LINEAR ALGEBRA. PART 1.

LECTURERS: DR K. KIRKINA AND DR E. ZORIN

Version: March 7, 2024.

Why study Linear Algebra?

Linear algebra has its origins in two areas of mathematics which appear extremely often in both practical and theoretical applications: systems of linear equations, and linear transformations. For example, systems of linear equations need to be solved in:

(i) logistics and network management (the costs of delivery of multiple things to multiple locations),
(ii) linear differential equations (for example, in biological or economic modelling, where they often arise as effective local approximations to nonlinear models).

Similarly, linear transformations are used in:

(i) video game display or more generally any 3d visualisation (rotations and reflections of objects or data sets are examples of linear transformations),
(ii) web search engines (link analysis),
(iii) product recommendations (e.g. Netflix),
(iv) Big Data processing for AI,
(v) quantum mechanics (the observables of a quantum mechanical system are the eigenvalues of linear differential transformations).

We study linear algebra not just because linearity appears naturally in theory and practice, but also because **nonlinearity** is so much harder and very often we seek to solve nonlinear problems by approximating them by linear problems (for example, the first derivative of a function is nothing other than the best linear approximation to it at the point of differentiation). A system of linear algebraic equations can **always** be solved explicitly, when a solution exists, by a simple algorithm (Gaussian elimination). An invertible linear transformation can often be understood by looking at its eigenvectors and their eigenvalues. By contrast, even simple nonlinear equations can be extremely difficult to solve explicitly and we often have to resort to finding approximate solutions.

Our aim in this module is to develop linear algebra by first looking at everything in terms of vectors (in $\mathbb{R}^n$ or $\mathbb{C}^n$) and matrices. Then we will see that all the language developed, and many of the results proved, also apply wherever "linearity" can be made sense of. In particular, when we replace coordinate vectors by other objects which can be added together and scalar multiplied (like functions, or even matrices). This is where Abstract Linear Algebra comes into play, allowing us to apply the methods of linear to a much wider range of situations than we started with initially.

The main message of this introduction is that: **linear algebra is used everywhere, and is relatively easy (with effort and practice)**. You will find it crops up in many places throughout the rest of your degree, so it well worth spending time now on becoming good at linear algebra.

1. Systems of Linear Equations

You have already met some linear algebra in the First Year of your degree. It is important that you re-familiarise yourself with this, as this module builds directly from that. These notes provide a small amount of overlap and reminder-giving, but you are still expected to be competent NOW with all the techniques taught last year.

The aim of this chapter is to completely understand the set of solutions to any system of p linear equations in n unknowns, and to do this in the case where the variables are real variables or complex variables. Let's start with an example.

Suppose we want to understand the solutions to the following system of two linear equations in four unknowns.

$$\begin{aligned} 3x_1 - x_2 + 2x_3 + x_4 &= 1, \\ -x_1 + x_2 + x_4 &= 0. \end{aligned}$$

We know we can write this in matrix form:

$$\begin{pmatrix} 3 & -1 & 2 & 1 \\ -1 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

We can use Gaussian elimination (i.e., row reduction) to express some variables in terms of others, and find the **general solution**. By the general solution we mean an expression which gives all possible solutions. By replacing row two (r_2) with $3r_2 + r_1$ we get

$$\begin{pmatrix} 3 & -1 & 2 & 1 \\ 0 & 2 & 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

We can now **eliminate** x_2 using the second row:

$$x_2 = \tfrac{1}{2} - x_3 - 2x_4.$$

Now substitute this into the equation given by the first row to **eliminate** x_1:

$$x_1 = \tfrac{1}{3}(\tfrac{1}{2} + x_2 - 2x_3 - x_4) = \tfrac{1}{3}(\tfrac{3}{2} - 3x_3 - 3x_4) = \tfrac{1}{2} - x_3 - x_4.$$

So we could write the general solution as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} - x_3 - x_4 \\ \frac{1}{2} - x_3 - 2x_4 \\ x_3 \\ x_4 \end{pmatrix}.$$

We could also write this as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -1 \\ -1 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -1 \\ -2 \\ 0 \\ 1 \end{pmatrix}.$$

This way of writing the general solution demonstrates several very important principles which are common to **all** linear systems of equations.

(i) There are two free variables, x_3 and x_4, in the solution. This number of free variables equals the number of original variables minus the number of independent equations: each equation eliminated one more variable.
(ii) The first vector is a particular solution to the original system.
(iii) The second two vectors are solutions to the **homogeneous system, the system for which the right hand side is the zero vector**:

$$\begin{pmatrix} 3 & -1 & 2 & 1 \\ 0 & 2 & 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

In fact the last two terms comprise the general solution to this homogeneous system.

Our aim in Chapter 1 is to demonstrate that this form of the general solution is common to all systems of linear equations. To do this we need some new concepts and terminology.

We can write a general system of p linear equations in n unknowns in the form

$$\begin{aligned} A_{11}x_1 + \ldots + A_{1n}x_n = &\quad y_1 \\ \vdots &\quad \vdots \\ A_{p1}x_1 + \ldots + A_{pn}x_n = &\quad y_p, \end{aligned} \tag{1.1}$$

where the coefficients A_{jk} can be either real or complex numbers. Note that our interest is in solving for $x_1, \ldots, x_n$ under the assumption that $y_1, \ldots, y_p$ are known (and A_{jk} are given).

Note that some systems of the form (1.1) have no solutions at all. For example, consider the following simple system of linear equations:

$$\begin{cases} x_1 + x_2 &= 0, \\ x_1 + x_2 &= 1. \end{cases}$$

Whatever solution (x_1, x_2) to this system might have been, its existence would imply $0 = 1$.

At the same time, the general properties of systems of linear equations are the same regardless of whether the coefficients A_{jk} are real or complex numbers. In fact one of the beauties of linear algebra is that if the coefficients A_{jk} and the quantities y_k are all real numbers, then the solutions x_j are guaranteed to be real[1](indeed, if A_{jk} and y_k are all rational numbers then the solutions x_j will all be rational numbers) – of course, on the assumption that these solutions exist at all. In other words, linear equations can be solved inside the **field** in which the problem is specified. Recall that we usually use the symbols $\mathbb{R}$ and $\mathbb{C}$ to indicate the fields of real or complex numbers. To avoid having to fix a choice when this is irrelevant, we will simply use $\mathbb{F}$ to mean either $\mathbb{R}$ or $\mathbb{C}$. We refer to elements of $\mathbb{F}$ as **scalars** and call $\mathbb{F}$ the **field of scalars**.

Remark 1.1. A number system is a field if we stay inside it whenever we do addition, subtraction, multiplication and division, so the set of integers $\mathbb{Z}$ is not a field. It is easy to write down a systems of linear equations involving only integers but whose solutions are not integers (for example, $2x = 1$).

In matrix notation (1.1) is written

$$\begin{pmatrix} A_{11} & \dots & A_{1n} \\ \vdots & & \vdots \\ A_{p1} & \dots & A_{pn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix}, \tag{1.2}$$

which we simply write as $A\mathbf{x} = \mathbf{y}$. The idea here is to solve for $\mathbf{x} \in \mathbb{F}^n$ (the vector of unknowns) when $\mathbf{y} \in \mathbb{F}^p$ is known and A is the $p \times n$ matrix above. Since there might be more than one solution, we actually want to understand the **solution set** to the system (1.2), i.e., given $\mathbf{y}$ and A, find

$$S = \{\mathbf{x} \in \mathbb{F}^n : A\mathbf{x} = \mathbf{y}\}.$$

1.1. **Linear combinations and linear subspaces.** The first and most crucial property of linear systems is how the solution set S behaves for **homogeneous systems**, i.e., when $\mathbf{y} = \mathbf{0} = (0, \dots, 0)$, the **zero vector** in $\mathbb{F}^p$. To explain this we need the following concepts.

Definition 1.1. *Given vectors $v_1, \dots, v_q \in \mathbb{F}^n$ and scalars $\alpha_1, \dots, \alpha_q \in \mathbb{F}$ we call the sum*

$$\alpha_1 v_1 + \ldots + \alpha_q v_q = \sum_{j=1}^{q} \alpha_j v_j,$$

*a **linear combination** of $v_1, \dots, v_q$.*

This definition is relevant because we will show that **any linear combination of solutions to a homogenous linear system is again a solution**. We say the solution set is **closed under taking linear combinations**. The next definition gives this property a name and explains what it means.

Definition 1.2. *A subset $S \subset \mathbb{F}^n$ is called a **subspace** (or **linear subspace**) of $\mathbb{F}^n$ if it contains the zero vector $\mathbf{0} \in \mathbb{F}^n$ and it is closed under taking all possible linear combinations.*

[1]Compare this with nonlinear equations: the equation $x^2 = -1$ does not have any real solutions even though its coeffiecients and right hand side are real.

This means that given any vectors $v_1, \ldots, v_q \in S$ *and scalars* $\alpha_1, \ldots, \alpha_q \in \mathbb{F}$ *the linear combination*

$$\alpha_1 v_1 + \ldots + \alpha_q v_q$$

also belongs to S.

We say a subspace S *is a* ***proper subspace*** *if* $S \neq \mathbb{F}^n$.

Remark 1.2. Note that if S is non-empty the condition of being closed under linear combinations implies $\mathbf{0} \in S$. So another way of stating the definition is that S must be non-empty and contain all linear combinations of elements of S.

Remark 1.3. Because every linear combination is the result of successive linear combinations of just two vectors, i.e.,

$$\alpha_1 v_1 + \ldots \alpha_q v_q = (\ldots((\alpha_1 v_1 + \alpha_2 v_2) + \alpha_3 v_3) + \ldots + \alpha_q v_q)$$

to check that S is closed under taking linear combinations we only have to check

$$\alpha_1 v_1 + \alpha_2 v_2 \in S \quad \text{for all } v_1, v_2 \in S,\ \alpha_1, \alpha_2 \in \mathbb{F}.$$

Example 1.1.

(i) $\mathbb{F}^n$ is clearly a subspace of itself but (obviously) not a proper subspace. The set $\{\mathbf{0}\} \subset \mathbb{F}^n$ is a subspace, called the **trivial subspace**. It is the smallest subspace of $\mathbb{F}^n$.

(ii) The set $S = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 = 0\}$ is a subspace of $\mathbb{R}^2$. To show this, we use the remark above and suppose we have any two vectors in this set, call them

$$v_1 = (a_1, b_1),\ v_2 = (a_2, b_2),\ \text{i.e., } a_j + b_j = 0,\ \text{for } j = 1, 2.$$

Then for any scalars $\alpha_1, \alpha_2 \in \mathbb{R}$ we observe that

$$\alpha_1 v_1 + \alpha_2 v_2 = (\alpha_1 a_1 + \alpha_2 a_2, \alpha_1 b_1 + \alpha_2 b_2).$$

We claim that this vector belongs to S, i.e., the coordinates x_1, x_2 of this vector satisfy $x_1 + x_2 = 0$. This is, of course, pretty obvious

$$\alpha_1 a_1 + \alpha_2 a_2 + \alpha_1 b_1 + \alpha_2 b_2 = \alpha_1(a_1 + b_1) + \alpha_2(a_2 + b_2) = 0.$$

Notice that geometrically this is a straight line in the plane passing through the origin.

(iii) The set $S = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 = 1\}$ is not a subspace of $\mathbb{R}^2$. This is easy to check: it does not contain the zero vector $(0, 0)$. Geometrically this is a straight line in the plane but it does not pass through the origin.

(iv) The set $S = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 - x_2^2 = 0\}$ is not a subspace. Note that the zero vector $\mathbf{0} = (0, 0)$ does belong to S. But if we take $v_1 = (1, 1)$ and $v_2 = (1, -1)$, the linear combination $v_1 + v_2 = (2, 0)$ does not belong to S. Geometrically this set is a union of two straight lines, since $(x_1)^2 - (x_2)^2 = (x_1 - x_2)(x_1 + x_2)$, so this is zero if either $x_1 - x_2 = 0$ or $x_1 + x_2 = 0$.

The relevance of the concept of linear subspace to solution sets is the following theorem.

Theorem 1.3. *The solution set to any linear system of the form $A\mathbf{x} = \mathbf{y}$ of p equations in n unknowns is a linear subspace of $\mathbb{F}^n$ if and only if $\mathbf{y} = \mathbf{0}$ (i.e., if and only if the linear system is homogeneous),*

The proof is easy, but let's do it anyway.

Proof. First, suppose $\mathbf{y} = \mathbf{0}$ and $v_1, \dots, v_q$ are arbitrary solution vectors: $Av_j = \mathbf{0}$ for all j. Then for any scalars $\alpha_1, \dots, \alpha_q$ we have

$$\begin{aligned} A(\alpha_1 v_1 + \ldots + \alpha_q v_q) &= \alpha_1 A v_1 + \ldots + \alpha_q A v_q \\ &= \alpha_1 \mathbf{0} + \ldots + \alpha_q \mathbf{0} = \mathbf{0}. \end{aligned} \tag{1.3}$$

The first line of this equation is the key point here: matrix multiplication turns linear combinations into linear combinations. We will emphasize this again later when we talk about matrices as linear maps.

So we've shown that the solution set to a homogeneous linear system is a subspace. Now for the converse, we are claiming that if the solution set is a subspace the system must be homogeneous. This is easy: every subspace contains the zero vector $\mathbf{0}$ and $A\mathbf{0} = \mathbf{0}$, so $\mathbf{y} = \mathbf{0}$. □

Beware that we are using one symbol $\mathbf{0}$ to denote zero vectors of different sizes (i.e., we use the same notation for $\mathbf{0} \in \mathbb{F}^n$ and $\mathbf{0} \in \mathbb{F}^q$ even though when $n \neq q$ these are different vectors). This can cause confusion if you aren't paying careful attention, but it is common practice.

Example 1.2. Consider the matrix

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

If we choose $\mathbf{y} = (0,0) = \mathbf{0}$ then the linear system $A\mathbf{x} = \mathbf{y}$ is

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

i.e,

$$x_1 + x_2 = 0, \quad x_2 + x_3 = 0.$$

If we assume $\mathbb{F} = \mathbb{R}$ then each equation describes a plane in $\mathbb{R}^3$, passing through the origin, and the solution set is the intersection of these planes (since both equations hold there). This is the line along which $x_1 = -x_2 = x_3$, so

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_3 \\ -x_3 \\ x_3 \end{pmatrix} = x_3 \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

This gives us the vector equation of the line, using the parameter x_3.

Now consider the same equation but with a general right hand side:

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

The general solution can be written as

$$x_2 = y_2 - x_3, \quad x_1 = y_1 - x_2 = y_1 - y_2 + x_3,$$

so that

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} y_1 - y_2 + x_3 \\ y_2 - x_3 \\ x_3 \end{pmatrix} = \begin{pmatrix} y_1 - y_2 \\ y_2 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}. \tag{1.4}$$

This is a again a vector equation for a line, but it only passes through the origin (where $x_1 = x_2 = x_3 = 0$) when $y_1 = y_2 = 0$. This illustrates the previous theorem: for a linear system, its solution set contains the origin if and only if the system is homogeneous.

Our second theorem describes the solution set to every linear system, homogeneous or inhomogeneous.

Theorem 1.4. *Let $A\mathbf{x} = \mathbf{y}$ be a general linear system of p equations in n unknowns. Then the solution set S is either empty or has the form*

$$S = \{w + \mathbf{x}_0 : Aw = \mathbf{0},\ A\mathbf{x}_0 = \mathbf{y}\},$$

where $\mathbf{x}_0$ is an arbitrary choice of a single (also called "particular") solution to $A\mathbf{x} = \mathbf{y}$ and w ranges over all solutions to the homogeneous system.

The explanation comes from the fact that if $\mathbf{x} = u$ and $\mathbf{x} = v$ are solutions to the equation $A\mathbf{x} = \mathbf{y}$ then $A(u - v) = Au - Av = \mathbf{y} - \mathbf{y} = \mathbf{0}$. Hence if we fix $\mathbf{x}_0 = v$, then u has the form $u = (u - \mathbf{x}_0) + \mathbf{x}_0$ and $w = u - \mathbf{x}_0$ satisfies $Aw = \mathbf{0}$.

Example 1.3. The previous example illustrates this, through (1.4). The general solution has the form

$$\mathbf{x} = w + \mathbf{x}_0,$$

where

$$w = x_3 \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \quad \mathbf{x}_0 = \begin{pmatrix} y_1 - y_2 \\ y_2 \\ 0 \end{pmatrix}.$$

The particular solution $\mathbf{x}_0$ depends upon $\mathbf{y}$, while w is the general solution to the homogeneous system.

Example 1.4. Not every linear system has a solution for every $\mathbf{y}$. Consider $A\mathbf{x} = \mathbf{y}$ when

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

The linear equations are then

$$x_1 + x_2 + x_3 = 0, \quad x_3 = 0, \quad 0 = 1.$$

Since the third equation is never true, there are no solutions: the solution set is the empty set $\emptyset$.

You might think this is a pointless example: no-one would ever write down a system which required $0 = 1$. But consider

$$A = \begin{pmatrix} 1 & 1 & 1 \\ -1 & -1 & 0 \\ 1 & 1 & 2 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

It is not immediately obvious that this has no solutions, but in fact by Gaussian elimination (row reduction to echelon form) it has the same solution set as the first example.

1.2. **Linear independence, bases and dimension.** The previous theorem shows that understanding the solution set to homogeneous linear systems is the key to understanding all linear systems which have solutions. We know from Theorem 1.3 that solution sets to homogeneous systems are linear subspaces. Our aim now is to show that every linear subspace $S \subset \mathbb{F}^n$ can be constructed from linear combinations using only finitely many vectors in S. First we need a reminder of an idea introduced in first year Algebra.

