Core Mathematics III

University of Dundee

Prof Gunnar Hornig • ghornig@dundee.ac.uk

2025-11-06

Table of contents

Introduction

Welcome to MA21001 at the University of Dundee.

These notes are available at dundeemath.github.io/MA21001/ as HTML and also as a PDF.

Licence

Part I

Algebra

1 Before we start ...

1.1 Linear Algebra

Algebra: The department of mathematics which investigates the relations and properties of numbers by means of general symbols; and, in a more abstract sense, a calculus of symbols combining according to certain defined laws (Oxford English Dictionary)

Linear algebra is the branch of mathematics concerned with the study of vectors, vector spaces (also called linear spaces), linear transformations, and systems of linear equations. Vector spaces are a central theme in modern mathematics; thus, linear algebra is widely used in both abstract algebra and functional analysis. Linear algebra also has a concrete representation in analytic geometry and is generalised in operator theory. It has extensive applications in the natural sciences and the social sciences since a linear model can often approximate nonlinear models.

The name Algebra (from Arabic: al-jabr) is derived from the treatise written by the Persian mathematician Muhammad ibn Musa al-Kwarizmi titled *Al-Kitab al-Jabr wa-l-Muqabala*, meaning “The Compendious Book on Calculation by Completion and Balancing”, which provided symbolic operations for the systematic solution of linear and quadratic equations [1].

1.2 The Language of Mathematics

To formulate mathematics, we first have to define the notions we are using before we make any statement. This allows us to prove general statements, which is the essence of mathematics. Mathematical notions are defined in Axioms and Definitions, while statements are called Theorems or Corollaries:

1.2.1 Axioms

a fundamental definition or statement which can’t be derived from any simpler statement. Axioms are the starting points of a mathematical theory. Often, axioms are so ”self-evident” that they are not explicitly mentioned.

Example: ”Peano Axioms” for the natural numbers.

1. 0 is a natural number.
2. For every natural number n, the successor, S(n), is a natural number.
3. For every natural number n, $S(n) \neq 0$. That is, there is no natural number whose successor is 0.
4. For all natural numbers m and n, if S(m) = S(n), then m = n. That is, S is an injection.

[1] From Wikipedia, the free encyclopaedia

The original Peano axioms (published in 1889 by the Italian mathematician Giuseppe Peano) also included a set of axioms for the notion of equality and one further "axiom of induction".

1. For every natural number x, x = x. That is, equality is reflexive.
2. For all natural numbers x and y, if x = y, then y = x. That is, equality is symmetric.
3. For all natural numbers x, y and z, if x = y and y = z, then x = z. That is, equality is transitive.
4. For all a and b, if a is a natural number and a = b, then b is also a natural number. That is, the natural numbers are closed under equality.
5. If K is a set such that 0 is in K, and for every natural number n, if n is in K, then S(n) is in K, then K contains every natural number. (axiom of induction)

1.2.2 Definition

A new mathematical notion is given (defined) in terms of existing notions.

Example: *A rational number is any number that can be expressed as the quotient a/b of two integers, with the denominator b not equal to zero.* Note that this definition requires to know what integers are and the notion of division (quotient).

1.2.3 Theorem

A statement which relates specific (perviously defined) mathematical properties. It usually requires proof using definitions and mathematical logic to show the statement is true.

1.2.4 Corollary, Lemma

An immediate consequence of a theorem (Corollary) or a minor theorem (Lemma). Sometimes, it comes with a proof; sometimes, it is omitted because it is considered trivial.

Example 1.1. As an example let's define something new :

Definition 1.1. An umpf is a natural number, which is a multiple of 3. A gumpf is a natural number which is a multiple of 6. (These are made-up notions, just for demonstration)

Lemma 1.1. *Every gumpf is an umpf.*

Proof. If x is a gumpf, then $x = 6n$, where n is a natural number. Hence $x = 3 \cdot 2 \cdot n = 3 \cdot (2 \cdot n)$. Therefore, x is an umpf. □ □

The end of a proof is often indicated by a square or the abbreviation q.e.d. (Latin: quod erat demonstrandum) "that which was to be shown".

