j$. The identity, $\mathrm{Id} = \epsilon_1 + \cdots + \epsilon_r$, is just a restatement of the decomposition $V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_r}$. □

10.6. **The unitary and orthogonal groups.** (Time allowing)

11. Appendix: Zorn's Lemma

A set S is called **partially ordered set**, or a **poset** for short, if it is equipped with a relation $\leq$. The relation does not need to be defined for any two elements. We use the notation $x \leq y$. This relation is required to satisfy the following properties

- $x \leq x$ for all $x \in X$.
- $x \leq y$ and $y \leq x$ implies $x = y$.
- $x \leq y$ and $y \leq z$ implies $x \leq z$.

The simplest example is perhaps the set $\mathbb{R}$ of real numbers where $x \leq y$ is the usual relation. For another example, take as a set $\mathbb{N}$ of natural numbers and say that $m \leq n$ if $m|n$. Note that the same definition doesn't work for $\mathbb{Z}$ as $2|-2$ and $-2|2$, for example, but we cannot conclude that $2 = -2$. Yet another example, is a set A and where S is the set of all subsets of A. We say for two subsets U, V that $U \leq V$ if $U \subset V$.

Let S then be a poset. Let $T \subseteq S$ be a subset. We say that an element $s \in S$ is an **upper bound** of T if for all $t \in T$ we have $t \leq s$. In general an upper bound need not exist. For example, in $\mathbb{R}$ the set $\mathbb{R}$ itself doesn't have an upper bound. A **chain** C in S is a subset of S such that for any two elements x, y in C we have either $x \leq y$ or $y \leq x$. For example $\{1, 2, 6, \dots n!, \dots\}$ is a chain in $\mathbb{N}$ as is $\{2, 4, 8, \dots, 2^n, \dots\}$.

Lemma 11.0.1 (Zorn's Lemma)**.** *Let S be a non-empty poset such that every non-empty chain in S has an upper bound. Then S has a maximal element. Namely, there is an element $s \in S$ such that $s \leq t$ implies $s = t$.*

Zorn's lemma is equivalent to the axiom of choice but the proof of Zorn's lemma from the axiom of choice is not easy and is best left to a course in set theory. Based on Zorn's lemma, we can draw the following corollary.

Corollary 11.0.2. *Every vector space has a basis.*

Proof. Let V be a vector space over a field $\mathbb{F}$. If $V = \{0\}$ the empty set is a basis.[8] and that matches the fact that V then "ought to" have dimension 0. Otherwise, V has some independent set. Indeed, if $v \in V$ is a non-zero vector the set $\{v\}$ is a linearly independent set.

Let S then be the set of all linearly independent subsets of V. It is a non-empty poset where we define for two subsets U, V that $U \leq V$ if $U \subset V$. We claim that every chain in S has an upper bound. Let C be a chain, say $C = \{U_i : i \in I\}$ such that each U_i is a linearly independent set and for each $i \neq j$ either $U_i \subset U_j$ or $U_j \subset U_i$. Note that this implies that for any finitely many subsets $U_1, \dots, U_n$ of C one of them, say U_{i_0} contains all the others (argue by induction on n).

In any case, let $U = \sup_{i \in I} U_i$ be the union of all the sets in the chain C. Note that we do not require the upper bound to belong to the chain. However, we claim that $U \in S$, namely, that U is a linearly independent set. Indeed, let $v_1, \dots, v_n$ be vectors in U. Then, from the definition of U, for each i there is some $U_i \in C$ such that $v_i \in U_i$. By the remark above, there is thus some $U_{i_0} \in C$ such that $v_1, \dots, v_n$ all belong to U_{i_0}. But U_{i_0} is a linearly independent set so if $\sum \alpha_i v_i = 0$ then every $\alpha_i = 0$. This shows that U is a linearly independent set and so $U \in C$. Clearly U is an upper bound for C.

The set S satisfies the conditions of Zorn's lemma and we thus conclude that it has a maximal element B. A moment reflection shows that B is a maximal linear independent set of V, i.e., a basis of V. □

A modification of the proof above yields the following stronger (and useful fact).

Theorem 11.0.3. *Let V be a vector space and $S \subset V$ a linearly independent set then S is contained in a basis of V.*

[8] If this causes you a splitting headache that's natural.

We leave the proof as an exercise.

Zorn's lemma is used in many proofs in algebra, but most of them require more advanced concepts. However, the following theorem is not too hard to prove and is left as an exercise.

Theorem 11.0.4. *Let R be a commutative non-zero ring then R has a maximal ideal.* [9]

[9] Or, more generally, any non-zero ring has a maximal left ideal and a maximal right ideal.

INDEX