Definition 1.5. *A collection of vectors $v_1, \ldots, v_q \in \mathbb{F}^n$ is **linearly dependent** if there exists scalars $\alpha_1, \ldots, \alpha_q \in \mathbb{F}$, not all zero, for which*

$$\alpha_1 v_1 + \ldots + \alpha_q v_q = \mathbf{0}. \tag{1.5}$$

*Otherwise we say $v_1, \ldots, v_q$ are **linearly independent**.*

Equally, $v_1, \ldots, v_q$ are linearly independent if $\alpha_1 v_1 + \ldots + \alpha_q v_q = \mathbf{0}$ implies $\alpha_1, \ldots, \alpha_q$ are all zero.

The idea here is that linear dependence means one of these vectors can be written as a linear combination of the others. Specifically, if $\alpha_1 \neq 0$ the we can rearrange (1.5) to get

$$v_1 = -\frac{1}{\alpha_1}(\alpha_2 v_2 + \ldots + \alpha_q vq).$$

On the other hand, if (1.5) only holds when all the α_j's are zero, we cannot write any v_j as a linear combination of the others, so all are independent of each other.

Remark 1.4. Any collection containing the zero vector is linearly dependent, for if (say) $v_1 = \mathbf{0}$ then taking $\alpha_1 = 1$ and $\alpha_j = 0$ for $j \neq 1$ satisfies (1.5) with a non-zero α_1.

Example 1.5.

(i) Any one element set $\{v\} \subset \mathbb{F}^n$ with $v \neq 0$ is a linearly independent set, for the only way that $\alpha v = \mathbf{0}$ can hold is when $\alpha = 0$.

(ii) In $\mathbb{F}^n$ let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ denote the vectors

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots \quad \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

These are linearly independent, since

$$\alpha_1 \mathbf{e}_1 + \ldots + \alpha_n \mathbf{e}_n = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

Clearly this equals $\mathbf{0}$ if and only if $\alpha_j = 0$ for all $j = 1, \ldots, n$.

(iii) In $\mathbb{C}^2$ the vectors

$$u = \begin{pmatrix} i \\ 1 \end{pmatrix}, \quad v = \begin{pmatrix} -1 \\ i \end{pmatrix}$$

are linearly dependent because $iu - v = \mathbf{0}$. However, if we set

$$w = \begin{pmatrix} 1 \\ i \end{pmatrix}$$

then $\alpha_1 u + \alpha_2 w = 0$ if and only if both $i\alpha_1 + \alpha_2 = 0$ and $\alpha_1 + i\alpha_2 = 0$. In other words we need

$$\begin{pmatrix} i & 1 \\ 1 & i \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \mathbf{0}.$$

Using methods from last year (either substitution or calculating the determinant of the 2×2 matrix) we see that this holds if and only if $\alpha_1 = 0 = \alpha_2$. Hence u, w are linearly independent.

Given any collection of vectors, not necessarily finite, we can create a subspace by taking all linear combinations. This has a name.

Definition 1.6. *For any non-empty collection of vectors $\mathcal{C} \subset \mathbb{F}^n$ we call the set of all linear combinations of vectors from $\mathcal{C}$ the **span** of $\mathcal{C}$, denoted $\mathrm{Sp}(\mathcal{C})$. When $\mathcal{C}$ is a finite collection $v_1, \ldots, v_q$ we may also write $\mathrm{Sp}(v_1, \ldots, v_n)$. We define $\mathrm{Sp}(\emptyset) = \{\mathbf{0}\}$.*

You could try to prove the following lemma, as an exercise testing your understanding of the definitions above.

Lemma 1.7. *For any collection $\mathcal{C} \subset \mathbb{F}^n$, $\mathrm{Sp}(\mathcal{C})$ is a subspace of $\mathbb{F}^n$. In fact it is the smallest subspace which contains $\mathcal{C}$ (i.e., if $S \subset \mathbb{F}^n$ is any subspace with $\mathcal{C} \subset S$ then $\mathrm{Sp}(\mathcal{C}) \subset S$).*

For any subspace $S \subset \mathbb{F}^n$ we can say $\mathcal{C}$ **spans** S when $S = \mathrm{Sp}(\mathcal{C})$, or we call $\mathcal{C}$ a **spanning set** for S, or we can say S **is spanned by** $\mathcal{C}$.

Example 1.6. In $\mathbb{F}^3$ we have $\mathbf{e}_1 = (1,0,0)$ and $\mathbf{e}_2 = (0,1,0)$. For these

$$\begin{aligned}\mathrm{Sp}(\mathbf{e}_1,\mathbf{e}_2) &= \{\alpha(1,0,0)+\beta(0,1,0) : \alpha,\beta\in\mathbb{F}\}\\ &= \{(\alpha,\beta,0) : \alpha,\beta\in\mathbb{F}\}\\ &= \{(x_1,x_2,x_3) : x_3 = 0\}.\end{aligned}$$

Notice that this example works for either $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$.

Now consider $\mathrm{Sp}(\mathbf{e}_1,\mathbf{e}_2,u)$ where $u = (1,1,0)$. Since $u = \mathbf{e}_1 + \mathbf{e}_2$ it is clear that $\mathrm{Sp}(\mathbf{e}_1,\mathbf{e}_2,u) = \mathrm{Sp}(\mathbf{e}_1,\mathbf{e}_2)$. But since $\mathbf{e}_1 = u - \mathbf{e}_2$ and $\mathbf{e}_2 = u - \mathbf{e}_1$ it is also clear that

$$\mathrm{Sp}(\mathbf{e}_1,\mathbf{e}_2,u) = \mathrm{Sp}(\mathbf{e}_2,u) = \mathrm{Sp}(\mathbf{e}_1,u).$$

If we replace u by $v = (1,1,1)$, then $\mathbf{e}_3 = v - \mathbf{e}_1 - \mathbf{e}_2$ so

$$\mathrm{Sp}(\mathbf{e}_1,\mathbf{e}_2,v) = \mathrm{Sp}(\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3) = \mathbb{F}^3,$$

i.e., every vector in $\mathbb{F}^3$ is a linear combination of $\mathbf{e}_1,\mathbf{e}_2,v$.

Example 1.7. In $\mathbb{R}^2$ consider the vectors

$$v_1 = \begin{pmatrix}3\\-1\end{pmatrix}, \quad v_2 = \begin{pmatrix}-2\\3\end{pmatrix}, \quad v_3 = \begin{pmatrix}0\\1\end{pmatrix}.$$

Let us answer the questions (i) are these linearly independent, (ii) do they span $\mathbb{R}^2$? To answer (i) we work from the definition of linear independence and look at the equation

$$\alpha\begin{pmatrix}3\\-1\end{pmatrix} + \beta\begin{pmatrix}-2\\3\end{pmatrix} + \gamma\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}.$$

The vectors will be linearly independent if the only solution to this equation is $\alpha = \beta = \gamma = 0$. In matrix form this equation is

$$\begin{pmatrix}3 & -2 & 0\\-1 & 3 & 1\end{pmatrix}\begin{pmatrix}\alpha\\\beta\\\gamma\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}.$$

By Gaussian elimination (replacing the second row, r_2, by $3r_2 + r_1$) this system has the same solution set as

$$\begin{pmatrix}3 & -2 & 0\\0 & 7 & 3\end{pmatrix}\begin{pmatrix}\alpha\\\beta\\\gamma\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}.$$

So the general solution is

$$\beta = -\tfrac{3}{7}\gamma, \quad \alpha = \tfrac{2}{3}\beta = -\tfrac{2}{7}\gamma.$$

Since we can choose γ to be non-zero and get a solution, these vectors are not linearly independent. For example, if we take $\gamma = 1$ the solution above shows that

$$-\tfrac{2}{7}v_1 - \tfrac{3}{7}v_2 + v_3 = \mathbf{0}.$$

Now consider the second question: do these vectors span $\mathbb{R}^2$? From the defintion this is true when for every $(x, y) \in \mathbb{R}^2$ we can find $\alpha, \beta, \gamma \in \mathbb{R}$ with

$$\alpha \begin{pmatrix} 3 \\ -1 \end{pmatrix} + \beta \begin{pmatrix} -2 \\ 3 \end{pmatrix} + \gamma \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

This is again a system of equations for α, β, γ (think of x, y as fixed but unknown). In matrix form this is

$$\begin{pmatrix} 3 & -2 & 0 \\ -1 & 3 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

By Gaussian elimination this system has the same solution set as

$$\begin{pmatrix} 3 & -2 & 0 \\ 0 & 7 & 3 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} x \\ 3y + x \end{pmatrix}.$$

It is clear at this point that no matter what x, y we choose there are solutions α, β, γ (indeed, the solution set has one free parameter which we could choose to be γ, as above). So we conclude that every vector in $\mathbb{R}^2$ can be written as a linear combination of v_1, v_2, v_3 and therefore they do span $\mathbb{R}^2$.

Now we come to a very important definition.

Definition 1.8. *Given a non-trivial subspace $S \subset \mathbb{F}^n$ a collection $\mathcal{B} = \{v_1, \ldots, v_q\} \subset S$ of vectors in S is called a* ***basis of*** *S if it is both linearly independent and spans S. In other words $\mathcal{B}$ has the two properties:*

(i) *if $\alpha_1 v_1 + \ldots + \alpha_q v_q = \mathbf{0}$ then $\alpha_1, \ldots, \alpha_q$ are all zero,*
(ii) *for any $v \in S$ there exist scalars $\alpha_1, \ldots, \alpha_q$ for which $v = \alpha_1 v_1 + \ldots + \alpha_q v_q$.*

Remark 1.5. The trivial subspace $\{\mathbf{0}\}$ does play a regular role in linear algebra, for example as the solution set to any linear system $A\mathbf{x} = \mathbf{0}$ where A is an invertible square matrix, but the above definition doesn't allow us to describe a basis. For formal logical reasons (and to make some statements a bit simpler) it is consistent to say that the empty set $\emptyset$ is a basis for $\{\mathbf{0}\}$, and actually this fits in fairly well with subsequent results about bases. Note that while it is incorrect to say $\{\mathbf{0}\}$ has no basis, it is correct to say that a basis for $\{\mathbf{0}\}$ has no elements.

It should be clear that there is no such thing as a unique basis for a non-trivial vector space. But once a basis is fixed, with an ordering, it provides a unique way of representing any vector.

Lemma 1.9. *Let $S \subset \mathbb{F}^n$ be a subspace with an* ***ordered*** *basis $\mathcal{B} = (v_1, \ldots, v_q)$. Then each $u \in S$ can be uniquely written in the form $u = \alpha_1 v_1 + \cdots + \alpha_q v_q$ with $\alpha_i \in \mathbb{F}$, i.e., the α_i are unique to u. These are called the* ***coordinates of*** *u* ***with respect to the ordered basis*** *$\mathcal{B}$.*

The reason follows from the definition of linear independence. Suppose

$$u = \alpha_1 v_1 + \cdots + \alpha_q v_q = \beta_1 v_1 + \cdots + \beta_q v_q.$$

Then by subtraction

$$(\alpha_1 - \beta_1)v_1 + \ldots + (\alpha_q - \beta_q)v_q = \mathbf{0}.$$

So linear independence forces $\alpha_j = \beta_j$ for all j.

Example 1.8.

(i) $\mathbb{F}^n$ itself has a **standard basis** consisting of the vectors

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots \quad \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

We have seen above that these are linearly independent, and clearly every vector is a linear combination of these:

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = \alpha_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \ldots + \alpha_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = \sum_{j=1}^{n} \alpha_j \mathbf{e}_j.$$

The lemma above simply asserts that the vector $(\alpha_1, \ldots, \alpha_n)$ is uniquely determined by it coordinates $\alpha_1, \ldots, \alpha_n$.

(ii) Another example of a basis of $\mathbb{F}^n$ is to take

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \mathbf{e}_1 + \mathbf{e}_2 + \ldots + \mathbf{e}_n, \quad v_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \mathbf{e}_2 + \ldots + \mathbf{e}_n, \quad \ldots \quad v_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = \mathbf{e}_n.$$

It is not hard to see these are linearly independent. To write a general vector as a linear combination of this basis we need to find scalars $\beta_1, \ldots, \beta_n$ so that

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = \beta_1 \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \ldots + \beta_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

In other words, we have to solve the linear system

$$\begin{pmatrix} 1 & 0 & \ldots & 0 \\ 1 & 1 & \ldots & 0 \\ \vdots & & & \vdots \\ 1 & 1 & \ldots & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

Notice that the columns of the matrix are the vectors $v_1, \ldots, v_n$. It is not hard to show that the solution to this system is

$$\beta_1 = \alpha_1, \quad \beta_2 = \alpha_2 - \alpha_1, \quad \ldots \quad \beta_n = \alpha_n - \alpha_{n-1}.$$

These are the coordinates of the vector $u = (\alpha_1, \ldots, \alpha_n)$ with respect to the basis $v_1, \ldots, v_n$.

For practical reasons it is useful to know that one can always swap in a preferred vector to a basis. To be precise:

Lemma 1.10 (Steinitz exchange lemma)**.** *Given a non-trivial subspace $S \subset \mathbb{F}^n$ and a basis $\{v_1, \ldots, v_q\}$ for S, for any non-zero $u \in S$ there is some j for which swapping u for v_j produces another basis, i.e., $\{v_1, \ldots, v_{j-1}, u, v_{j+1}, \ldots, v_q\}$ is again a basis.*

Proof. Since $\{v_j : j = 1, \ldots, q\}$ is a basis for S we can write

$$u = \sum_{k=1}^{q} \alpha_k v_k \tag{1.6}$$

for some $\alpha_k \in \mathbb{F}$, $1 \le k \le q$. Now, since $u \neq \mathbf{0}$, there is a first nonzero coefficient, call it α_j. Then we can write

$$v_j = \alpha_j^{-1} u + \sum_{k \neq j} (-\alpha_j^{-1} \alpha_k) v_k.$$

Now $\mathcal{C} = \{v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n, u\}$ still spans S. To show linear independence, suppose

$$\mathbf{0} = \sum_{k \neq j} \beta_k v_k + \gamma u$$

for some $\beta_k, \gamma \in \mathbb{F}$ with $k \neq j$. Substituting for u from equation (1.6) we get

$$\mathbf{0} = \sum_{k \neq j} \beta_k v_k + \gamma \left(\sum_{k=1}^{q} \alpha_k v_k \right) = \sum_{k \neq j} (\beta_k + \gamma \alpha_k) v_k + \gamma \alpha_j v_j.$$

By linear independence of $\{v_k : k = 1, \ldots, v_q\}$, we conclude that $\beta_k + \gamma \alpha_k = 0$ for each $k \neq j$ and $\gamma \alpha_j = 0$. Since $\alpha_j \neq 0$ by our earlier choice, we must have $\gamma = 0$, and then we get $\beta_k = 0$ for all $k \neq j$ as well. Thus $\mathcal{C}$ is linearly independent. $\square$

The proof above have practical value. It shows that to swap u into the basis we first look at the expression

$$u = \alpha_1 v_1 + \ldots + \alpha_q v_q.$$

We can swap u with any v_j for which $\alpha_j \neq 0$.

The notion of a basis allows us to give a clear meaning to "dimension", at least for linear subspaces.

Theorem 1.11. *Every subspace $S \subset \mathbb{F}^n$ has a basis and every basis of S has the same finite number of elements. This number is called the **dimension of** S, denoted* $\dim(S)$.

Proof. Be begin by providing a method to construct a basis, and after that we will show that this method actually terminates in a finite number of steps.

We may assume S is non-trivial (since we already defined the basis for $\{\mathbf{0}\}$). In that case there is a non-zero vector $v_1 \in S$. If $\mathrm{Sp}(v_1) = S$ we are done, since v_1 is linearly independent and spans S. If $\mathrm{Sp}(v_1) \neq S$ then there exists another vector $v_2 \notin \mathrm{Sp}(v_1)$ which is therefore independent of v_1. Now proceed like this: at the k-th step we have linearly independent vectors $v_1, \dots, v_k$. If $S = \mathrm{Sp}(v_1, \dots, v_k)$ then $v_1, \dots, v_k$ is a basis. If not there exists $v_{k+1} \notin \mathrm{Sp}(v_1, \dots, v_k)$ and $v_1, \dots, v_{k+1}$ are linearly independent.

Now we proceed with a more tricky part: to show that the procedure described above stops after finitely many steps and we obtain a linearly independent collection $v_1, \dots, v_q$ for which $S = \mathrm{Sp}(v_1, \dots, v_q)$.

Suppose the procedure described above does not stop. Take the first n vectors in the sequence. Using the Steinitz Exchange Lemma repeatedly, we replace each element of the standard basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ by an element of $\{v_1, \dots, v_n\}$ to conclude that this is a basis. This contradicts the assumption that the sequence doesn't terminate, since that implies there is a linearly independent v_{n+1}.

The argument that we can successively apply Steinitz Exchange to always remove an $\mathbf{e}_j$ goes as follows. Suppose we have done the first exchange and swapped out $\mathbf{e}_j$ for v_1. Then we can write

$$v_2 = \alpha v_1 + \sum_{k \neq j} \alpha_k \mathbf{e}_k.$$

There must be some non-zero α_k, since otherwise $v_2 = \alpha v_1$ which would contradict linear independence of the v_j's. We take the first non-zero α_k and swap out $\mathbf{e}_k$ for v_2. Repeating this argument allows us to always swap a standard basis vector for one of the v_j's until all are replaced.

Once we have shown that the sequence is finite we obtain a basis $v_1, \dots, v_q$ for S (since there are no vectors left in S outside $\mathrm{Sp}(v_1, \dots, v_q)$). Any other basis must have the same number of elements using the exchange argument above. Notice that this proof shows that there are no more than n linearly independent vectors in any subspace S of $\mathbb{F}^n$, i.e., $\dim(S) \leq n$. $\square$

Example 1.9.

(i) The dimension of $\mathbb{F}^n$ is n. Note in particular that $\mathbb{C}$ has dimension one. Sometimes this is emphasized by saying it has complex dimension one. Students are used to thinking about the Argand plane representation of the complex numbers, which suggests $\mathbb{C}$ should have dimension two, but that is its real dimension (dimension as a vector space over the field $\mathbb{R}$ - this will be explained later in the course when we discuss the definition of a vector space over a field).