1.3 Mathematical notation

Symbols

$\mathbb{N}$: natural numbers, e.g.$1, 2, ...$
$\mathbb{Z}$: integers: $.., -2, -1, 0, 1, 2, ...$
$\mathbb{Q}$: rational numbers: $0, 1/3, -5/11, ...$
$\mathbb{R}$: real numbers: $0, \pi, \sqrt{2}, 2/3, -2, ...$
$\mathbb{C}$: complex numbers: $0, i, \pi + i, \sqrt{2} - 4i, -2/3, ...$
$\Rightarrow$ logical implication, $A \Rightarrow B$: if A then B
$\Leftrightarrow$ logical equivalence, $A \Leftrightarrow B$: if A then B and if B then A
$\vee$ logical *or*, $A \vee B$: either A or B or both
$\wedge$ logical *and*, $A \wedge B$: A and B
$\in$ *is an element of*, $x \in \mathbb{R}$: x is an element of $\mathbb{R}$
$\forall$ *for all*, e.g.: $x^n \in \mathbb{R}\ \forall\, n$: x^n is real for all n
$\exists$ *there exist(s)*, e.g.: $\forall x \in \mathbb{Q}\ \exists\ p, q \in \mathbb{Z}$ with $x = p/q$

See also the list of Greek letters in the Appendix Section 6.1.

1.4 Mathematical proofs

A sequence of logical arguments which shows that a statement is true in all cases, without a single exception. Types of proofs are:

- direct proof.
  Example: The product of two even numbers is even. Proof: Let a,b two even numbers, that is $a = 2n$ and $b = 2m$, where $m, n \in \mathbb{Z}$. $\Rightarrow ab = 2n2m = 2(2nm)$ is an even number.

- proof by induction.
  This type of proof is often used to prove an infinite series of statements A_n, $n = 1, 2, 3 \ldots$. If we can prove that a) A_1 is true and b) if A_n is true then A_{n+1} is true, then A_n is true for all $n \in \mathbb{N}$.

  Example: $\sum_{n=1}^{N} n = \frac{1}{2}N(N + 1)$

  Proof.

  a) The statement is true for $n = 1$: $1 = \frac{2}{2}$. b) If the formula is true for N=k then it is also true for N=k+1: $\sum_{n=1}^{N} n = \frac{1}{2}N(N + 1) \Rightarrow \sum_{n=1}^{N+1} n = \frac{1}{2}N(N + 1) + (N + 1) = (N + 2)(N + 1)/2.$ □

  □

- proof by contradiction. The type of proof uses $\neg(\neg A) = A$; that is, it is shown that the opposite statement $\neg A$ is false (or $\neg(\neg A)$ is true) and hence A is true.
  A famous example is the proof that $\sqrt{2}$ is not a rational number: A: "$\sqrt{2}$ is not a rational number". Let us assume $\neg A$ is true, that is, $\sqrt{2}$ is a rational number. Then $\sqrt{2} = p/q$ with $p, q \in \mathbb{Z}$ with no common divisior. $\Rightarrow 2q^2 = p^2 \Rightarrow p^2$ is even and therefore p is even. Then $p = 2n \Rightarrow p^2 = 2^2n^2$ and

hence $q^2 = p^2/2 = 2n^2$ is even. Therefore, q must be even in contradiction to the assumption that p, q had no common divisor.

2 Vector spaces

2.1 Motivation

You have used vectors in previous modules to do elementary geometry. These vectors were introduced as lines connecting two points with a direction:

$$\vec{a} = \overrightarrow{PQ},$$

where P is the starting point of the vector and Q is the endpoint. The space of all these vectors is called the Euclidean space. It is usually introduced as a two- or three-dimensional space, but it is possible to generalise this to more dimensions. The Euclidean space was introduced as a simple model for the physical space around us. One can furthermore introduce orthogonal coordinates for the points in this space, e.g.

$$P = (P_x, P_y, P_z) = (3, 2, 1).$$

and this then allows us to define the components of a vector:

$$\vec{a} = \overrightarrow{PQ} = \begin{pmatrix} Q_x - P_x \\ Q_y - P_y \\ Q_z - P_z \end{pmatrix}$$