(ii) Consider the solution set S to the homogeneous linear system $x_1 + x_2 + x_3 = 0$. We claim that S is two-dimensional, spanned by $v_1 = (1, 0, -1)$ and $v_2 = (0, 1, -1)$. To see this, clearly $v_1, v_2 \in S$. Moreover they are linearly independent, since

$$\alpha_1(1,0,-1) + \alpha_2(0,1,-1) = (0,0,0) \;\Leftrightarrow\; (\alpha_1, \alpha_2, -\alpha_1 - \alpha_2) = (0,0,0) \;\Leftrightarrow\; \alpha_1 = 0 = \alpha_2.$$

Further, every solution has the form

$$(x_1, x_2, -x_1 - x_2) = x_1(1, 0, -1) + x_2(0, 1, -1),$$

so every solution belongs to $\mathrm{Sp}(v_1, v_2)$.

As a consequence of the proof of the existence of a basis, we derive a collection of very useful consequences.

Lemma 1.12. *Suppose S is a subspace of $\mathbb{F}^n$ of dimension q.*

(i) *Any linearly independent subset of S has no more than q elements. Hence, any subset of S containing more than q vectors is linearly dependent.*
(ii) *Any linearly independent subset $\mathcal{C} \subset \mathbb{F}^n$ can be extended to a basis of $\mathbb{F}^n$ (i.e., there is a basis $\mathcal{B}$ which contains $\mathcal{C}$). In particular, any basis of S can be extended to a basis of $\mathbb{F}^n$.*
(iii) *Any finite spanning set for S contains a basis. Hence no subset containing fewer than q vectors can span S.*
(iv) *Any linearly independent subset of S containing q vectors must span S (hence is a basis for S). Similarly, any set of size q which spans S must be linearly independent (hence is a basis for S).*
(v) *If $q = 0$ then $S = \{\mathbf{0}\}$. If $q = n$ then $S = \mathbb{F}^n$.*

Proof.

(i) If $v_1, \ldots, v_k$ are linearly independent in S then using the argument in the proof above we can complete them to make a basis for S, but this means $k \leq q$. Hence $k > q$ implies they are not independent.
(ii) This is just the argument above with $S = \mathbb{F}^n$.
(iii) Given a finite spanning set $\{v_1, \ldots, v_k\}$ either these are independent, hence a basis already, or we can remove one vector which depends on the others. This does not change the span (since the removed vector is in the span of the others). We keep doing this until we reach an independent set. The span has not changed, so this new set is a basis. It follows that if we have fewer than q vectors their span cannot equal S, otherwise they would contain a basis with fewer than $q = \dim(S)$ vectors.
(iv) Given $v_1, \ldots, v_q$ independent, if they do not span S then at least one more independent vector is required to span S, yielding a basis with more than q elements, which is a contradiction to Theorem 1.11. On the other hand if $v_1, \ldots, v_q$ span S and are dependent, then we can remove vectors as above to get a basis with fewer than q vectors, again a contradiction.
(v) The basis for $\{\mathbf{0}\}$ is $\emptyset$, which has zero elements. $\mathbb{F}^n$ has basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$.

□

Example 1.10.

(i) In Example 1.7 we found that those three vectors in $\mathbb{R}^2$ are not linearly independent. The Lemma above tells us that this is always the case ($S = \mathbb{R}^2$, so $q = 2$): we don't have to work this out the hard way. We also saw that the vectors spanned $\mathbb{R}^2$.

Looking at those vectors it is clear that any two of them are linearly independent and therefore by the Lemma above they must span $\mathbb{R}^2$.

(ii) The three vectors

$$\begin{pmatrix} i \\ 1 \end{pmatrix}, \begin{pmatrix} \sqrt{3}+2i \\ \sqrt{3}-12i \end{pmatrix}, \begin{pmatrix} e^{4i} \\ \pi \end{pmatrix}$$

in $\mathbb{C}^2$ must be linearly dependent, since $\mathbb{C}^2$ has dimension 2. Notice that this would be hard work to try to establish directly by finding a linear combination which sums to zero, so the lemma is saving us work.

(iii) The three vectors

$$\begin{pmatrix} i \\ 1 \\ \sqrt{3}-2i \\ 1 \end{pmatrix}, \begin{pmatrix} e^{4i} \\ \pi \\ 0 \\ 14 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -1 \\ 2 \end{pmatrix}$$

in $\mathbb{C}^4$ cannot be a basis for $\mathbb{C}^4$. Again, this would be hard work to prove directly.

(iv) The three vectors

$$(1, i, 0),\ (1, 0, 1),\ (0, i, 1)\ \in \mathbb{C}^3$$

are linearly independent (check this for yourself) and therefore form a basis for $\mathbb{C}^3$.

The punch-line from all this discussion of independence, span and bases is that we now have a very satisfying way of describing the solution set to any linear system.

Theorem 1.13. *Let $A\mathbf{x} = \mathbf{y}$ be any linear system of p equations in n unknowns. If its solution set is non-empty, then every solution has the form*

$$\mathbf{x} = \alpha_1 v_1 + \ldots + \alpha_q v_q + \mathbf{x}_0, \quad \alpha_1, \ldots, \alpha_q \in \mathbb{F}, \tag{1.7}$$

where $\{v_1, \ldots, v_q\}$ is a basis for the solution set of the homogeneous system $A\mathbf{x} = \mathbf{0}$ and $\mathbf{x}_0$ is a particular solution to $A\mathbf{x} = \mathbf{y}$.

The expression (1.7) is often referred to as the **general solution** to the system of equations. Notice that it has q "free parameters" $\alpha_1, \ldots, \alpha_q$: the dimension q of the solution set to the homogeneous system measures the degree of freedom in the general solution to the inhomogeneous system. In particular, the solution is unique if and only if $q = 0$, i.e., if and only if the solution to $A\mathbf{x} = \mathbf{0}$ is unique (i.e., $\mathbf{x} = \mathbf{0}$ is the only solution).

Note also that a different choice of basis or particular solution will give a general solution which looks different, but any two such expressions still describe every element of the solution set. This is illustrated in the next example.

Example 1.11. Consider the linear system

$$x_1 + x_2 + x_3 = 1, \quad x_2 - x_3 = 0,$$

i.e., determined by

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

This system clearly has a "particular" solution $(1,0,0)$, and the solution set to the homogeneous system $A\mathbf{x} = \mathbf{0}$ is easily seen to be spanned by the single vector $v_1 = (-2,1,1)$. Hence the general solution can be written in the form

$$(x_1, x_2, x_3) = \alpha(-2,1,1) + (1,0,0) = (1-2\alpha, \alpha, \alpha),$$

which has one degree of freedom (the choice of α). If we chose instead the particular solution $(-1,1,1)$ and replaced v_1 by $(2,-1,-1)$ we would get a different looking expression for the general solution:

$$(x_1, x_2, x_3) = \beta(2,-1,-1) + (-1,1,1) = (2\beta - 1, 1-\beta, 1-\beta).$$

1.3. **Sum and direct sum of subspaces.** We now want to consider ways of constructing subspaces from other subspaces. Suppose $S_1, S_2 \subset \mathbb{F}^n$ are subspaces. The natural operations on sets are intersection $S_1 \cap S_2$ and union $S_1 \cup S_2$. We can show that the intersection of subspaces is always another subspace, but this is rarely true for unions. For example, take

$$S_1 = \{(x_1, 0) : x_1 \in \mathbb{R}\}, \quad S_2 = \{(0, x_2) : x_2 \in \mathbb{R}\}.$$

Then $S_1 \cap S_2 = \{(0,0)\}$, which is the trivial subspace, but $S_1 \cup S_2$ is not a subspace because, for example, taking $(1,0) \in S_1$ and $(0,1) \in S_2$ we can make the linear combination $(1,1) = (1,0) + (0,1)$, which does not belong to $S_1 \cup S_2$. However, since the span of any subset of $\mathbb{F}^n$ is a subspace, we can create a subspace from the union.

Definition 1.14. *For subspaces $S_1, \ldots, S_q \subset \mathbb{F}^n$ we define their **sum**, denoted $S_1 + \ldots + S_q$ (or $\sum_{j=1}^q S_j$) to be the span of their union, i.e.,*

$$S_1 + \ldots + S_q = \mathrm{Sp}(S_1 \cup \ldots \cup S_q) = \{\alpha_1 v_1 + \ldots + \alpha_q v_q : \alpha_j \in \mathbb{F},\ v_j \in S_j\}.$$

*When $S_j \cap (\sum_{k \neq j} S_k) = \{\mathbf{0}\}$ for all $1 \leq j \leq q$ we call this the **direct sum**, denoted $S_1 \oplus \ldots \oplus S_q$ (or $\bigoplus_{j=1}^q S_j$).*

Theorem 1.15. *For any collection of subspaces $S_1, \ldots, S_q \subset \mathbb{F}^n$ both their common intersection $S_1 \cap \ldots \cap S_q$ and their sum (or direct sum) $S_1 + \ldots + S_q$ are also subspaces.*

This is quite easy to prove. By Lemma 1.7 any span is a subspace, so this proves it for the sum. For the intersection, every subspace contains $\mathbf{0}$ so this belongs to the intersection. Also, if we take any vectors in the intersection then they belong to every S_j so any linear combination of them also belongs to every S_j, hence it belongs to the intersection.

Example 1.12.

(i) If $S_1 \subset \mathbb{F}^n$ is the subspace of solutions to the linear equation $\sum_{j=1}^n \alpha_j x_j = 0$ and S_2 is the subspace of solutions to the linear equation $\sum_{j=1}^n \beta_j x_j = 0$, then clearly their intersection $S_1 \cap S_2$ is the subspace of vectors $\mathbf{x} \in \mathbb{F}^n$ which satisfy both equations. Hence every solution set for a system of p equations in n unknowns is the intersection of p subspaces of $\mathbb{F}^n$.

(ii) A simple example of direct sum is the following. In $\mathbb{F}^3$ recall the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. Let $S_j = \mathrm{Sp}(\mathbf{e}_j)$: each of these is a line we call an axis when

$\mathbb{F} = \mathbb{R}$. Notice that $S_j \cap S_k = \{\mathbf{0}\}$ for $j \neq k$. Now $S_1 \oplus S_2 = \mathrm{Sp}(\mathbf{e}_1, \mathbf{e}_2)$ is the plane defined by $x_3 = 0$.

(iii) Here is a simple example of a sum of subspaces which is not a direct sum. In $\mathbb{R}^3$ take $V_1 = \mathrm{Sp}(\mathbf{e}_1, \mathbf{e}_2)$ and $V_2 = \mathrm{Sp}(\mathbf{e}_2, \mathbf{e}_3)$. Then $V_1 + V_2 = \mathrm{Sp}(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = \mathbb{R}^3$ but $V_1 \cap V_2 = \mathrm{Sp}(\mathbf{e}_2)$ so the sum is not direct.

The following fact is useful for computing the dimension of subspace sums.

Lemma 1.16. *Let S_1, S_2 be subspaces of $\mathbb{F}^n$. Then*

$$\dim(S_1 + S_2) = \dim(S_1) + \dim(S_2) - \dim(S_1 \cap S_2).$$

In particular, for a direct sum $\dim(S_1 \oplus S_2) = \dim(S_1) + \dim(S_2)$.

Example 1.13. In part (iii) of the previous example, this equation reads:

$$\dim(\mathbb{R}^3) = \dim(V_1) + \dim(V_2) - \dim(\mathrm{Sp}(\mathbf{e}_2)), \text{ i.e., } 3 = 2 + 2 - 1.$$

One of the main reasons for being interested in direct sums is expressed in the following result.

Lemma 1.17. *Let $S_1 \oplus \ldots \oplus S_q$ be a direct sum of subpaces and $v_j \in S_j$ be non-zero for $j = 1, \ldots q$. Then $v_1, \ldots, v_q$ are linearly independent.*

Proof. Suppose $\sum_{k=1}^{q} \alpha_k v_k = \mathbf{0}$. Then for each j

$$\alpha_j v_j = -\sum_{k \neq j} \alpha_k v_k \in S_j \cap (\sum_{k \neq j} S_k).$$

By the definition of direct sum this intersection is the trivial subspace, so $\alpha_j v_j = \mathbf{0}$. Hence each α_j must be zero. $\square$

2. Matrices and Linear Maps

2.1. **Linear maps.** A $p \times n$ matrix A can be thought of as a map from $\mathbb{F}^n$ to $\mathbb{F}^p$ (i.e., with domain $\mathbb{F}^n$ and codomain $\mathbb{F}^p$) which sends u to Au. As we have already suggested, this map has very special properties which are at the heart of the idea of linearity in algebra.

Definition 2.1. *A map $L : \mathbb{F}^n \to \mathbb{F}^p$ is called a linear map if it maps linear combinations to linear combinations in the following way:*

$$L(\alpha u + \beta v) = \alpha L(u) + \beta L(v), \quad \text{for all } u, v \in \mathbb{F}^n, \ \alpha, \beta \in \mathbb{F}.$$

Lemma 2.2. *A map $L : \mathbb{F}^n \to \mathbb{F}^p$ is linear if and only if it can be represented by a $p \times n$ matrix A, i.e., $L(u) = Au$ for some $p \times n$ matrix A.*

It is quite useful to see why this is true. First, we already know that given A the map $L(u) = Au$ is linear. Now suppose L is linear. Define A to be the matrix whose j-th column is the p-vector $L(\mathbf{e}_j)$, i.e.,

$$A = \begin{pmatrix} L(\mathbf{e}_1) & \dots & L(\mathbf{e}_n) \end{pmatrix}.$$

We claim that for every $u \in \mathbb{F}^n$, $L(u) = Au$. To see this, write u in the standard basis: $u = \sum_{j=1}^n \alpha_j \mathbf{e}_j$. Now apply L and use the linearity:

$$L(\sum_{j=1}^n \alpha_j \mathbf{e}_j) = \sum_{j=1}^n \alpha_j L(\mathbf{e}_j) = \begin{pmatrix} L(\mathbf{e}_1) & \dots & L(\mathbf{e}_n) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = Au.$$

Remark 2.1. For reasons which will become more meaningful later when we introduce abstract linear algebra, we will distinguish between the linear map and the matrix that represents it. For example, the linear map

$$L : \mathbb{R}^2 \to \mathbb{R}^3; \quad L(x_1, x_2) = (x_1, x_2, 0),$$

is represented by the matrix

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

The set of all $p \times n$ matrices with entries from the scalar field $\mathbb{F}$ will be denoted by $M_{p \times n}(\mathbb{F})$. It will be convenient to have a notation for indicating the entries of a matrix: we will denote the entry of the j-th row and k-th column of a matrix A by A_{jk} and then write $A = (A_{jk})$.

Remark 2.2. We will have cause later on to be careful about the choice of scalar field. Any linear map $L : \mathbb{R}^n \to \mathbb{R}^p$ has a natural extension to a linear map $L : \mathbb{C}^n \to \mathbb{C}^p$: the matrix doesn't change and we are just using the fact that $\mathbb{R} \subset \mathbb{C}$. But the converse is not true: there are plenty of linear maps $L : \mathbb{C}^n \to \mathbb{C}^p$ which cannot be restricted to a map of $\mathbb{R}^n$ into $\mathbb{R}^p$. For example, $L : \mathbb{C}^2 \to \mathbb{C}$, $L(x_1, x_2) = ix_1$, represented by

$$A = \begin{pmatrix} i & 0 \end{pmatrix},$$

clearly does not map $\mathbb{R}^2$ into $\mathbb{R}$.

Recall that we can add matrices of the same size, and indeed make linear combinations of them: for $A, B \in M_{p\times n}(\mathbb{F})$ we define

$$\alpha A + \beta B = (\alpha A_{jk} + \beta B_{jk}).$$

If L, M are the corresponding linear maps $\mathbb{F}^n \to \mathbb{F}^p$ then $\alpha A + \beta B$ represents the linear combination of linear maps

$$(\alpha L + \beta M) : \mathbb{F}^n \to \mathbb{F}^p; \quad (\alpha L + \beta M)(u) = \alpha L(u) + \beta M(u).$$

Matrix multiplication is **defined** to correspond to composition of linear maps.

Lemma 2.3. *Given two linear maps $L : \mathbb{F}^n \to \mathbb{F}^p$ and $M : \mathbb{F}^p \to \mathbb{F}^q$, represented by matrices $A \in M_{p\times n}(\mathbb{F})$ and $B \in M_{q\times p}(\mathbb{F})$ respectively, their composite $(M \circ L)(u) = M(L(u))$ is represented by the matrix product BA.*

2.2. **Image and kernel; rank and nullity.** We can use linear maps to re-phrase the problem of existence and uniqueness for linear systems of equations. To a system of p linear equations in n unknowns, $A\mathbf{x} = \mathbf{y}$ we assign the linear map $L : \mathbb{F}^n \to \mathbb{F}^p$ given by $L(\mathbf{x}) = A\mathbf{x}$. Recall that the **image** (or range) of this map L is the set

$$\mathrm{Im}(L) = \{\mathbf{y} \in \mathbb{F}^p : \mathbf{y} = L(\mathbf{x}) \text{ for some } \mathbf{x} \in \mathbb{F}^n\}.$$

We also introduce a new set, called the **kernel** (or **null-space**) for a linear map L, defined to be

$$\mathrm{Ker}(L) = \{\mathbf{x} \in \mathbb{F}^n : L(\mathbf{x}) = \mathbf{0}\}.$$

Notice that when $L(\mathbf{x}) = A\mathbf{x}$ the kernel of L is exactly the solution set to the homogeneous system $A\mathbf{x} = \mathbf{0}$. This is one of the reasons why we are interested in it.