The identification between points and vectors in an Euclidean space becomes particularly simple if we can choose P to be the origin of the coordinate system, $P = (0, 0, 0)$,

$$\vec{a} = \overrightarrow{OQ} = \begin{pmatrix} Q_x \\ Q_y \\ Q_z \end{pmatrix}.$$

Note that while the coordinates of points are usually denoted by a row of numbers (tuplet), the components of vectors are written as a column. An alternative way to write a vector is to write it as a sum of multiples of unit vectors,

$$\vec{v} = v_x \mathbf{e}_x + v_y \mathbf{e}_y + v_z \mathbf{e}_z = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}.$$

An addition as well as a scalar multiplication was defined for these vectors. Later, the scalar product between two vectors and the cross-product (vector product) were also defined.

The operations on vectors of this Euclidean space E, the addition and scalar multiplication, had certain properties, for instance

- $\vec{u} + \vec{v} \in E$, closure condition for addition
- $\vec{u} + \vec{v} = \vec{v} + \vec{u}$, the addition is commutative
- $\lambda(\vec{u} + \vec{v}) = \lambda\vec{u} + \lambda\vec{v}$, distributive law
- …

In the following, we will generalise this space and define an abstract notion of a vector space, which has a much wider range of applications. To understand the crucial properties of such a definition, consider two examples.

Example 2.1. The captain of a sailing boat can calculate the velocity of his ship (speed and direction over ground) as the sum of two velocities: the velocity of his boat relative to the water (speed through water) and the velocity of the water (drift due to tides or ocean currents). The sum of the two velocities is found in the same way as above, as the addition of two vectors; however, the coordinate system used is usually a polar one (the angle as determined by a compass and speed in knots) rather than a Cartesian one (x, y-components). The velocities obtained this way are part of a *velocity space*. The elements of this space, the velocities, follow the same rules of addition and scalar multiplication as the vectors in our Euclidean space. Remark: In this example, the velocity space is a tangent space to a manifold (the surface of the ocean), a concept explained in more detail in Differential Geometry.

Example 2.2. The set of all polynomials with degree $\leq n$ is denoted by $\mathscr{P}_n$. We can add polynomials in $\mathscr{P}_n$ to obtain a polynomial in $\mathscr{P}_n$ again. Polynomials $p(x) = a_0+a_1x+...+a_nx^n$ and $q(x) = b_0+b_1x+...+b_nx^n \in \mathscr{P}_n$ add up to

$$p(x) + q(x) = (a_0 + b_0) + (a_1 + b_1)x + ... + (a_n + b_n)x^n \in \mathscr{P}_n.$$

Similarly, we can multiply polynomials by a scalar

$$\lambda p(x) = \lambda a_0 + \lambda a_1 x + ... + \lambda a_n x^n \in \mathscr{P}_n$$

Note that we can identify a polynomial of degree $\leq n$ with the vector of coefficients

$$p(x) \sim \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

The two examples above illustrate that numerous diverse examples exist where we find a structure similar to the Euclidean space. The elements in these spaces often don't look like "vectors"; they can be polynomials, velocities, matrices, etc.~, but addition and scalar multiplication work in the same way, so they follow the same rules as vectors. This is the motivation to define an abstract notion of a vector space and define it based on how its elements behave under addition and scalar multiplication, rather than what they represent.

2.2 Definition of a Vector Space

Definition 2.1: Vector Space over R

A vector space over $\mathbb{R}$ is a set V, the elements of which are called vectors $\vec{v} \in V$ along with two operations, an addition $\vec{v} + \vec{w}$ and a scalar multiplication $a\vec{v}$, $a \in \mathbb{R}$, subject to the following axioms for all $\vec{u}, \vec{v}, \vec{w} \in V$ and $a, b \in \mathbb{R}$:

$$\vec{u} + \vec{v} \in V \tag{2.1}$$

$$a\vec{u} \in V \tag{2.2}$$

$$\vec{u} + \vec{v} = \vec{v} + \vec{u} \tag{2.3}$$

$$(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w}) \tag{2.4}$$

$$\text{there exists a vector } \vec{0} \in V \text{ such that } \vec{u} + \vec{0} = \vec{0} + \vec{u} = \vec{u} \tag{2.5}$$