Lemma 2.4. *For a linear map $L : \mathbb{F}^n \to \mathbb{F}^p$, the image $\mathrm{Im}(L)$ is a subspace of the codomain $\mathbb{F}^p$, while the kernel $\mathrm{Ker}(L)$ is a subspace of domain $\mathbb{F}^n$. Moreover, when L is represented by a matrix A the image $\mathrm{Im}(L)$ is the subspace spanned by the columns of A.*

The dimension of the image of L is called the **rank** of L. The dimension of the kernel (null-space) of L is called the **nullity** of L. We write

$$\dim(\mathrm{Im}(L)) = \mathrm{rank}(L), \quad \dim(\mathrm{Ker}(L)) = \mathrm{null}(L).$$

Proof. Suppose $v_1, v_2 \in \mathrm{Im}(L)$. By definition this means there are $u_1, u_2 \in \mathbb{F}^n$ for which $L(u_j) = v_j$. Now by linearity

$$\alpha v_1 + \beta v_2 = L(\alpha u_1 + \beta u_2) \in \mathrm{Im}(L),$$

whenever $\alpha, \beta \in \mathbb{F}$. Also $\mathbf{0} = L(\mathbf{0})$ so $\mathbf{0} \in \mathrm{Im}(L)$. Hence $\mathrm{Im}(L)$ is a subspace. Now suppose $u_1, u_2 \in \mathrm{Ker}(L)$, then by linearity

$$L(\alpha u_1 + \beta u_2) = \alpha L(u_1) + \beta L(u_2) = \mathbf{0},$$

so $\alpha u_1 + \beta u_2 \in \mathrm{Ker}(L)$, and $\mathbf{0} \in \mathrm{Ker}(L)$ so $\mathrm{Ker}(L)$ is a subspace. Finally, recall from above that the columns of A are $L(\mathbf{e}_1), \ldots, L(\mathbf{e}_n)$. Now $v \in \mathrm{Im}(L)$ if and only if $v = L(u)$ for some $u \in \mathbb{F}^n$ and we know we can write

$$u = \sum \alpha_j \mathbf{e}_j, \quad \text{for some } \alpha_j \in \mathbb{F}.$$

Linearity gives
$$L(\sum \alpha_j \mathbf{e}_j) = \sum \alpha_j L(\mathbf{e}_j) \in \mathrm{Sp}(L(\mathbf{e}_1), \dots, L(\mathbf{e}_n)),$$
and we conclude that $\mathrm{Im}(L) = \mathrm{Sp}(L(\mathbf{e}_1), \dots, L(\mathbf{e}_n))$. □

Recall that a map $L : \mathbb{F}^n \to \mathbb{F}^p$ is **one-to-one** (or injective) when $L(u) = L(v)$ implies $u = v$, and **onto** (or surjective) when $\mathrm{Im}(L) = \mathbb{F}^p$. We say it is **bijective** when it is both one-to-one and onto. For linear maps these properties are very closely related because of the following theorem, which is one of the most important in finite-dimensional linear algebra.

Rank-Nullity Theorem. *For a linear map $L : \mathbb{F}^n \to \mathbb{F}^p$, $n = \mathrm{rank}(L) + \mathrm{null}(L)$.*

The basic idea to help to understand the meaning of this result and the proof below. We start with a basis $\{u_1, \dots, u_q\} \subset \mathbb{F}^n$ for the kernel $\mathrm{Ker}(L)$ (hence $q = \mathrm{null}(L)$). By Lemma 1.12(ii) we can extend this to a basis of $\mathbb{F}^n$: call these additional vectors $v_1, \dots, v_{n-q}$. We can prove that $L(v_1), \dots, L(v_{n-q})$ is a basis for $\mathrm{Im}(L)$, and therefore $\mathrm{rank}(L) = n - q = n - \mathrm{null}(L)$.

Proof of the Rank-Nullity Theorem. Choose a basis $\{u_1, \dots, u_q\} \subset \mathbb{F}^n$ for the kernel $\mathrm{Ker}(L)$ (hence $q = \mathrm{null}(L)$). By Lemma 1.12(ii) we can extend this to a basis of $\mathbb{F}^n$: call these additional vectors $v_1, \dots, v_{n-q}$. We claim that $L(v_1), \dots, L(v_{n-q})$ is a basis for $\mathrm{Im}(L)$, and therefore $\mathrm{rank}(L) = n - q = n - \mathrm{null}(L)$. To prove this we must show they are linearly independent and span $\mathrm{Im}(L)$. First, suppose
$$\sum_{j=1}^{n-q} \alpha_j L(v_j) = \mathbf{0}.$$
By linearity this means $\sum_{j=1}^{q} \alpha_j v_j \in \mathrm{Ker}(L)$, but $\mathrm{Ker}(L)$ is spanned by $u_1, \dots, u_q$ and the v_j are independent of these, so it must be that $\alpha_j = 0$ for all j and therefore $L(v_1), \dots, L(v_{n-q})$ are linearly independent.

Next, suppose $w \in \mathrm{Im}(L)$, i.e., $w = L(v)$ for some $v \in \mathbb{F}^n$. We can write
$$v = \sum_{j=1}^{q} \alpha_j u_j + \sum_{j=1}^{n-q} \beta_j v_j,$$
for some choice of scalars $\alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_{n-q}$. Hence
$$w = L(v) = L(\sum_{j=1}^{q} \alpha_j u_j + \sum_{j=1}^{n-q} \beta_j v_j) = \sum_{j=1}^{n-q} \beta_j L(v_j),$$
using linearity and the fact that $L(u_j) = \mathbf{0}$ for all u_j. Hence every $w \in \mathrm{Im}(L)$ is a linear combination of the vectors $L(v_1), \dots, L(v_{n-q})$, i.e., they span $\mathrm{Im}(L)$. Hence $L(v_1), \dots, L(v_{n-q})$ is a basis for $\mathrm{Im}(L)$ as claimed. □

Lemma 2.5. *A linear map $L : \mathbb{F}^n \to \mathbb{F}^p$ is one-to-one if and only if* $\mathrm{Ker}(L) = \{\mathbf{0}\}$. *Hence L is: (i) one-to-one if and only if* $\mathrm{null}(L) = 0$, *(ii) onto if and only of* $\mathrm{rank}(L) = p$, *(iii) bijective if and only if* $\mathrm{null}(L) = 0$ *and $p = n$.*

Proof. (i) Suppose L is one-to-one. Since $L(\mathbf{0}) = \mathbf{0}$ this means $u \in \mathrm{Ker}(L)$ implies $L(u) = L(\mathbf{0})$, so that $u = \mathbf{0}$. Hence $\mathrm{Ker}(L) = \{\mathbf{0}\}$. Conversely, if $\mathrm{Ker}(L) = \{\mathbf{0}\}$ and $L(u) = L(v)$ then $L(u - v) = \mathbf{0}$ by linearity, so $u - v = \mathbf{0}$, i.e., $L(u) = L(v)$ implies $u = v$, so L is one-to-one. Since $\{\mathbf{0}\}$ is the only zero dimensional subspace, $\mathrm{Ker}(L) = \{\mathbf{0}\}$ if and only if $\mathrm{null}(L) = 0$. (ii) Since $\mathbb{F}^p$ is the only p dimensional subspace of $\mathbb{F}^p$, L is onto precisly when $\mathrm{rank}(L) = p$. (iii) Finally, L is bijective if and only if both $\mathrm{null}(L) = 0$ and $\mathrm{rank}(L) = p$, but by the Rank-Nullity Theorem their sum is n, so $p = n$. □

We can apply these results to get a complete understanding of the existence and uniqueness of solutions to linear systems of equations.

Corollary 2.6. *A linear system $L(\mathbf{x}) = \mathbf{y}$ possesses a solution $\mathbf{x}$ if and only if $\mathbf{y} \in \mathrm{Im}(L)$. When it does possess a solution, the solution is unique if and only if L is a one-to-one map, which happens if and only if $\mathrm{Ker}(L) = \{\mathbf{0}\}$ (equally, $\mathrm{null}(L) = 0$).*

Proof. The first sentence is obvious. To say the solution is unique means if $L(\mathbf{x}_1) = L(\mathbf{x}_2)$ then $\mathbf{x}_1 = \mathbf{x}_2$, which is exactly the statement that L is one-to-one. □

Remark 2.3. Notice that for a linear system uniquenes depends only on the map L regardless of the choice of $\mathbf{y}$ (provided $\mathbf{y} \in \mathrm{Im}(L)$ to have a solution at all). The same is not generally true for nonlinear functions. Consider, for example, the nonlinear map

$$F : \mathbb{R}^2 \to \mathbb{R}^2; \quad F(x_1, x_2) = (x_1^2 + x_2^2, x_2).$$

There is a unique solution to $F(x_1, x_2) = (1, 1)$, given by $(x_1, x_2) = (0, 1)$, but two solutions to $F(x_1, x_2) = (1, a)$ whenever $|a| < 1$, given by $(x_1, x_2) = (\pm\sqrt{1 - a^2}, a)$. Geometrically the solutions are points on the unit circle $x_1^2 + x_2^2 = 1$ lying on a horizontal line of constant x_2. When $x_2 = 1$ the line is tangent to the circle and there is a single point of intersection; when $x_2 = a$ for $|a| < 1$ the line cuts the circle with two points of intersection.

Everything we need to know about counting the number of independent solutions to a system of linear equations is contained in the following very important corollary.

Corollary 2.7. *For a homogeneous linear system of p equations in n unknowns, $A\mathbf{x} = \mathbf{0}$, the number of linearly independent solutions equals $n - \mathrm{rank}(A)$.*

Remark 2.4. The statement above follows directly from the Rank-Nullity Theorem if we consider $\mathrm{rank}(A)$ to be the column rank of A, which can be thought of as either the number of linearly independent columns or the dimension of the subspace of $\mathbb{F}^p$ spanned by the columns. However, it is a true fact that for matrices this equals the row rank (we require inner products to prove this - see the exercises for Chapter 3). Since each row of A represents an equation, we can re-phrase the Rank-Nullity Theorem as saying that, for a system of linear homogeneous equations in n unknowns,

> *the number of linearly independent solutions is equal to $n - r$ where r is the number of linearly independent equations.*

One can see this very clearly in the Gaussian elimination algorithm. It produces r rows in echelon form.

Example 2.1. Consider the homogeneous linear system

$$x_1 + 2x_2 - x_4 = 0, \quad x_1 - x_2 - x_3 = 0,$$

which can be written as $L_1(x_1, x_2, x_3, x_4) = \mathbf{0}$ where $L_1 : \mathbb{R}^4 \to \mathbb{R}^2$ corresponds to the matrix

$$A_1 = \begin{pmatrix} 1 & 2 & 0 & -1 \\ 1 & -1 & -1 & 0 \end{pmatrix}.$$

The last two columns are clearly linearly independent, and since $p = 2$ in this example there cannot be more than two linearly independent columns, so $\operatorname{rank}(A_1) = 2$. By the Rank-Nullity Theorem the dimension of the solution space $\{\mathbf{x} \in \mathbb{R}^4 : L_1(\mathbf{x}) = \mathbf{0}\}$ is $4 - 2 = 2$.

Now by row reduction we obtain a matrix in echelon form

$$A_2 = \begin{pmatrix} 1 & 2 & 0 & -1 \\ 0 & -3 & -1 & 1 \end{pmatrix}$$

corresponding to a different linear map $L_2 : \mathbb{R}^4 \to \mathbb{R}^2$ but with $\operatorname{Ker}(L_1) = \operatorname{Ker}(L_2)$, i.e., $L_1(\mathbf{x}) = \mathbf{0}$ and $L_2(\mathbf{x}) = \mathbf{0}$ have the same solution spaces. The form of A_2 makes it clear that there are two linearly independent equations (rows), so our second point of view again shows that the number of linearly independent solutions to $L_1(\mathbf{x}) = \mathbf{0}$ is $4 - 2 = 2$.

2.3. **Invertible linear maps; change of basis.** Recall that any map between sets is invertible if and only if it is bijective. From Lemma 2.5 it follows that a linear map $L : \mathbb{F}^n \to \mathbb{F}^p$ is invertible if and only if both $n = p$ and $\operatorname{null}(L) = 0$.

Lemma 2.8. *If $L : \mathbb{F}^n \to \mathbb{F}^n$ is invertible, represented by an $n \times n$ matrix A, then $L^{-1} : \mathbb{F}^n \to \mathbb{F}^n$ is also linear and is represented by the inverse matrix A^{-1}.*

The first part follows immediately from the linearity condition. Suppose $L(u) = a$ and $L(v) = b$, so that $u = L^{-1}(a)$, $v = L^{-1}(b)$. Then

$$\begin{aligned} L(\alpha u + \beta v) = \alpha L(u) + \beta L(v) \quad &\Leftrightarrow \quad \alpha u + \beta v = L^{-1}(\alpha L(u) + \beta L(v)) \\ &\Leftrightarrow \quad \alpha L^{-1}(a) + \beta L^{-1}(b) = L^{-1}(\alpha a + \beta b). \end{aligned}$$

The second part follows from Lemma 2.3.

An invertible linear map can be interpreted as a change of basis.

Lemma 2.9. *Given a basis $v_1, \ldots, v_n$ for $\mathbb{F}^n$, a linear map $L : \mathbb{F}^n \to \mathbb{F}^n$ is invertible if and only if $L(v_1), \ldots, L(v_n)$ is again a basis. In particular, an $n \times n$ matrix A is invertible if and only if its columns provide a basis for $\mathbb{F}^n$.*

Proof. Suppose L is invertible. We will show that $L(v_1), \ldots, L(v_n)$ are linearly independent (and since $\mathbb{F}^n$ has dimension n it follows from Lemma 1.12(iv) that this is a basis). Suppose there are scalars α_j for which $\sum_{j=1}^n \alpha_j L(v_j) = \mathbf{0}$. Applying L^{-1} to both sides and using linearity gives $\sum_{j=1}^n \alpha_j v_j = \mathbf{0}$, hence all $\alpha_j = 0$ since $v_1, \ldots, v_n$ is a basis.

Conversely, suppose $L(v_1), \ldots, L(v_n)$ is a basis. By the Rank-Nullity Theorem it suffices to show that $\mathrm{Ker}(L) = \{\mathbf{0}\}$. So suppose $L(v) = \mathbf{0}$ for some $v \in \mathbb{F}^n$. Then we can write $v = \sum_{j=1}^n \alpha_j v_j$ for some scalars α_j, and by linearity

$$\mathbf{0} = L(v) = \sum_{j=1}^n \alpha_j L(v_j).$$

Since $L(v_1), \ldots, L(v_n)$ is a basis all $\alpha_j = 0$ and therefore $v = \mathbf{0}$.

The final statement follows from the fact that the columns of a matrix A are given by $A\mathbf{e}_1, \ldots, A\mathbf{e}_n$, i.e., the images of the standard basis under the linear map $L(v) = Av$. $\square$

Example 2.2. Consider the invertible linear maps of $\mathbb{R}^2$ into itself given by the following two matrices:

$$A_1 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The first corresponds to the map

$$L_1(x_1, x_2) = (-x_2, x_1).$$

By the proof above, this maps the (ordered) standard basis $(\mathbf{e}_1, \mathbf{e}_2)$ to $(\mathbf{e}_2, -\mathbf{e}_1)$ (the columns of A_1). In other words, it rotates the standard basis anti-clockwise by $\pi/2$. The second matrix corresponds to the map

$$L_2(x_1, x_2) = (x_2, x_1).$$

This maps the standard basis to the ordered basis $(\mathbf{e}_2, \mathbf{e}_1)$. This is a reflection across the line $x_1 = x_2$.

2.4. **Eigenvectors and eigenvalues.** Linear maps from $\mathbb{F}^n$ in to itself, $L : \mathbb{F}^n \to \mathbb{F}^n$, correspond to square $n \times n$ matrices. We can understand such linear maps using the ideas of eigenvectors and eigenvalues.

Definition 2.10. *For a linear map $L : \mathbb{F}^n \to \mathbb{F}^n$ an* ***eigenvector of*** *L (or equally, of the matrix A representing L) is a non-zero vector v for which $L(v) = \lambda v$ for some scalar $\lambda \in \mathbb{F}$. In that case λ is called an* ***eigenvalue of*** *L (or A).*

The set of all eigenvalues of L is called the ***spectrum*** *of L. It can be thought of as the set of all $\lambda \in \mathbb{F}$ for which $L - \lambda I_n$ is not invertible (has non-trivial kernel).*

Geometrically, an eigenvector v generates a direction (the line $\mathrm{Sp}(v)$) in which L acts by scaling.

Example 2.3. By considering just the case of $\mathbb{R}^2$ we can see a wide enough range of possible behaviours. We will think in terms of the matrix representing the linear map.

(i) There are matrices with distinct eigenvalues. For example, any diagonal matrix

$$A = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \quad \lambda_1 \neq \lambda_2$$

is easily seen to have eigenvectors $(1, 0)$ (with eigenvalue λ_1) and $(0, 1)$ (with eigenvalue λ_2). Notice that the eigenvectors are linearly independent and span $\mathbb{R}^2$.

(ii) Taking $\lambda_1 = \lambda_2$ in the previous example gives a matrix with only one eigenvalue but two linearly independent eigenvectors. Notice that in this case every non-zero vector is an eigenvector. This includes the case where $\lambda_1 = 0 = \lambda_2$, for which L is just the **zero map**.

(iii) Because we have chosen $\mathbb{F} = \mathbb{R}$ in this example, there are matrices with no eigenvalues. For example, the matrix

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

represents a rotation in $\mathbb{R}^2$ by $\pi/2$, and this clearly does not act by scaling in any direction.

(iv) There are matrices with only one linearly independent eigenvector. For example, the matrix

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

has an eigenvector $(1, 0)$, and every other eigenvector is just a scalar multiple of this.

An important point about eigenvectors is that they are never unique: if v is an eigenvector of L for eigenvalue λ then so is every scalar multiple of v. In fact the set of all eigenvectors for the same eigenvalue is almost a subspace. Let $I_n : \mathbb{F}^n \to \mathbb{F}^n$ denote the identity map (or matrix): $I_n(v) = v$ for all v. Clearly $L(v) = \lambda v$ is the same as $L(v) = \lambda I_n(v)$, which is the same as $(L - \lambda I_n)(v) = \mathbf{0}$.

Definition 2.11. *For a given eigenvalue λ of a linear map $L : \mathbb{F}^n \to \mathbb{F}^n$ we call the set*

$$\operatorname{Ker}(L - \lambda I_n) = \{v \in \mathbb{F}^n : (L - \lambda I_n)(v) = \mathbf{0}\}$$

*the **eigenspace for eigenvalue** λ of L. It consists of all eigenvectors of eigenvalue λ together with the zero vector $\mathbf{0}$. The dimension of the eigenspace* $\operatorname{Ker}(L - \lambda I_n)$ *is called the **geometric multiplicity** of the eigenvalue λ.*

We emphasize that $\mathbf{0}$ is **not** an eigenvector, even though it belongs to every eigenspace. An immediate consequence is that the eigenspace for any eigenvalue is a linear subspace of $\mathbb{F}^n$.

Example 2.4. The linear map $L_1 : \mathbb{C}^2 \to \mathbb{C}^2$, $L_1(x_1, x_2) = (-x_2, x_1)$, represented by the matrix A_1 from Example 2.2 can be shown to have eigenvalues $\pm i$. The eigenspaces are therefore

$$\operatorname{Ker}(L - iI_2) = \operatorname{Ker}\left(\begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix} \right) = \operatorname{Sp}\left(\begin{pmatrix} i \\ 1 \end{pmatrix} \right),$$

and

$$\operatorname{Ker}(L + iI_2) = \operatorname{Ker}\left(\begin{pmatrix} i & -1 \\ 1 & i \end{pmatrix} \right) = \operatorname{Sp}\left(\begin{pmatrix} i \\ -1 \end{pmatrix} \right),$$

Lemma 2.12. *Let $\lambda_1, \ldots, \lambda_q$ be distinct eigenvalues for a linear map $L : \mathbb{F}^n \to \mathbb{F}^n$. Then the corresponding eigenspaces $S_1, \ldots, S_q$ form a direct sum $S_1 \oplus \ldots \oplus S_q$. In particular, eigenvectors $v_1, \ldots, v_q$ for distinct eigenvalues are linearly independent.*

Proof. This can be proved by finite induction. First we show that $j \neq k$ S_j, S_k form a direct sum for any $j \neq k$, which is the same as saying $S_j \cap S_k = \{\mathbf{0}\}$. To show this, suppose $v \in S_j \cap S_k$ for any $j \neq k$. Then $L(v) = \lambda_j v$ and $L(v) = \lambda_k v$. Hence $(\lambda_j - \lambda_k)v = \mathbf{0}$. But $\lambda_j \neq \lambda_k$, so $v = \mathbf{0}$. Now make the inductive assumption that any $l-1$ eigenspaces form a direct sum. We will show that this implies any l form a direct sum. Then by induction on l all $S_1, \ldots, S_q$ form a direct sum.

So assume that $S_1, \ldots, S_{l-1}$ form a direct sum and consider $S_l \cap (\bigoplus_{k \neq l} S_k$. Notice that by relabelling this deals with all possible choices of l eigenspaces. Let $v \in S_l \cap (\bigoplus_{k \neq l} S_k)$. Then $v = \sum_{k \neq l} \alpha_k v_k$ for some scalars α_k, and applying L gives

$$\lambda_l(\sum_{k \neq l} \alpha_k v_k) = \sum_{k \neq l} \lambda_k \alpha_k v_k.$$

Re-arranging we obtain

$$\sum_{k \neq l} (\lambda_l - \lambda_k)\alpha_k v_k = \mathbf{0}.$$

But since the v_k are independent (by Lemma 1.17 and $\lambda_l \neq \lambda_k$ it follows that $v = \mathbf{0}$.

The last line of the statement follows just by an application of Lemma 1.17. ☐

2.5. Diagonalizability.

Definition 2.13. *A linear map $L : \mathbb{F}^n \to \mathbb{F}^n$ is said to be* ***diagonalizable over*** *$\mathbb{F}$ when the $n \times n$ matrix A representing it can be diagonalized, i.e., when there exists an invertible $n \times n$ matrix $P \in M_{n \times n}(\mathbb{F})$ for which $P^{-1}AP$ is a diagonal matrix.*

Notice the emphasis here on the choice of scalar field $\mathbb{F}$. We will see later that a linear map $L : \mathbb{R}^n \to \mathbb{R}^n$ can fail to be diagonalizable over $\mathbb{R}$ even though it is diagonalizable over $\mathbb{C}$.

You have already seen in first year Algebra the theorem (and proof) that an $n \times n$ matrix with n distinct eigenvalues is diagonalizable. This is a special case of the following result.

Theorem 2.14. *A linear map $L : \mathbb{F}^n \to \mathbb{F}^n$ is diagonalizable over $\mathbb{F}$ if and only if $\mathbb{F}^n$ has a basis which consists of eigenvectors of L. This is equivalent to saying that $\mathbb{F}^n$ is a direct sum of the eigenspaces of L. This occurs if and only if the sum of all the dimensions of the eigenspaces of L equals n.*

The proof is important as it shows us exactly how to construct an invertible matrix which diagonalizes the matrix A representing L.

Proof. First suppose the matrix A representing L is diagonalizable, so that

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \lambda_n \end{pmatrix}, \tag{2.1}$$

for scalars $\lambda_1, \dots, \lambda_n$ (not necessarily all different). Denote the diagonal matrix on the right by D, and re-arrange the equation into $AP = PD$. Now apply both sides to each standard basis vector $\mathbf{e}_j$ to get

$$AP\mathbf{e}_j = PD\mathbf{e}_j = \lambda_j P\mathbf{e}_j,$$

using the fact that $D\mathbf{e}_j$ is the j-th column $\lambda_j\mathbf{e}_j$ of D. Since $P\mathbf{e}_j$ is the j-th column of P, we see that this column is an eigenvector of A with eigenvalue λ_j. Since P is invertible these columns form a basis of $\mathbb{F}^n$. Conversely, if $v_1, \dots, v_n$ is a basis of $\mathbb{F}^n$ with $L(v_j) = \lambda_j v_j$ then the matrix P with columns $v_1, \dots, v_n$ in that order satisfies (2.1).

Now let $\mu_1, \dots, \mu_q$ denote the distinct eigenvalues, $S_1, \dots, S_q$ the corresponding eigenspaces, and choose a basis $\mathcal{B}_j$ for each S_j. Let $n_j = \dim(S_j) = \#\mathcal{B}_j$. If $\mathbb{F}^n = \bigoplus_{j=1}^q S_j$ then $\cup_{j=1}^q \mathcal{B}_j$ is a basis of eigenvectors for $\mathbb{F}^n$ since then

$$n = \sum_{j=1}^q \dim(S_j) = \sum_{j=1}^q n_j = \# \cup_{j=1} \mathcal{B}_j,$$

i.e., this union consists of n linearly independent vectors by Lemma 2.12. Conversely, if $\cup_{j=1}^q \mathcal{B}_j$ has n elements then this is a basis of eigenvectors for $\mathbb{F}^n$ and therefore $\mathbb{F}^n = \bigoplus_{j=1}^q S_j$. $\square$

We will leave examples until the next section.

2.6. **The characteristic polynomial.** It is a central fact about matrices that an $n \times n$ matrix is not invertible if and only if its determinant is zero. Hence the determinant plays a central role in the study of eigenvalues, since the spectrum of an $n \times n$ matrix A is therefore the set

$$\{\lambda \in \mathbb{F} : \det(A - \lambda I_n) = 0\}.$$

The next result is a reminder about properties of determinants from first year Algebra which we need for the discussion about eigenvalues (the next section provides a summary of more facts about determinants that were taught in first year).

Theorem 2.15 (Determinants). *For $n \times n$ matrices A, B:*

(i) $\det(A) = 0$ *if and only if* $\operatorname{rank}(A) < n$ *(equally, A is not invertible),*

(ii) $\det(AB) = \det(A)\det(B)$, *hence* $\det(B^{-1}) = \frac{1}{\det(B)}$ *when B is invertible, and therefore* $\det(B^{-1}AB) = \det(A)$ *for every invertible matrix B (similar matrices have the same determinant).*

(iii) *If A is upper or lower triangular,* $\det(A)$ *is the product of diagonal elements* $A_{11}A_{22}\dots A_{nn}$. *It follows that* $\det(\alpha I_n) = \alpha^n$, *hence* $\det(\alpha A) = \alpha^n \det(A)$ *for any scalar* α.

(iv) $\det(A^\top) = \det(A)$, *where* $A^\top$ *is the transpose of* A, *i.e.,* $A^\top = (A_{kj})$ *when* $A = (A_{jk})$.

It follows quite quickly from the definition of the determinant (see below) that the quantity $\det(A - \lambda I_n)$ is a polynomial in λ of degree n.

Definition 2.16. *The polynomial* $c_A(\lambda) = \det(\lambda I_n - A)$ *is called the* ***characteristic polynomial*** *of* A. *We write its terms as*

$$c_A(\lambda) = \lambda^n + c_1(A)\lambda^{n-1} + \dots + c_n(A) = \lambda^n + \sum_{j=1}^{n} c_j(A)\lambda^{n-j}.$$

If a linear map L is represented by matrix A then we also call c_A the characteristic polynomial of L.

Remark 2.5. There is a sign choice to be made when defining the characteristic polynomial. We choose $\det(\lambda I_n - A)$ instead of $\det(A - \lambda I_n)$ to make the leading order term λ^n. There is no difference when n is even but the signs are opposite when n is odd.

The coefficients $c_j(A)$ of the characteristic polynomial are very interesting functions in their own right. We don't have time to discuss them in detail, but we can state some facts.

Lemma 2.17. *For $n \times n$ matrices, each $c_j(A)$ is a polynomial function of degree j in the entries of A. In particular, $c_1(A)$ is the linear function*

$$c_1(A) = -\sum_{j=1}^{n} A_{jj},$$

while $c_n(A) = (-1)^n \det(A)$. *For every j, $c_j(B^{-1}AB) = c_j(A)$ for every invertible matrix B. In particular, similar matrices have the same characteristic polynomial.*

The last part follows immediately from the properties of determinants, since

$$\lambda I_n - B^{-1}AB = B^{-1}(\lambda I_n - A)B \Rightarrow \det(\lambda I_n - B^{-1}AB) = \det(\lambda I_n - A).$$

Remark 2.6. The sum $\sum_{j=1}^{n} A_{jj}$ of diagonal entries of A is called the **trace** of A and denoted $\mathrm{tr}(A)$. Hence $c_1(A) = -\mathrm{tr}(A)$. It is a deep fact, well beyond the scope of this module, that for each j, $c_j(A)$ can be expressed in terms of the polynomials $\mathrm{tr}(A), \mathrm{tr}(A^2), \dots, \mathrm{tr}(A^j)$. For example,

$$2c_2(A) = \mathrm{tr}(A)^2 - \mathrm{tr}(A^2).$$

Example 2.5. For any 2×2 matrix

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

a simple calculation shows that

$$c_A(\lambda) = \lambda^2 - (A_{11} + A_{22})\lambda + (A_{11}A_{22} - A_{12}A_{21}) = \lambda^2 - \mathrm{tr}(A)\lambda + \det(A).$$

So there is never a need to individually calculate the characteristic polynomial of a 2×2 matrix if you remember this simple expression.

We are now in a position to state the main relationship between eigenvalues and the characteristic polynomial.

Theorem 2.18. *Let $L : \mathbb{F}^n \to \mathbb{F}^n$ be a linear map represented by an $n \times n$ matrix A.*

(i) *The eigenvalues of L are the roots in $\mathbb{F}$ of its characteristic polynomial $c_A(\lambda)$. The multiplicity of the root is called the* ***algebraic multiplicity*** *of the eigenvalue.*
(ii) *For a given eigenvalue λ of L, its algebraic multiplicity is bigger or equal than its geometric multiplicity (recall that geometric multiplicity is defined in Definition 2.11).*
(iii) *If A is upper or lower triangular (in particular, if it is diagonal) then*

$$c_A(\lambda) = (\lambda - A_{11}) \dots (\lambda - A_{nn})$$

and therefore the eigenvalues are the diagonal entries $A_{11}, \dots, A_{nn}$ of A.
(iv) *There are at most n distinct eigenvalues for L.*
(v) *If L has n distinct eigenvalues (i.e., $c_A(\lambda)$ has n distinct roots in $\mathbb{F}$) then L is diagonalizable.*

Proof. We prove here only point ii and leave the rest for self-practice.

Let q be the geometric multiplicity of an eigenvalue λ_1 of L. By definition, this means that the dimension of the corresponding eigenspace $\mathrm{Ker}(L - \lambda_1 I_n)$ is q, hence we can pick a base of this eigenspace S_1 consisting of q elements. We proceed by completing S_1 to a base S of the whole space $\mathbb{F}^n$ (and we number elements of S so that S_1 are the first q elements of S). Then, the first q columns of the matrix $S^{-1}LS$ are $\lambda_1 \cdot \mathbf{e}_1, \dots, \lambda_1 \cdot \mathbf{e}_q$, so its characteristic polynomial $\lambda S^{-1}LS - I$ is divisible by $(\lambda - \lambda_1)^q$. Then, by Lemma 2.17 the characteristic polynomial of L is also divisible by $(\lambda - \lambda_1)^q$, which proves the claim. $\square$

Example 2.6.
(i) Without performing a single calculation we know that the matrix

$$\begin{pmatrix} -2 & \pi & 0 & e \\ 0 & \pi^2 & \frac{3}{4} & 27 \\ 0 & 0 & \log(2) & 8 \\ 0 & 0 & 0 & 117 \end{pmatrix}$$

has eigenvalues $-2, \pi^2, \log(2), 117$ because it is upper triangular and these are the diagonal entries. These are all distinct, so this matrix is diagonalizable over $\mathbb{R}$.
(ii) The linear map $L_1(x_1, x_2) = (-x_2, x_1)$ from Example 2.2 has characteristic polynomial

$$c_{A_1}(\lambda) = \lambda^2 + 1,$$

since $\mathrm{tr}(A_1) = 0$ and $\det(A_1) = 1$. So as a linear map from $\mathbb{R}^2$ to $\mathbb{R}^2$ this has no eigenvalues and is not diagonalizable over $\mathbb{R}$. But since $\lambda^2+1 = (\lambda-i)(\lambda+i)$, as a linear map from $\mathbb{C}^2$ into $\mathbb{C}^2$ this is diagonalizable. The eigenspaces were calculated in Example 2.4 and we can choose a basis of eigenvectors as

$$\begin{pmatrix} i \\ 1 \end{pmatrix} \text{ for } \lambda = i, \quad \begin{pmatrix} i \\ -1 \end{pmatrix} \text{ for } \lambda = -i.$$

It follows from the section on diagonalizability that

$$\begin{pmatrix} i & i \\ 1 & -1 \end{pmatrix}^{-1} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} i & i \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$

(iii) The matrix

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

is upper triangular so its eigenvalues are $1, 1, 0$. In other words, the eigenvalue 1 has algebraic multiplicity two, while 0 has algebraic multiplicity one. These are not distinct, so we have to check whether or not there is a basis of eigenvectors. If there is, this is diagonalizable over $\mathbb{R}$. The eigenspaces are

$$\mathrm{Ker}(A - I_3) = \mathrm{Ker}\left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{pmatrix} \right) = \mathrm{Sp}(\mathbf{e}_1, \mathbf{e}_2),$$

and

$$\mathrm{Ker}(A) = \mathrm{Sp}(\mathbf{e}_2 - \mathbf{e}_3).$$

So there is a basis of eigenvectors, given by $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_2 - \mathbf{e}_3$. Using these as columns for P in (2.1) we conclude that

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

2.7. **Determinants: a reminder.** This section will not be lectured: it is a summary of the facts about determinants which were taught in first year.

Definition 2.19. *Let $A = (A_{jk})$ be an $n \times n$ matrix with entries in $\mathbb{F}$. The **determinant** of A is defined by*

$$\det(A) = \begin{vmatrix} A_{11} & \dots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \dots & A_{nn} \end{vmatrix} = \sum_{\sigma \in S_n} (-1)^{N(\sigma)} A_{1\sigma(1)} \dots A_{n\sigma(n)}. \tag{2.2}$$

Here

$$S_n = \{\sigma : \{1, 2, \dots, n\} \to \{1, 2, \dots, n\} \mid \sigma \text{ is invertible}\}$$

is the set of all permutations of $\{1, 2, \ldots, n\}$ *(the* ***symmetric group****). Also,*

$$N(\sigma) = \#\{(j, k) \mid 1 \leq j < k \leq n \text{ and } \sigma(j) > \sigma(k)\},$$

is called the number of ***inversions*** *of* σ*.*

This definition for the determinant of A is known as the **Leibniz formula**. Some basic facts which follow from it are as follows.

Theorem 2.20. *For a square* $n \times n$ *matrix* $A \in M_{n\times n}(\mathbb{F})$ *with columns* $A_1, \ldots, A_n$ *its determinant* $\det(A)$ *has the following two properties.*

(i) *Swapping columns changes the sign of* $\det(A)$*. In particular, if any two columns of* A *are scalar mutiples of each other then* $\det(A) = 0$*.*
(ii) *If the* j*-th column* A_j *is replaced* $\alpha A_j + \beta A_k$ *then the new matrix has determinant* $\alpha \det(A)$*.*

The properties above are equally true if "columns" is replaced by "rows".

Remark 2.7. In fact these two properties characterize the determinant up to a scale factor. That is to say, think of the determinant as a function of its columns:

$$\det : \mathbb{F}^n \times \ldots \times \mathbb{F}^n \to \mathbb{F}; \quad \det(A_1, \ldots, A_n) = \det(A).$$

There is exactly one such function with the property that it is linear in each column, changes sign when columns are swapped, and for which $\det(\mathbf{e}_1, \ldots, \mathbf{e}_n) = 1$. This last condition is needed to get the right scale factor.