$$\text{given } \vec{v} \text{ there exists a vector } -\vec{v} \in V \text{ such that } \vec{v} + (-\vec{v}) = \vec{0} \tag{2.6}$$

$$1\vec{u} = \vec{u} \tag{2.7}$$

$$a(\vec{u} + \vec{v}) = a\vec{u} + a\vec{v} \tag{2.8}$$

$$(a + b)\vec{u} = a\vec{u} + b\vec{u} \tag{2.9}$$

$$a(b\vec{u}) = (ab)\vec{u} \tag{2.10}$$

Axiom 2.1 and 2.2 ensure that both operations do not lead to elements not in V, 2.3 to 2.6 ensure that the addition is commutative and associative and that there is a neutral element $\vec{0}$ and an inverse element with respect to the addition. Axiom 2.7 states that the neutral element of the multiplication in $\mathbb{R}$ is also the neutral element for the scalar multiplication, while the remaining conditions are distributive laws.

Remark on Notation

Alternative notations for vectors encountered in the literature are bold face symbols $\mathbf{v}$ or $\bar{v}$ (not to be confused with complex conjugation) or $\underline{v}$. The zero vector $\vec{0}$ is often written just as 0. Note that there are two kinds of additions in this definition, both denoted by the same symbol $+$. The addition of two real numbers $(a + b)$ and the addition of two vectors $\vec{v} + \vec{w}$. Also, for the multiplication, we have $ab \in \mathbb{R}$ as well as $a\vec{v} \in V$. No symbol is used for multiplication, since the dot as well as the cross will be later used to define two different types of multiplication between two vectors.
If we want to avoid writing a vector as a column, as this always takes up a lot of space, we can use the transpose (denoted by a superscript T):

$$(1, 2)^T = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Example 2.3. The set $\mathbb{R}^2 = \{(x_1, x_2)^T | x_1 \in \mathbb{R}, x_2 \in \mathbb{R}\}$ is a vector space, the elements of which are written as columns $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, if addition and multiplication are defined as

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \end{pmatrix}; \qquad a\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 \\ ax_2 \end{pmatrix}.$$

We have to check that all the axioms 2.1 to 2.10 are satisfied.

- axioms 2.1 and 2.2 are satisfied because the elements are again element in V.
- axioms 2.3 to 2.4 are satisfied because the addition of real numbers in each entry of the vector is associative and commutative,
- 2.5 and 2.6 are correct if we use the following zero vector and negative element:

$$\vec{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad -\vec{x} = \begin{pmatrix} -x_1 \\ -x_2 \end{pmatrix}.$$

- axiom 2.7 is obvious from the definition of the scalar multiplication,
- axiom 2.8

$$(r + s)\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} (r + s)v_1 \\ (r + s)v_2 \end{pmatrix} = \begin{pmatrix} rv_1 + sv_1 \\ rv_2 + sv_2 \end{pmatrix} = r\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + s\begin{pmatrix} v_1 \\ v_2 \end{pmatrix};$$

- axiom 2.9

$$r\left(\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}\right) = \begin{pmatrix} r(v_1 + w_1) \\ r(v_2 + w_2) \end{pmatrix} = \begin{pmatrix} rv_1 + rw_1 \\ rv_2 + rw_2 \end{pmatrix} = r\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + r\begin{pmatrix} w_1 \\ w_2 \end{pmatrix};$$

- axiom 2.10

$$(rs)\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} (rs)v_1 \\ (rs)v_2 \end{pmatrix} = \begin{pmatrix} r(sv_1) \\ r(sv_2) \end{pmatrix} = r\left(s\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}\right).$$

Example 2.4. The set

$$V = \left\{ \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in \mathbb{R}^2 \middle| v_1^2 + v_2^2 \le 1 \right\}$$

with the addition and scalar multiplication as in the previous example is **not** a vector space. The first two axioms (closure conditions) are not satisfied. For example

$$\mathbf{u} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} \in V, \text{ since } \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{2} \le 1, \text{ but } \quad 2\mathbf{u} = 2\begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \notin V.$$

Example 2.5. The set

$$V = \left\{ \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \middle| v_1 \in \mathbb{R}, v_2 \in \mathbb{R} \right\}$$

with the same scalar multiplication as before

$$a\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 \\ ax_2 \end{pmatrix},$$

but the addition

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 + 1 \\ x_2 + y_2 + 1 \end{pmatrix}$$

is **not** a vector space. Axioms 8 and 9 are not satisfied.