For the practical purpose of calculating a determinant, most people use the **Laplace (or cofactor)** expansion, which uses the minors and cofactors.

Definition 2.21. *Let* $A = (A_{jk}) \in M_{n\times n}(\mathbb{F})$*. For each* $1 \leq j, k \leq n$ *we obtain an* $(n-1) \times (n-1)$ *matrix by deleting the* j*-th row and* k*-th column of* A*. The determinant of this matrix, written* $M_{jk} \in \mathbb{F}$*, is called the* (j, k)***-minor of*** A*. The* (j, k)***-cofactor of*** A *is then defined to be* $C_{jk} := (-1)^{j+k} M_{jk} \in \mathbb{F}$*.*

Theorem 2.22 (Laplace formula for the determinant)**.** *Let* $A = (A_{jk}) \in M_{n\times n}(\mathbb{F})$*. For each fixed all* $1 \leq j \leq n$ *we have the* ***expansion along the*** j***-th row***

$$\det(A) = \sum_{k=1}^{n} A_{jk} C_{jk}. \tag{2.3}$$

For each fixed $1 \leq k \leq n$ *we have the* ***expansion along the*** k***-th column***

$$\det(A) = \sum_{j=1}^{n} A_{jk} C_{jk}. \tag{2.4}$$

The formulas (2.3) and (2.4) lead to an expression which provides with us with the inverse of a matrix when its determinant is not zero. This is not very practical for $n > 5$ but does have some theoretical use.

Definition 2.23. *Let $A = (A_{jk}) \in M_{n\times n}(\mathbb{F})$. The* ***cofactor matrix of*** *A is the matrix whose (j,k)-entry is the (j,k)-cofactor of A: $\operatorname{cof}(A) := (C_{jk})$. The* ***classical adjoint (or adjugate matrix) of*** *A is the transpose of the cofactor matrix: $\operatorname{adj}(A) := \operatorname{cof}(A)^\top = (C_{kj})$.*

Theorem 2.24 (Adjugate, determinant and inverse)**.** *Let $A = (A_{ij}) \in M_{n\times n}(\mathbb{F})$. Then*

$$A \operatorname{adj}(A) = \det(A) I_n = \operatorname{adj}(A) A. \tag{2.5}$$

In particular, if $\det(A) \neq 0$ then A is invertible and $A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A)$.

2.8. **Cayley-Hamilton Theorem.** Suppose $p(x) = \alpha_m x^m + \alpha_{m-1}x^{m-1} + \cdots + \alpha_1 x + \alpha_0$ is a polynomial with coefficients in $\mathbb{F}$. Given an $n \times n$ matrix A with entries in $\mathbb{F}$, we can define $p(A)$ to be the following matrix sum:

$$p(A) := \alpha_m A^m + \alpha_{m-1}A^{m-1} + \cdots + \alpha_1 A + \alpha_0 I,$$

where I is the $n\times n$ identity matrix. With this definition in hand, we can state the following remarkable theorem.

Theorem 2.25 (The Cayley-Hamilton Theorem)**.** *Let A be an $n \times n$ matrix with entries in $\mathbb{F}$, and let $c_A(x)$ be the characteristic polynomial of A. Then $c_A(A) = O$ (where O is the $n \times n$ matrix with zero entries).*

Proof. If $B = xI - A$, then $c_A(x) = \det(B)$. So, $c_A(x)$ is a polynomial of degree n, so that

$$c_A(x) = c_n + c_{n-1}x + \cdots + x^n. \tag{2.6}$$

Consider the adjugate matrix of B. By Theorem 2.24 it has the property

$$B \operatorname{adj}(B) = \det(B) I.$$

The (i,j) entry of $\operatorname{adj}(B)$ is the (j,i) cofactor of B, and each cofactor is $\pm$ determinant of some $(n-1)\times(n-1)$ supmatrix of B. Therefore, each entry of $\operatorname{adj}(B)$ is a polynomial of degree at most $n-1$:

$$\operatorname{adj}(B) = \begin{pmatrix} p_{11}(x) & \cdots & p_{1n}(x) \\ \vdots & \ddots & \vdots \\ p_{n1}(x) & \cdots & p_{nn}(x) \end{pmatrix}$$

where each $p_{ij}(x)$ is a polynomial of degree at most $n-1$. Hence we can write $\operatorname{adj}(B)$ as

$$\operatorname{adj}(B) = B_0 + B_1 x + \cdots + B_{n-1}x^{n-1} \tag{2.7}$$

where $B_0, \ldots, B_{n-1}$ are $n \times n$ matrices.

Further, we have

$$\det(B) I = B \operatorname{adj}(B) = (A - xI) \operatorname{adj}(B) = A \operatorname{adj}(B) - x \operatorname{adj}(B). \tag{2.8}$$

It follows from (2.6) that

$$\det(B) I = c_0 I + c_1 I x + \cdots + c_n I x^n. \tag{2.9}$$

Equation (2.7) implies that

$$A \operatorname{adj}(B) - x \operatorname{adj}(B) = AB_0 + AB_1 x + \cdots + AB_{n-1}x^{n-1} - B_0 x - B_1 x^2 - \cdots - B_{n-1}x^n. \tag{2.10}$$

Substituting (2.9) and (2.10) into (2.8) and equating coefficients of powers of x, we get

$$\begin{aligned}
c_0 I &= AB_0 \\
c_1 I &= AB_1 - B_0 \\
c_2 I &= AB_2 - B_1 \\
&\vdots \\
c_{n-1} I &= AB_{n-1} - B_{n-2} \\
c_n I &= -B_{n-1}.
\end{aligned}$$

Now we multiply these equations on the left by I, A, A^2, ...,A^n respectively. As a result, we obtain

$$\begin{aligned}
c_0 I &= AB_0 \\
c_1 A &= A^2 B_1 - AB_0 \\
c_2 A^2 &= A^3 B_2 - A^2 B_1 \\
&\vdots \\
c_{n-1} A^{n-1} &= A^n B_{n-1} - A^{n-1} B_{n-2} \\
c_n A^n &= -A^n B_{n-1}.
\end{aligned}$$

Adding up these equations, we find that all the terms on the right side cancel out, so that

$$c_A(A) = c_0 I + c_1 A + \cdots + c_n A^n = 0$$

as required. □

Example 2.7. Let us employ the Cayley-Hamilton Theorem to find the inverse of the matrix

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

Since $\det(A) = 5$, A^{-1} exists. First we compute the characteristic polynomial of A:

$$c_A(x) = \begin{vmatrix} 3-x & 0 & 4 \\ 1 & 1-x & 2 \\ 1 & 0 & 3-x \end{vmatrix} = -(x-1)^2(x-5) = -x^3 + 7x^2 - 11x + 5.$$

According to the Cayley-Hamilton Theorem,

$$-A^3 + 7A^2 - 11A + 5I = O.$$

Multiplying this by A^{-1} on the left, we find that

$$-A^2 + 7A - 11I + 5A^{-1} = O \quad \text{or} \quad A^{-1} = \frac{1}{5}(A^2 - 7A + 11I) = \begin{pmatrix} 3/5 & 0 & -4/5 \\ -1/5 & 1 & -2/5 \\ -1/5 & 0 & 3/5 \end{pmatrix}.$$

2.9. **Minimal Polynomial.** Let $L : \mathbb{F}^n \to \mathbb{F}^n$ be a linear map and let $p(x) = \alpha_m x^m + \alpha_{m-1}x^{m-1} + \cdots + \alpha_1 x + \alpha_0$ be a polynomial with coefficients in $\mathbb{F}$. Let us define linear transformations L^2, L^3, ... by

$$L^2(v) = (L \circ L)(v), \quad L^3(v) = (L \circ L^2)(v), \ldots, \quad L^k(v) = (L \circ L^{k-1})(v), \ldots$$

Then we can define the linear transformation $p(L) : \mathbb{F}^n \to \mathbb{F}^n$ by the formula

$$p(L) := \alpha_m L^m + \alpha_{m-1}L^{m-1} + \cdots + \alpha_1 L + \alpha_0 I.$$

Definition 2.26. *Let $L : \mathbb{F}^n \to \mathbb{F}^n$ be a linear map. The **minimal polynomial** of L, $d_L(x)$, is the monic polynomial $p(x)$ of least degree such that $p(L) = 0$.*

Example 2.8. The characteristic polynomial of the identity transformation I_n is $c_I(x) = (1-x)^n$, whereas the minimal polynomial is $d_{I_n}(x) = x - 1$.

Lemma 2.27. *The minimal polynomial of a linear map L is unique, and it divides any other polynomial that annihilates L.*

Proof. Let $d_L(x)$ have degree m. Let $p(x)$ be a polynomial with $p(L) = 0$ (i.e. $p(L)$ is a zero transformation). Then the degree of $p(x) \geq m$. Therefore, by the remainder theorem for polynomials, we can write

$$p(x) = q(x)d_L(x) + r(x)$$

where $\deg(r) < \deg(d_L) = m$. Now substitute $x = L$, and we have $O = q(L)0 + r(L)$, so $r(L) = 0$. This contradicts the assertion that $d_L(x)$ has the smallest degree and $\deg(r) < \deg(d_L)$, unless $r(x) \equiv 0$. Hence d_L divides $p(x)$. If $\deg(d_L) = \deg(p)$, then they can only differ by a multiplicative constant. Insisting that $d_L(x)$ is monic fixes that constant, and thus the minimal polynomial is unique. □

Theorem 2.28. *Let $L : \mathbb{F}^n \to \mathbb{F}^n$ be a linear map. A scalar λ is an eigenvalue of L if and only if λ is a root of the minimal polynomial of L, i.e. $d_L(\lambda) = 0$.*

Proof. Assume that λ is a root of $d_L(x)$. Then $d_L(x) = (x - \lambda)p(x)$ for some polynomial $p(x)$. Since the degree of $p(x)$ is smaller than the degree of $d_L(x)$, we conclude that $p(L) \neq 0$. Therefore, there is $v \in V$ such that $w = p(L)(v) \neq \mathbf{0}$. Hence,

$$\mathbf{0} = d_L(L)(v) = (L - \lambda I)p(L)(v) = (L - \lambda I_n)(w),$$

which means that λ is an eigenvalue of L.

Conversely, let $v \in V$ be an eigenvector of L with eigenvalue λ, so that $L(v) = \lambda v$. Since $d_L(L) = 0$, we have $d_L(L)(v) = \mathbf{0}$. Therefore,

$$\begin{aligned} \mathbf{0} &= d_L(L)(v) \\ &= (L^k + a_{k-1}L^{k-1} + \cdots + a_0 I)(v) \\ &= (\lambda^k + a_{k-1}\lambda^{k-1} + \cdots + a_0)v \\ &= d_L(\lambda)v. \end{aligned}$$

Finally, since $v \neq \mathbf{0}$, we conclude that $d_L(\lambda) = 0$. □

Example 2.9. Consider two matrices

$$A = \begin{pmatrix} 1 & 1 & 4 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 & 4 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

These have the same characteristic polynomial: $c_A(x) = c_B(x) = -(x-1)(x-2)^2$ and, therefore, the same eigenvalues, 1 and 2. Also, the Cayley-Hamilton Theorem (see below) says that $c_A(A) = c_B(B) = 0$. Do they have the same minimal polynomial?

The answer is 'No'. Since 1 and 2 are eigenvalues, both $x-1$ and $x-2$ must divide the minimal polynomials $d_A(x)$ and $d_B(x)$. We have

$$(A-I)(A-2I) = \begin{pmatrix} 0 & 1 & 4 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & 4 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$$(B-I)(B-2I) = \begin{pmatrix} 0 & 1 & 4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Therefore, $d_A(x) = (x-1)(x-2)^2$ and $d_B(x) = (x-1)(x-2)$. Thus, $d_A(x)$ has the same degree as the characteristic polynomial $c_A(x)$, and $d_B(x)$ has a smaller degree than the characteristic polynomial $c_B(x)$.

Definition 2.29. *The* ***minimal multiplicity*** $m_\lambda \in \mathbb{N}$ *of an eigenvalue* λ *of a linear transformation* L *or a matrix* A *is the multiplicity of* λ *as a root of the minimal polynomial* $d_L(x)$ *or* $d_A(x)$*.*

By using Lemma 2.27, we see that the minimal multiplicity of any given eigenvalue of a linear transformation $L : \mathbb{F}^n \to \mathbb{F}^n$ is less than or equal to its algebraic multiplicity.

2.10. **Jordan Normal Form.** We have seen that not every matrix can be diagonalized, even over $\mathbb{C}$. In this chapter we will see what can be done with an arbitrary matrix over $\mathbb{C}$, even if it is not diagonalizable.

3. The Jordan Theorem

For $\lambda \in \mathbb{C}$ the **elementary Jordan block** of size l with the eigenvalue λ is the $l \times l$ matrix

$$J_{\lambda,l} = \begin{pmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \lambda & 1 \\ 0 & & & \lambda \end{pmatrix}$$

with λ's on the diagonal, 1's immediately above the diagonal, and 0's everywhere else. Note that the characteristic polynomial $c_J(x) = (\lambda - x)^l$ so λ is the only eigenvalue of J.

Definition 3.1. *The $n \times n$ matrix J is said to have the* ***Jordan normal form*** *if*

$$J = \begin{pmatrix} \boxed{J_1} & & & 0 \\ & \boxed{J_2} & & \\ & & \ddots & \\ 0 & & & \boxed{J_k} \end{pmatrix} \tag{3.1}$$

where for each $i = 1, \ldots, k$

$$J_i = J_{\lambda_i, l_i} \tag{3.2}$$

for some complex numbers $\lambda_1, \ldots, \lambda_k$ and some postive integers $l_1, \ldots, l_k$.

Note that $l_1 + \ldots + l_k = n$. Also note that J is diagonal if and only if $l_1 = \ldots = l_k = 1$.
Now let A be an $n \times n$ complex matrix. Then the characteristic polynomial of A has exactly n roots (not necessarily distinct), so the sum of the algebraic multiplicities of the eigenvalues of A equals n. Each distinct eigenvalue has at least one eigenvector associated with it. Recall that the geometric multiplicity of a given eigenvalue of A is the maximal number of linearly independent eigenvectors associated with that eigenvalue.

Theorem 3.2. (i) *For any complex matrix $n \times n$ matrix A there exist matrices J and P such that J has a Jordan normal form* (3.1)*, P is invertible and*

$$A = P\,J\,P^{-1} . \tag{3.3}$$

(ii) *The collection of pairs $(\lambda_1, l_1), \ldots, (\lambda_k, l_k)$ is determined by the given matrix A uniquely up to reordering these pairs.*

(iii) *The matrix P can be chosen so that the diagonal blocks of J with the same eigenvalue λ appear consecutively (one after another), and the sizes of these consecutive blocks with the same λ do not increase as one goes down the diagonal.*

Note that the eigenvalues $\lambda_1, \ldots, \lambda_k$ may be any complex numbers. Hence there is no canonical ordering for those of $\lambda_1, \ldots, \lambda_k$ which are distinct from each other. Furthermore, we will see that even for a fixed J the choice of the matrix P may be not unique.

3.1. **Multiplicities of Eigenvalues.** The two matrices A and J in (3.3) have the same eigenvalues. Moreover the algebraic, geometric and minimal multiplicities of the eigenvalues of A and J are respectively the same. This can be proved directly by definition. Alternatively, observe that all these multplicitites have been defined in terms of linear transformations. But A and J can be regarded as matrices of the same linear transformation of the coordinate vector space $\mathbb{C}^n$ relative to two different bases. For the matrix J all the multiplicities can be easily computed.
The algebraic multiplicity. The number of times any given $\lambda \in \mathbb{C}$ appears on the diagonal of J equals the algebraic multiplicity a_λ . To see this note that the determinant of an upper triangular matrix is just the product of the diagonal elements. Using the notation (3.2),

the characteristic polynomial of the Jordan normal form matrix J is

$$c_J(x) = \prod_{i=1}^{k} (\lambda_i - x)^{l_i} .$$

In other words, a_λ is the **total size** of all Jordan blocks of J with the given eigenvalue λ.

The geometric multiplicity. The **number** of the elementary Jordan blocks of J with the same eigenvalue λ equals the geometric multiplicity g_λ. By definition, g_λ this is the maximal number of linearly independent eigenvectors associated to λ. Each elementary Jordan block has only one eigenvector associated to it. Indeed, the vector $(1, 0, \ldots, 0)^T$ is the only eigenvector of

$$\begin{pmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \lambda & 1 \\ 0 & & & \lambda \end{pmatrix} .$$

Hence our statement about g_λ is true when the matrix J consists of only one elementary Jordan block. To generalize this result to an arbitrary J, we can use the induction on the number k of the Jordan blocks and the following general argument. Let X and Y be $p \times p$ and $q \times q$ matrices respectively. Let u and v be column vectors with p and q coordinates respectively. Consider the matrix and the vector

$$A = \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} u \\ v \end{pmatrix} .$$

For any $\lambda \in \mathbb{C}$ the equation $A\,w = \lambda\,w$ is equivalent to the pair of equations

$$\begin{cases} X\,u = \lambda\,u\,, \\ Y\,v = \lambda\,v\,. \end{cases}$$

Hence w is an eigenvecor of A if and only if one of the following three conditions is met:
(i) u is an eigenvector of X while $v = 0$;
(ii) v is an eigenvector of Y while $u = 0$;
(iii) u and v are respectively eigenvectors of X and Y with the **same** eigenvalue.
The minimal multiplicity. Finally, the **maximal size** of any elementary Jordan block with the eigenvalue λ is the minimal multiplicity m_λ. To prove this we will make repeated use of another general observation: if we multiply two matrices in block diagonal form and (so long as the blocks are of the right shape) then the answer is still in block diagonal form. Let X and Z be $p \times p$ matrices, and let Y and W be $q \times q$ matrices. Then for the $(p+q) \times (p+q)$ matrices

$$\begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix} \begin{pmatrix} Z & 0 \\ 0 & W \end{pmatrix} = \begin{pmatrix} XZ & 0 \\ 0 & YW \end{pmatrix} .$$

Definition 3.3. *If a matrix A satisfies $A^m = O$ but $A^{m-1} \neq O$, then we say that A is* ***nilpotent*** *of degree m.*

Example 3.1. Let

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

Then

$$A^2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad A^3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and $A^4 = O$. Hence this A is nilpotent of degree 4.

Therefore if $J_{\lambda,l}$ is an elementary Jordan block of size l with the eigenvalue λ, that is

$$J_{\lambda,l} = \begin{pmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \lambda & 1 \\ 0 & & & \lambda \end{pmatrix},$$

then

$$J_{\lambda,l} - \lambda I_l = J_{0,l} = \left.\begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & 0 & 1 \\ 0 & & & 0 \end{pmatrix} \qquad \right\} l \text{ rows}$$

and this is nilpotent of degree l.

Now consider the matrix J having the Jordan normal form (3.1). Denote by $\sigma(J)$ the set of all distinct numbers amongst $\lambda_1, \ldots, \lambda_k$. Then $\sigma(J)$ is set of all distinct eigenvalues of J. For each $\lambda \in \sigma(J)$ denote by l_λ the largest of those numbers l_i where $\lambda_i = \lambda$. Put

$$p(x) = \prod_{\lambda \in \sigma(J)} (x-\lambda)^{l_\lambda}.$$

We need to prove that the minimal polynomial $d_J(x) = p(x)$.

Fix any $\lambda \in \sigma(J)$ and consider any Jordan block $J_i = J_{\lambda_i,l_i}$ such that $\lambda_i = \lambda$. Then

$$p(x) = (x-\lambda)^{l_\lambda} \cdot \prod_{\mu \in \sigma(J)\setminus\{\lambda\}} (x-\mu)^{l_\mu}$$

and

$$\begin{aligned} p(J_{\lambda_i,l_i}) &= (J_{\lambda_i,l_i} - \lambda\, I_{l_i}))^{l_\lambda} \cdot \prod_{\mu \in \sigma(J)\setminus\{\lambda\}} (J_{\lambda_i,l_i} - \mu\, I_{l_i})^{l_\mu} \\ &= J_{0,l_i}^{l_\lambda} \cdot \prod_{\mu \in \sigma(J)\setminus\{\lambda\}} J_{\lambda_i-\mu,l_i}^{l_\mu} = 0 \end{aligned}$$

because $l_i \le l_\lambda$ so that $J_{0,l_i}^{l_\lambda} = 0$ by the above argument. Thus $p(J_{\lambda_i,l_i}) = 0$ for every Jordan block of J. Hence $p(J) = 0$. So the polynomial $p(x)$ annihilates the matrix J.

It remains to prove that our $p(x)$ is minimal annihilating polynomial. Consider any other polynomial $q(x)$ which divides $p(x)$. Then

$$q(x) = \prod_{\lambda \in \sigma(J)} (x - \lambda)^{l'_\lambda}$$

for some exponents $l'_\lambda \le l_\lambda$ where at least one of inequalities is strict. Fix any such λ with $l'_\lambda < l_\lambda$. Take any Jordan block $J_i = J_{\lambda_i,l_i}$ such that $\lambda_i = \lambda$ and $l_i = l_\lambda$. That is, for our fixed λ we take the Jordan block of the maximal size. Then

$$q(J_{\lambda_i,l_i}) = J_{0,l_i}^{l'_\lambda} \cdot \prod_{\mu \in \sigma(J) \setminus \{\lambda\}} J_{\lambda_i - \mu, l_i}^{l'_\mu} \ne 0$$

because $l'_\lambda < l_\lambda = l_i$ and $J_{0,l_i}^{l'_\lambda} \ne 0$ by the above argument. Here we also use the observation that each matrix $J_{\lambda_i-\mu,l_i}$ is non-singular, as $\lambda_i \ne \mu$. Thus $q(J_{\lambda_i,l_i}) \ne 0$. Hence $q(J) \ne 0$.

Example 3.2. What are the algebraic, geometric, and minimal multiplicities of

$$A = \begin{pmatrix} 5 & 1 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 15 \end{pmatrix}$$

Solution. The eigenvalue 5 has algebraic multiplicity 2, geometric multiplicity 1 and minimal multiplicity 2. The eigenvalue 15 has algebraic multiplicity 1, geometric multiplicity 1 and minimal multiplicity 1. The eigenvectors are $(1,0,0)^T$ and $(0,0,1)^T$. The minimal polynomial is $(x-5)^2(x-15)$. Notice that

$$A - 5I = \left(\begin{array}{cc|c} 0 & 1 & 0 \\ 0 & 0 & 0 \\ \hline 0 & 0 & 10 \end{array}\right)$$

and

$$(A - 5I)^2 = \left(\begin{array}{cc|c} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \hline 0 & 0 & 10^2 \end{array}\right)$$

while

$$A - 15I = \left(\begin{array}{cc|c} -10 & 1 & 0 \\ 0 & -10 & 0 \\ \hline 0 & 0 & 0 \end{array}\right)$$

and so

$$(A - 5I)^2(A - 15I) = \left(\begin{array}{cc|c} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \hline 0 & 0 & 10^2 \end{array}\right) \left(\begin{array}{cc|c} -10 & 1 & 0 \\ 0 & -10 & 0 \\ \hline 0 & 0 & 0 \end{array}\right) = \left(\begin{array}{cc|c} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array}\right)$$

whereas

$$(A-5I)(A-15I)=\left(\begin{array}{cc|c}0&1&0\\0&0&0\\\hline 0&0&10\end{array}\right)\left(\begin{array}{cc|c}-10&1&0\\0&-10&0\\\hline 0&0&0\end{array}\right)=\left(\begin{array}{cc|c}0&-10&0\\0&0&0\\\hline 0&0&0\end{array}\right)$$

Determining Jordan normal form by the multiplicities. The two matrices A and J in (3.3) have the same eigenvalues. Moreover the algebraic, geometric and minimal multiplicities of the eigenvalues of A and J are respectively the same. This has been observed earlier in the current section.

On the other hand, for some (**not all**) Jordan normal form matrices J their inordered collections of pairs $(\lambda_1, l_1), \dots, (\lambda_k, l_k)$ are already determined by the algebraic, geometric and minimal multiplicities of the eigenvalues of J. Since these multiplicities are the same for A, we can find them for a given matrix A first, and then try to determine J.

Example 3.3. Find algebraic, geometric and minimal multiplicities of the eigenvalues of

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

Hence find the Jordan normal form of A.

Solution. The characteristic polynomial of A is $c_A(x)=-(x+2)^3$. So the only eigenvalue of A is -2, and the algebraic multiplicity of the eigenvalue is 3. Since

$$A+2I=\begin{pmatrix}1&1&-1\\0&0&0\\1&1&-1\end{pmatrix} \quad\text{and}\quad (A+2I)^2=\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix},$$

the minimal polynomial is $m_A=(x+2)^2$, and the minimal multiplicity is 2.

To find $\mathrm{Ker}(A+2I)$, we solve the system of equations

$$(A+2I)\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}1&1&-1\\0&0&0\\1&1&-1\end{pmatrix}\begin{pmatrix}x\\y\\x\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}.$$

Its solution can be written as

$$\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}a-b\\b\\a\end{pmatrix}=a\begin{pmatrix}1\\0\\1\end{pmatrix}+b\begin{pmatrix}-1\\1\\0\end{pmatrix}$$

for any $a,b\in\mathbb{C}$. Therefore,

$$\mathrm{Ker}(A+2I)=\mathrm{Sp}\left\{\begin{pmatrix}1\\0\\1\end{pmatrix},\begin{pmatrix}-1\\1\\0\end{pmatrix}\right\},$$

so that $\dim(\mathrm{Ker}(A+2I))=2$ and the geometric multiplicity is 2.

Since the geometric multiplicity of the only eigenvalue is 2, the Jordan normal form of A consists of two elementary Jordan blocks. Since the minimal multiplicity is 2, the maximum size of elementary Jordan blocks is 2. Therefore, the Jordan normal form of A has one elementary Jordan block of size 2 and one elementary Jordan block of size 1:

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

We have already used the notation $\sigma(J)$ for the Jordan normal form matrices. We will finish this section with a more general definition, to be used later on.

Definition 3.4. *For any matrix A with complex entries, the set of all pairwise distinct eigenvalues of A is called the* ***spectrum*** *of A and is denoted by $\sigma(A)$.*

4. Constructing the Jordan Normal Form

This proof of the Jordan Theorem is non-examinable. But you will need to know how to find, for any given matrix A with complex entries, a Jordan normal form matrix J and a coordinate change matrix P such that $A = P\,J\,P^{-1}$. This procedure is best learned by looking at the examples, and several of them are given in the next section. Before considering the examples, we will make a few general remarks on that procedure.

First consider the elementary Jordan block

$$J_{\lambda,l} = \begin{pmatrix} \lambda & 1 & & \lambda \\ & \ddots & \ddots & \\ & & \lambda & 1 \\ 0 & & & \lambda \end{pmatrix}$$

with any eigenvalue $\lambda \in \mathbb{C}$. This $l \times l$ matrix can be regarded as a linear transformation of the coordinate vector space $\mathbb{C}^l$. Let $e_1, \ldots, e_l$ be the vectors of the standard basis of $\mathbb{C}^l$. Then consider the linear transformation $J_{\lambda,l} - \lambda\,I_l = J_{0,l}$. Then

$$J_{\lambda,l} - \lambda\,I_l : e_l \mapsto e_{l-1} \mapsto \ldots \mapsto e_2 \mapsto e_1 \mapsto 0\,.$$

We can arrange the basis vectors into a column array called a **tower**:

$$\begin{matrix} e_l \\ e_{l-1} \\ \vdots \\ e_2 \\ e_1 \end{matrix}$$

Note that each application of $J_{\lambda,l} - \lambda\,I_l$ maps a basis vector one level down. The basis vector e_1 is mapped to the zero vector, but we do not show the zero level here.

More generally, consider the Jordan normal form matrix J of size $n \times n$ such that the eigenvalues λ are the same for all k Jordan blocks: $\lambda_1 = \ldots = \lambda_k$. Assume that the sizes of the diagonal blocks of J do not increase as one goes down the diagonal: $l_1 \geq \ldots \geq l_k$.

For each Jordan block we have its own tower. Let us put these k towers next after each other. We will get an array of vectors called a **pyramid**:

$$\begin{array}{cccc} e_{l_1} & & & \\ \vdots & e_{l_1+l_2} & & \\ \vdots & \vdots & \cdots & e_n \\ \vdots & \vdots & \cdots & \vdots \\ e_2 & e_{l_1+2} & \cdots & e_{n-l_k+2} \\ e_1 & e_{l_1+1} & \cdots & e_{n-l_k+1} \end{array} \tag{4.1}$$

Here to simplyfy the notation we use the equality $l_1 + \ldots + l_{k-1} = n - l_k$.

Note that like in the case of a single Jordan block considered earlier, each application of the linear transformation $J - \lambda\, I_n$ maps a basis vector one level down. The basis vectors $e_1, e_{l_1+1}, \ldots, e_{n-l_k+1}$ occuring at the level one of the pyramid are mapped to zero. In particular, the subspace $\operatorname{Ker}(J - \lambda\, I_n) \subseteq \mathbb{C}^n$ is spanned by the basis vectors occuring at the level one of the pyramid. More generally, for any $m = 1, 2, \ldots$ the subspace

$$\operatorname{Ker}(J - \lambda\, I_n)^m \subseteq \mathbb{C}^n \tag{4.2}$$

is spanned by the basis vectors occuring at the first m levels of the pyramid (4.1).

If we are given an arbitrary Jordan form matrix J with several distinct eigenvalues, for each $\lambda \in \sigma(J)$ we can separately consider the pyramid of those standard basis vectors of $\mathbb{C}^n$ which correspond to the Jordan blocks with the same eigenvalue λ. For this λ, the subspace (4.2) will be again spanned by by the basis vectors occuring at the first m levels of that particular pyramid. Thus we can sort the standard basis vectors by looking that the kernel subspaces (4.2) for each $\lambda \in \sigma(J)$ and every $m = 1, 2, \ldots$.

General procedure. Let A be any $n \times n$ matrix. We will first compute the characteristic polynomial $c_A(x)$ and hence determine the spectrum $\sigma(A)$. Then separately for each $\lambda \in \sigma(A)$ we will compute the subspaces

$$\operatorname{Ker}(A - \lambda\, I_n)^m \subseteq \mathbb{C}^n \tag{4.3}$$

for all $m = 1, 2, \ldots$. Note that

$$\operatorname{Ker}(A - \lambda\, I_n) \subseteq \operatorname{Ker}(A - \lambda\, I_n)^2 \subseteq \ldots$$

so that each subsequent kernel subspace contains all the preceding ones. By choosing certain basis vectors in these subspaces and collecting them together for all $\lambda \in \sigma(A)$ will finally construct a basis $v_1, \ldots, v_n$ of $\mathbb{C}^n$ such that for any index j

$$A\, v_j = \sum_{i=1}^{n} J_{ij}\, v_i \tag{4.4}$$

where $J = (J_{ij})_{i,j=1}^n$ is a Jordan normal form matrix.

In the particular case when the set $\sigma(A)$ consists of only one element $\lambda \in \mathbb{C}$, the basis vectors $v_1, \dots, v_n$ will be arranged as a pyramid similar to (4.1):

$$\begin{array}{cccc} v_{l_1} & & & \\ \vdots & v_{l_1+l_2} & & \\ \vdots & \vdots & \cdots & v_n \\ \vdots & \vdots & \cdots & \vdots \\ v_2 & v_{l_1+2} & \cdots & v_{n-l_k+2} \\ v_1 & v_{l_1+1} & \cdots & v_{n-l_k+1} \end{array} \tag{4.5}$$

In this case each application of the linear transformation $A - \lambda I_n$ will map a basis vector in (4.5) one level down. Hence for each $m = 1, 2, \dots$ the subspace (4.3) will be spanned by the basis vectors v_i occuring at the first m levels of the pyramid (4.5).

Note that in the case $\sigma(A) = \{\lambda\}$ considered above, it suffices to choose only the basis vectors $v_1, v_{l_1+l_2}, \dots, v_n$ occurring at the **top** of every tower of the pyramid (4.5). This is because the other basis vectors showing in (4.5) are obtained from the top ones by successive applications of the linear transoformation $A - \lambda I_n$. Also note that for any $m = 1, 2, \dots$ the number of basis vectors at the level m of the pyramid (4.5) is exactly

$$\dim \operatorname{Ker}(A - \lambda I_n)^m - \dim \operatorname{Ker}(A - \lambda I_n)^{m-1}. \tag{4.6}$$

For the matrix A with several distinct eigenvalues, for each $\lambda \in \sigma(A)$ we will separately find a pyramid of linearly independent vectors of $\mathbb{C}^n$. For this λ, the subspace (4.3) will be again spanned by the vectors occuring at the first m levels of that particular pyramid. By collecting together the vectors from the pyramids corresponding to all $\lambda \in \sigma(A)$ we will get the basis vectors $v_1, \dots, v_n$ in general. It will be called a **Jordan basis** for A.

The procedure outlined above will determine the Jordan normal form matrix J for any given A, and will provide a Jordan basis $v_1, \dots, v_n$ of $\mathbb{C}^n$ corresponding to A. Note that the choice of a Jordan basis is not unique. Having made any such choice, let us write the vectors $v_1, \dots, v_n$ it terms of their coordinates:

$$v_1 = \begin{pmatrix} p_{11} \\ \vdots \\ p_{n1} \end{pmatrix} \quad , \dots , \quad v_n = \begin{pmatrix} p_{1n} \\ \vdots \\ p_{nn} \end{pmatrix}.$$

By putting these n column vectors together, we get a certain $n \times n$ matrix

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{pmatrix}.$$

We claim that then $A = P\,J\,P^{-1}$ as required. Indeed, the equality (4.4) means that J is the matrix of the linear transformation A of $\mathbb{C}^n$ relative to the Jordan basis $v_1, \dots, v_n$. Denote this basis by $\mathcal{V}$. The matrix A itself is the matrix of the linear transformation A

of $\mathbb{C}^n$ relative to the standard basis

$$e_1 = \begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix} \quad , \ldots , \quad e_n = \begin{pmatrix} 0 \\ \vdots \\ 1 \end{pmatrix} .$$

Denote the standard basis by $\mathcal{E}$. Then P is the coordinate change matrix: by the definition of this matrix we have $P = C^{\mathcal{E}}_{\mathcal{V}}$ in the notation of Chapter **??**. Hence by using Lemma **??**

$$A = M_{\mathcal{E}}(A) = C^{\mathcal{E}}_{\mathcal{V}}\, M_{\mathcal{V}}(A)\, (\, C^{\mathcal{E}}_{\mathcal{V}}\,)^{-1} = P\, J\, P^{-1} .$$

5. Four Examples

In this section we will find the Jordan normal form matrix J and the coordinate change matrix P for four matrices, illustrating the method described in the previous section. Along the way we will find their eigenvalues, eigenvectors, characteristic and minimal polynomials, and geometric, algebraic and minimal multiplicities of their eigenvalues.

Example A.