Example 2.6. What we have shown for a vector space consisting of two vectors with two real components can be easily extended to n components. The vector space

$$I\!R^n = \left\{ \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \middle| v_i \in \mathbb{R}, i = 1, 2, .., n \right\}$$

together with the operation of addition

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{pmatrix},$$

and the scalar multiplication

$$\alpha\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} \alpha v_1 \\ \alpha v_2 \\ \vdots \\ \alpha v_n \end{pmatrix}, \quad \alpha \in \mathbb{R},$$

is called the vector space $\mathbb{R}^n$.

Although we haven't defined yet what a basis is, we mention here for future reference that the vectors,

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

form a basis of $\mathbb{R}^n$, the so called the standard basis. That is, every vector in $\mathbb{R}^n$ can be expressed as a linear combination of the basis vectors,

$$\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 + \ldots + x_n\vec{e}_n,$$

Or, in other words, the vector space is spanned by the basis vectors.

Example 2.7. The set of all $m \times n$ matrices where the entries are real numbers forms a vector space. We denote this space by $M^{(m\times n)}$. Addition and scalar multiplication are the usual addition and scalar multiplication of matrices. A basis of the space consists of the matrices $\mathrm{E}^{(i,j)}$ which have a 1 at entry (i,j) and zeros everywhere else. The space is of dimension $m \times n$.

Example 2.8. The solutions of a homogeneous linear differential equation of order two, e.g.

$$\frac{\mathrm{d}^2 y(t)}{\mathrm{d}t^2} + y(t) = 0.$$

form a vector space. The general solution is

$$y(t) = A\sin(t) + B\cos(t).$$

Addition and scalar multiplication in this vector space are the usual addition and scalar multiplication of functions. The space is spanned by two linearly independent solutions, $\sin(t)$ and $\cos(t)$. This is an example of a so-called *function space*. The same holds for a homogenous linear ODE of n-th order. In this case there are n linearly independent solutions.

Advanced material

Example 2.9. The set $\mathscr{F}$ of all infinitely differentiable functions, $\mathbb{R} \to \mathbb{R}$, i.e., those that have derivatives of all orders. Note that (among others) the following functions belong to $\mathscr{F}$:

$$e^{nx},\ \cos(2\pi nx),\ x^n \quad (\text{for all } n = 0, 1, 2, \ldots).$$

$\mathscr{F}$ is an example of an infinite-dimensional space because it contains an unlimited number of linearly independent elements.

The definition of a vector space over $\mathbb{R}$ can be extended to a vector space over a more general set of numbers, a field K. To understand what a 'field' is, we need to take a short detour into number systems.

Definition: Vector Space over K

A vector space over a field K is a set V, the elements of which are called vectors $\vec{v} \in V$ along with two operations, an addition $\vec{v} + \vec{w}$ and a scalar multiplication $a\vec{v}$, $a \in K$, subject to the same 10 axioms Equation 2.1 to Equation 2.10 as before only that now $a, b \in K$.

Example 2.10. The set 2.2 $\mathbb{C}^n = \{(c_1, ..., c_n)^T | c_1, .., c_n \in \mathbb{C}\}$ is a vector space over $\mathbb{C}$. Addition and scalar multiplication are the usual addition and scalar multiplication of complex numbers. A basis for this space is given by the n vectors $e_1 = (1, 0, ..., 0)$,..., $e_n = (0, ..., 0, 1)$. The space has n dimensions. Note that we can identify $\mathbb{C}$ with $\mathbb{R}^2$ due to $z = a + bi$ with $a, b \in \mathbb{R}$. Hence $\mathbb{C}^n \sim \mathbb{R}^{2n}$ is also a vector space over $\mathbb{R}$, but in this case it is 2n-dimensional with two basis vectors for the real and imaginary part of each complex dimension.