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

First we must find the eigenvalues, and so we calculate the characteristic polynomial:

$$c_A(x) = \begin{vmatrix} -x & 0 & 0 & 0 & 0 \\ 0 & -x & 0 & 1 & 0 \\ 0 & 0 & -x & 1 & 0 \\ -1 & -1 & 1 & -x & 0 \\ 0 & 0 & 0 & 0 & -x \end{vmatrix} .$$

Expanding this determinant in the first row and then in the last column, we get

$$(-x)^2 \begin{vmatrix} -x & 0 & 1 \\ 0 & -x & 1 \\ -1 & 1 & -x \end{vmatrix} .$$

Expanding in the first row again, we get

$$(-x)^2 \left(-x \begin{vmatrix} -x & 1 \\ 1 & -x \end{vmatrix} + 1 \cdot \begin{vmatrix} 0 & -x \\ -1 & 1 \end{vmatrix} \right) = (-x)^2 \left(-x(x^2-1) - x \right) = -x^5 .$$

So, $c_A(x) = -x^5$ and $\sigma(A) = \{0\}$. This eigenvalue has algebraic multiplicity $a_0 = 5$.

Now we start to look for the Jordan basis. We first form an initial basis of $\operatorname{Ker} A$, extend it to $\operatorname{Ker} A^2$, and so on until we have the basis for $\mathbb{C}^5$. We have

$$\operatorname{Ker} A = \left\{ \begin{pmatrix} a \\ b \\ c \\ d \\ e \end{pmatrix} : \ d = 0, \ -a - b + c = 0 \right\}.$$

We get this directly from A: the first and fourth row give trivial equations on a, b, c, d, e, the second and third give $d = 0$, and the fourth one produces $-a - b + c = 0$. Next,

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

Hence

$$\operatorname{Ker} A^2 = \left\{ \begin{pmatrix} a \\ b \\ c \\ d \\ e \end{pmatrix} : \ -a - b + c = 0 \right\}$$

Note that $\operatorname{Ker} A$ as described above is indeed a subspace of $\operatorname{Ker} A^2$. Finally, $A^3 = 0$ and

$$\operatorname{Ker} A^3 = \mathbb{C}^5 .$$

According to formula (4.6) the pyramid (4.5) for our matrix A has there levels. The levels $1, 2$ and 3 of it contain respectively

$$3, \ 4 - 3 = 1 \quad \text{and} \quad 5 - 4 = 1$$

basis vectors respectively. Hence the pyramid is

$$\begin{array}{lll} v_3 & & \\ v_2 & & \\ v_1 & v_4 & v_5 \end{array}$$

In particular, there are exactly three towers in this pyramid.

Let us read off some data from this pyramid before choosing the Jordan basis. We have seen that $\sigma(A) = \{0\}$. Hence $a_0 = 5$ is the total number of vectors in the pyramid. Now we see that $g_0 = 3$ is the number of towers of the pyramid, which is equal to the number of vectors at the lowest level, which is the dimension of the eigenspace of A associated to 0. We also see that $m_0 = 3$ is the number of non-empty levels in the pyramid, which is equal to the size of the largest tower. The minimal polynomial is $d_A(x) = x^3$.

We can also read off the Jordan normal form from this pyramid. There is only one eigenvalue 0. The sizes of the elementary Jordan blocks are the sizes of towers in the pyramid: these are $3, 1, 1$. The Jordan form of the matrix A is thus

$$J = \begin{pmatrix} \boxed{\begin{matrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{matrix}} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \end{matrix} & \boxed{0} & 0 \\ \begin{matrix} 0 & 0 & 0 \end{matrix} & 0 & \boxed{0} \end{pmatrix}.$$

Let us now find the basis in which A takes the form J. We do it by choosing the top basis vector for every tower, and by making sure that all our basis vectors are indeed linearly independent. We will start with choosing any vector v_3 so that $v_3 \in \operatorname{Ker} A^3$ but $v_3 \notin \operatorname{Ker} A^2$. Let us choose

$$v_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Then

$$v_2 = A\, v_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -1 \\ 0 \end{pmatrix} \quad \text{and} \quad v_1 = A\, v_2 = \begin{pmatrix} 0 \\ -1 \\ -1 \\ 0 \\ 0 \end{pmatrix}.$$

The vectotrs v_4 and v_5 can be now chosen to satisfy the only condition that together with v_1 they make a basis of $\operatorname{Ker} A$. Let us choose

$$v_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \qquad v_5 = \begin{pmatrix} 2 \\ -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

The change of basis matrix is

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

Now let us make a remark about uniqueness. The matrix J is unique up to reordering of the elementary Jordan blocks. But the matrix P is very not unique. For instance, the basis vector v_3 could have been replaced by any vector in $\operatorname{Ker} A^3$ which is not in $\operatorname{Ker} A^2$. Equivalently, it could have been replaced by any $v = av_3 + u$ where $a \in \mathbb{C}$ is not zero and

$u \in \operatorname{Ker} A^2$. Then v_2 would be replaced by Av, and v_1 by $A^2 v = aA^2 v_3$. The vectors v_4 and v_5 could be just any vectors in $\operatorname{Ker} A$ which together with $A^2 v$ make a basis of $\operatorname{Ker} A$.

Example B.

$$B = \begin{pmatrix} 10 & -4 & 0 \\ 1 & 5 & 9 \\ -1 & 1 & 9 \end{pmatrix}$$

and let us assume that we already know that the the characteristic polynomial of B is

$$c_B(x) = -(x-6)^2\,(x-12)\,.$$

Thus we know the eigenvalues of the matrix are 6 and 12 with the algebraic multiplicities $a_6 = 2$ and $a_{12} = 1$ respectively. Then we also have $g_{12} = 1$ so that the eigenspace V_{12} is one-dimensional. Let us find this eigenspace first. We have

$$B - 12\,I = \begin{pmatrix} -2 & -4 & 0 \\ 1 & -7 & 9 \\ -1 & 1 & -3 \end{pmatrix}.$$

Divide the first row by -2 and use it to eliminate everything in the first column:

$$\begin{pmatrix} -2 & -4 & 0 \\ 1 & -7 & 9 \\ -1 & 1 & -3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 2 & 0 \\ 1 & -7 & 9 \\ -1 & 1 & -3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 2 & 0 \\ 0 & -9 & 9 \\ 0 & 3 & -3 \end{pmatrix}$$

Divide the second row by -9, third by 3, and use second row to eliminate everything in the second column:

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

Therefore

$$V_{12} = \operatorname{Ker}(B - 12I) = \left\{ \begin{pmatrix} a \\ b \\ c \end{pmatrix} \;:\; a + 2b = 0,\; b - c = 0 \right\} = \left\{ \begin{pmatrix} -2c \\ c \\ c \end{pmatrix} \;:\; c \in \mathbb{C} \right\}.$$

As a basis vector of $\operatorname{Ker}(B - 12I)$ we can now choose

$$v_1 = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}.$$

The pyramid for to the eigenvalue 12 will then consist of this single vector. We also have $m_{12} = 1$, and the only elementary Jordan block with this eigenvalue is a 1×1 matrix

$$(12)\,.$$

Now we move on to the eigenvalue 6. Finding $\mathrm{Ker}(B-6I)$:

$$B-6I=\begin{pmatrix} 4 & -4 & 0 \\ 1 & -1 & 9 \\ -1 & 1 & 3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & -1 & 0 \\ 1 & -1 & 9 \\ -1 & 1 & 3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 9 \\ 0 & 0 & 3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

and

$$\mathrm{Ker}(B-6I)=\left\{ \begin{pmatrix} a \\ b \\ c \end{pmatrix} \;:\; a-b=0,\; c=0 \right\} = \left\{ \begin{pmatrix} b \\ b \\ 0 \end{pmatrix} \;:\; b\in\mathbb{C} \right\}.$$

Therefore $g_6=1$ and there is only one basis vector at the level 1 of the pyramid for this eigevalue. This pyramid consists of only one tower. The size of this tower is then $a_6=2$. Hence $m_6=2$ and there is only one elementary Jordan block with this eigenvalue:

$$\begin{pmatrix} 6 & 1 \\ 0 & 6 \end{pmatrix}.$$

The pyramid itself has the form

$$\begin{matrix} v_3 \\ v_2 \end{matrix}$$

To determine the Jordan basis vectors v_2 and v_3 we compute $\mathrm{Ker}(B-6I)^2$:

$$(B-6I)^2=\begin{pmatrix} 12 & -12 & -36 \\ -6 & 6 & 18 \\ -6 & 6 & 18 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & -1 & -3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\mathrm{Ker}(B-6I)^2=\left\{ \begin{pmatrix} a \\ b \\ c \end{pmatrix} \;:\; a-b-3c=0 \right\} = \left\{ \begin{pmatrix} b+3c \\ b \\ c \end{pmatrix} \;:\; b,c\in\mathbb{C} \right\}.$$

Note that $\mathrm{Ker}(B-6I)\subset\mathrm{Ker}(B-6I)^2$. We can choose v_3 to be any vector in $\mathrm{Ker}(B-6I)^2$ such that $v_3\notin\mathrm{Ker}(B-6I)$. Let us choose

$$v_3=\begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}.$$

Then

$$v_2=(B-6I)\,v_3=\begin{pmatrix} 12 \\ 12 \\ 0 \end{pmatrix}.$$

The minimal polynomial

$$d_B(x)=(x-6)^2\,(x-12)$$

while

$$J=\begin{pmatrix} 12 & 0 & 0 \\ 0 & 6 & 1 \\ 0 & 0 & 6 \end{pmatrix} \quad \text{and} \quad P=\begin{pmatrix} -2 & 12 & 3 \\ 1 & 12 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Example C.

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

Calculate the characteristic polynomial:

$$c_C(x) = \begin{vmatrix} 2-x & 0 & 0 & -1 & 0 \\ 0 & 3-x & 0 & 1 & 0 \\ 0 & -1 & 1-x & 1 & -1 \\ 0 & 0 & 0 & 3-x & 0 \\ 1 & 2 & 2 & 0 & 4-x \end{vmatrix}.$$

Expanding this determinant along the fourth row, we get

$$(3-x)\begin{vmatrix} 2-x & 0 & 0 & 0 \\ 0 & 3-x & 0 & 0 \\ 0 & -1 & 1-x & -1 \\ 1 & 2 & 2 & 4-x \end{vmatrix}.$$

Expanding along the first row, and then again first row, we get

$$(3-x)\,(2-x)\begin{vmatrix} 3-x & 0 & 0 \\ -1 & 1-x & -1 \\ 2 & 2 & 4-x \end{vmatrix} = (3-x)^2\,(2-x)\begin{vmatrix} 1-x & -1 \\ 2 & 4-x \end{vmatrix} =$$

$$(3-x)^2\,(2-x)\,(x^2-5x+4+2) = (3-x)^3\,(2-x)^2.$$

So the eigenvalues of C are 2 and 3 with the algebraic multiplicities $a_2 = 2$ and $a_3 = 3$. We will now consider these two eigenvalues separately. We will get two separate pyramids, with the total size of the pyramid associated to the eigenvalue λ being a_λ.

We start with $\lambda = 2$:

$$\mathrm{Ker}(C-2I) = \mathrm{Ker}\begin{pmatrix} 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & -1 & -1 & 1 & -1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 2 & 2 & 0 & 2 \end{pmatrix}.$$

Reducing,

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

which has the first column looking like the first column of an upper triangular matrix. Then we reduce the second column using the second row and get

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

Reading from the last row up, the kernel $\mathrm{Ker}(C-2I)$ will contain the vector

$$\begin{pmatrix} a \\ b \\ c \\ d \\ e \end{pmatrix}$$

if and only if $d=0$, then $-c+2d-e=0$ so that $c=-e$, then $b+d=0$ so that $b=0$, and finally $a+2b+2c+2e=0$ so that $a=0$. The set of solutions is

$$\mathrm{Ker}(C-2I) = \left\{ \begin{pmatrix} 0 \\ 0 \\ -e \\ 0 \\ e \end{pmatrix} : \; e \in \mathbb{C} \right\}$$

Next,

$$(C-2I)^2 = \begin{pmatrix} 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 2 & 0 \\ -1 & -2 & -1 & -1 & -1 \\ 0 & 0 & 0 & 1 & 0 \\ 2 & 4 & 2 & 3 & 2 \end{pmatrix} \rightsquigarrow \begin{pmatrix} -1 & -2 & -1 & -1 & -1 \\ 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 2 & 4 & 2 & 3 & 2 \end{pmatrix}$$

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

Hence $\mathrm{Ker}(C-2I)^2$ equals

$$\left\{ \begin{pmatrix} a \\ b \\ c \\ d \\ e \end{pmatrix} : \; -a-2b-c-d-e=0, \; b+2d=0, \; d=0 \right\} = \left\{ \begin{pmatrix} -c-e \\ 0 \\ c \\ 0 \\ e \end{pmatrix} : \; c,e \in \mathbb{C} \right\}.$$

This is two-dimensional. As $a_2 = 2$ the corresponding pyramid consists of a single tower of size 2:

$$\begin{matrix} v_2 \\ v_1 \end{matrix}$$

Note that $\mathrm{Ker}(C-2I) \subset \mathrm{Ker}(C-2I)^2$. We can choose v_2 to be any vector in $\mathrm{Ker}(C-2I)^2$ such that $v_2 \notin \mathrm{Ker}(C-2I)$. Let us choose

$$v_2 = \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

Then

$$v_1 = (C-2I)\, v_2 = \begin{pmatrix} 0 \\ 0 \\ -2 \\ 0 \\ 2 \end{pmatrix}.$$

We conclude that $m_2 = 2$, $g_2 = 1$ and the only elementary Jordan block for $\lambda = 2$ is

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

Now let us deal with the eigenvalue $\lambda = 3$. Finding $\mathrm{Ker}(C-3I)$:

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

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

Hence

$$\mathrm{Ker}(C-3I) = \left\{ \begin{pmatrix} 0 \\ 0 \\ -e \\ 0 \\ 2e \end{pmatrix} \; : \; e \in \mathbb{C} \right\}.$$

In particular, the space $\operatorname{Ker}(C-3I)$ is one-dimensional. So the pyramid will consist of only one tower. The size of this tower will be $a_3 = 3$. Finding $\operatorname{Ker}(C-3I)^2$:

$$(C-3I)^2 = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & -3 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & 3 & -1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 2 & -3 & 1 \\ 0 & 0 & -2 & 3 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\rightsquigarrow \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & -2 & 1 \\ 0 & 0 & -2 & 3 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & -2 & 1 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

so that

$$\operatorname{Ker}(C-3I)^2 = \left\{ \begin{pmatrix} 0 \\ b \\ -e \\ 0 \\ 2e \end{pmatrix} \;:\; b, e \in \mathbb{C} \right\}.$$

Finally, finding $\operatorname{Ker}(C-3I)^3$:

$$(C-3I)^3 = \begin{pmatrix} -1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & -2 & 3 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & -2 & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} -1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -1 & 1 \end{pmatrix}$$

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

Hence

$$\operatorname{Ker}(C-3I)^3 = \left\{ \begin{pmatrix} -2d \\ b \\ d-e \\ 2d \\ 2e \end{pmatrix} \;:\; b, d, e \in \mathbb{C} \right\}.$$

The pyramid corresponding to $\lambda = 3$ will have the form

$$\begin{matrix} v_5 \\ v_4 \\ v_3 \end{matrix}$$

Note that
$$\operatorname{Ker}(C-3I) \subset \operatorname{Ker}(C-3I)^2 \subset \operatorname{Ker}(C-3I)^3 .$$
We can choose v_5 to be any vector in $\operatorname{Ker}(C-3I)^3$ such that $v_5 \notin \operatorname{Ker}(C-3I)^2$. Choose
$$v_5 = \begin{pmatrix} -2 \\ 0 \\ 1 \\ 2 \\ 0 \end{pmatrix}.$$
Then
$$v_4 = (C-3I)\,v_5 = \begin{pmatrix} 0 \\ 2 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad v_3 = (C-3I)\,v_4 = \begin{pmatrix} 0 \\ 0 \\ -2 \\ 0 \\ 4 \end{pmatrix}.$$
So $g_3 = 1$ and $m_3 = 3$ while the elementary Jordan block with the eigenvalue 3 is
$$\begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}.$$
The minimal polynomial:
$$d_C(x) = (x-2)^2(x-3)^3 .$$
We found
$$J = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix} \quad \text{and} \quad P = \begin{pmatrix} 0 & -2 & 0 & 0 & -2 \\ 0 & 0 & 0 & 2 & 0 \\ -2 & 1 & -2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 2 \\ 2 & 1 & 4 & 0 & 0 \end{pmatrix}.$$

Example D.
$$D = \begin{pmatrix} 2 & 1 & 3 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$
We see that $c_D(x) = (2-x)^3$ so that the only eigenvalue of the matrix D is 2 occuring with the algebraic multiplicity 3. Finding $\operatorname{Ker}(D-2I)$:
$$\operatorname{Ker}(D-2I) = \operatorname{Ker}\begin{pmatrix} 0 & 1 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \left\{ \begin{pmatrix} a \\ -3c \\ c \end{pmatrix} \,:\, a, c \in \mathbb{C} \right\}.$$
Hence $g_2 = 2$ and the pyramid corresponding to this eigenvalue will have 2 towers. The total size of this pyramid is $a_2 = 3$ so it must have the form
$$\begin{matrix} v_2 & \\ v_1 & v_3 \end{matrix}$$

So $m_2 = 2$ and $d_D(x) = (x-2)^2$. To construct a Jordan basis we will choose the vector v_2 first. Note that

$$\operatorname{Ker}(D-2I)^2 = \operatorname{Ker} 0 = \mathbb{C}^3 .$$

We can choose v_2 to be any vector in $\mathbb{C}^3$ such that $v_2 \notin \operatorname{Ker}(D-2I)$. Choose

$$v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} .$$

Then

$$v_1 = (D-2I)\, v_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} .$$

It remains to choose $v_3 \in \operatorname{Ker}(D-2I)$ subject to the only requirement that v_1 and v_3 make a basis in $\operatorname{Ker}(D-2I)$. Let us choose

$$v_3 = \begin{pmatrix} 0 \\ -3 \\ 1 \end{pmatrix} .$$

Then we get

$$J = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \quad \text{and} \quad P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{pmatrix} .$$