2.3 Linear independence

In the same way as a coordinate system is introduced on a plane to allocate a unique pair of coordinates to any point in the plane, we would like to have a coordinate system for a vector space, which uniquely assigns to any vector of this space a coordinate tuple (set). To generate such a coordinate system, we need a set of 'basis' vectors which are, in a sense, independent. Consider the following example in the x-y plane.

Given the two unit vectors in $\mathbb{R}^2$,

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Any point in the plane, that is, any vector in $\mathbb{R}^2$, can be represented by a linear combination

$$\vec{v} = v_1\vec{e}_1 + v_2\vec{e}_2, v_1, v_2 \in \mathbb{R}$$

where (v_1, v_2) play the role of coordinates. These coordinates are unique; that is, there are no two different vectors with the same coordinate pair, nor are two different coordinate pairs assigned to the same vector.

The same is true if we replace our basis vectors by (check!)

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ \vec{e}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

However, if we use a triplet of vectors:

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \ \vec{e}_3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

$$\vec{v} = v_1\vec{e}_1 + v_2\vec{e}_2 + v_3\vec{e}_3, v_1, v_2, v_3 \in \mathbb{R}$$

Then there are vectors with non-unique coordinates:

$$\vec{v} = 2\vec{e}_1 + 3\vec{e}_2 - 1\vec{e}_3 = 1\vec{e}_1 + 2\vec{e}_2 + 0\vec{e}_3$$

More generally, if such a non-unique case occurs, one can write

$$\begin{aligned}
\mathbf{v} &= v_1\vec{e}_1 + v_2\vec{e}_2 + v_3\vec{e}_3 \\
\mathbf{v} &= v_1'\vec{e}_1 + v_2'\vec{e}_2 + v_3'\vec{e}_3 \\
\Rightarrow \vec{0} &= (v_1 - v_1')\vec{e}_1 + (v_2 - v_2')\vec{e}_2 + (v_3 - v_3')\vec{e}_3, \\
\Rightarrow \vec{0} &= a_1\vec{e}_1 + a_2\vec{e}_2 + a_3\vec{e}_3, \quad (a_1, a_2, a_3) \neq (0, 0, 0)
\end{aligned}$$

where at least one of the three brackets is non-zero since we assumed that $(v_1, v_2, v_3) \neq (v_1', v_2', v_3')$. This is called a non-trivial linear combination.

Definition 2.2: Linear Combination

An expression of the form $a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n$ is known as a *linear combination of the n vectors* $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n$*. The numbers* $a_1, a_2, ..., a_n$ *are known as the coefficients of the linear combination.*

Definition 2.3: Linear Independence

A set of vectors $\vec{v}_1, \vec{v}_2, ..., \vec{v}_n$ is called *linearly independent* if none of its elements is a linear combination of the others. That is the equation

$$a_1\vec{v}_1 + a_2\vec{v}_2 + ... + a_n\vec{v}_n = 0$$

has no solution $(a_1, a_2, .., a_n)$ other than the trivial solution $(0, 0, ..., 0)$. Otherwise, the set is called *linearly dependent*.

I.e., a set of vectors is linearly dependent if we can find a non-trivial linear combination that yields the zero vector. Note that a simple relation between just two of the vectors, e.g. $\mathbf{v}_1 = 3\mathbf{v}_2$, is enough to make the complete set $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ linearly dependent.

Example 2.11. Are the vectors $(1, -1, 0)^T, (2, 1, 1)^T$ and $(-1, 2, 1)^T$ linearly independent?

We solve

$$a_1\begin{pmatrix}1\\-1\\0\end{pmatrix} + a_2\begin{pmatrix}2\\1\\1\end{pmatrix} + a_3\begin{pmatrix}-1\\2\\1\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}.$$

for (a_1, a_2, a_3). We find that $a_1 = a_2 = a_3 = 0$; so the 3 given vectors are linearly independent. If we replace the first entry in the first vector by 3, the set becomes linearly dependent. A solution is $a_1 = 1$, $a_2 = -1$, $a_3 = 1$.

Example 2.12. Are the vectors $(1, 0, 0)^T$, $(1, 1, 0)^T$ and $(1, 0, 1)^T$ linearly independent?

We solve

$$a_1\begin{pmatrix}1\\0\\0\end{pmatrix} + a_2\begin{pmatrix}1\\1\\0\end{pmatrix} + a_3\begin{pmatrix}1\\0\\1\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix},$$

for (a_1, a_2, a_3). The third component of this equation implies $a_3 = 0$. From the second and first component follow $a_2 = 0$ and $a_1 = 0$. Hence, the vectors are linearly independent.

On the other hand, the vectors

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

are linearly dependent, since the third component reads $a_3 0 = 0$ which is satisfied for any $a_3 = \lambda$. The second component requires $a_2 = -a_3$ and the first $a_1 = -a_2$. So there are non-trivial solutions $(\lambda, -\lambda, \lambda)$, e.g. $(1, -1, 1)$.

Example 2.13. A set of two vectors is linearly independent if they are not parallel: $a_1\vec{v}_1 \neq a_2\vec{v}_2$ for a_1, a_2 non-zero. A set of three vectors in $\mathbb{R}^2$ is always linearly dependent. A set of three vectors in $\mathbb{R}^3$ is linearly independent if they do not all lie in the same plane.

Example 2.14. Show that the vectors $(1, 0, 2, 1)^T, (0, 1, -1, 2)^T, (2, 1, 3, 4)^T$ are linearly dependent in $\mathbb{R}^4$.

2.4 Span and Basis

Definition 2.4: Span

The set of vectors $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ in V **span** V if every vector $\vec{v} \in V$ is a linear combination of $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$. The set of all vectors of the form $\vec{v} = a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n$ is called the span of $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$, and denoted by $\mathrm{span}(\vec{v}_1, \vec{v}_2, ..., \vec{v}_n)$.

Example 2.15.

$$\begin{aligned}\mathrm{span}\left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}2\\0\end{pmatrix}\right) &= \left\{ \mathbf{v} = a_1\begin{pmatrix}1\\0\end{pmatrix} + a_2\begin{pmatrix}2\\0\end{pmatrix} \middle| \, a_1, a_2 \in \mathbb{R} \right\} \\ &= \left\{ \mathbf{v} = \begin{pmatrix}a_1 + 2a_2\\0\end{pmatrix} \middle| \, a_1, a_2 \in \mathbb{R} \right\} = \mathbb{R}^1,\end{aligned}$$

spans the whole of $\mathbb{R}$. This space contains only the vectors with 0 in the second component.

Example 2.16. The span of two linearly independent vectors in $\mathbb{R}^3$ is a plane. The span of three linearly independent vectors in $\mathbb{R}^3$ is the whole of $\mathbb{R}^3$.

Example 2.17. The set $\{(1, 0)^T, (0, 1)^T\}$ spans $\mathbb{R}^2$.

$$\begin{aligned}\mathrm{span}\left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right) &= \left\{ \vec{v} = a_1\begin{pmatrix}1\\0\end{pmatrix} + a_2\begin{pmatrix}0\\1\end{pmatrix} \middle| \, a_1, a_2 \in \mathbb{R} \right\} \\ &= \left\{ \vec{v} = \begin{pmatrix}a_1\\a_2\end{pmatrix} \middle| \, a_1, a_2 \in \mathbb{R} \right\} = \mathbb{R}^2.\end{aligned}$$

Thus, any vector $(a_1, a_2)^T \in \mathbb{R}^2$ can be expressed as a linear combination of $(1, 0)^T$ and $(0, 1)^T$.

If we add another vector, e.g.

$$\mathrm{span}\left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}\right) = \mathbb{R}^2,$$

then the space may or may not become any bigger. In this case, the space remains the same since the third vector is a linear combination of the first two vectors, that is, the set is linearly dependent. The smallest linearly independent set which spans a given vector space is called a basis.

Definition 2.5: Basis

A basis of a vector space is a linearly independent set of vectors which span the vector space.

Example 2.18. The set of vectors $\{(1,0,0)^T, (0,1,0)^T, (0,0,1)^T\}$ forms a basis of $\mathbb{R}^3$, the set of vectors $\{(1,0,0,0)^T, (0,1,0,0)^T,$
$(0,0,1,0)^T, (0,0,0,1)^T\}$ forms a basis of $\mathbb{R}^4$ and so on. This is called the **standard basis** of $\mathbb{R}^n$ for $n \in \mathbb{N}$.

Example 2.19. The vectors

$$\vec{f}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad \vec{f}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{f}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

do not form a basis for $\mathbb{R}^3$ since they do not span the whole $\mathbb{R}^3$. The vector $\vec{e}_1$, for instance, has no representation in this set.

Example 2.20. The vectors

$$\vec{f}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{f}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{f}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \vec{f}_4 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

are also not a basis. Although they span the whole $\mathbb{R}^3$, they are not linearly independent.

Example 2.21. Consider the set of all $n \times n$ matrices with real entries $\mathrm{M}^{\mathrm{n}\times\mathrm{n}}$. Let E_{ij} be the $n \times n$ matrix with a '1' in position (i,j) and '0' elsewhere. Then the collection of all such matrices (for all $i, j = 1, 2, \ldots, n$) is a basis for $\mathbb{R}^{(n\times n)}$.

Consider the case $n = 2$, then the basis matrices are

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = aE_{11} + bE_{12} + cE_{21} + dE_{22}, \quad \text{(spanning)}.$$

Also, we can show that the four E matrices are linearly independent.

$$\alpha_1 E_{11} + \alpha_2 E_{12} + \alpha_3 E_{21} + \alpha_4 E_{22} = \mathbf{0}$$

means that

$$\begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \text{i.e. } \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0.$$

Since the four matrices span the space and are linearly independent, they form a basis for $\mathbb{R}^{(n\times n)}$.

Theorem 2.6: Uniqueness w.r.t. a basis

Let $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a basis for a vector space V. Each vector from V can be uniquely expressed as a linear combination of these vectors.

Proof

If there were two different representations of a vector $\vec{v}$ with respect to this basis: Eg. $\vec{v} = a_1\vec{v}_1 + \cdots + a_n\vec{v}_n$ and $\vec{v} = b_1\vec{v}_1 + \cdots + b_n\vec{v}_n$, then the difference $\mathbf{0} = (a_1 - b_1)\vec{v}_1 + \cdots + (a_n - b_n)\vec{v}_n$ would be a non-trivial linear combination representing the zero vector, which is impossible since the vectors $\vec{v}_1, \ldots, \vec{v}_n$ were linearly independent. □

Example 2.22. In $\mathbb{R}^3$,

$$\begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - 3\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 4\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 2\vec{e}_1 - 3\vec{e}_2 + 4\vec{e}_3.$$

So, any linear combination of the vectors $\vec{e}_1, \vec{e}_2, \vec{e}_3$ is a vector in $\mathbb{R}^3$ and any vector in $\mathbb{R}^3$ can be written as a linear combination of $\vec{e}_1, \vec{e}_2, \vec{e}_3$ and the expression is unique.

Example 2.23. A different basis, other than the standard basis, for the $\mathbb{R}^3$ is, for instance

$$\vec{f}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{f}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{f}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

We can check that we still can express any vector with respect to (w.r.t.) this basis. We express the arbitrary vector $\vec{r}$ first w.r.t. the standard basis and show that we can translate this into a representation w.r.t. the new basis:

$$\begin{aligned} \vec{r} &= r_1\vec{e}_1 + r_2\vec{e}_2 + r_3\vec{e}_3 \\ &= r_1(\vec{e}_1 + \vec{e}_2) + (r_2 - r_1)\vec{e}_2 + r_3\vec{e}_3 \\ &= r_1\vec{f}_1 + (r_2 - r_1)\vec{f}_2 + r_3\vec{f}_3. \end{aligned}$$

And this expression is unique. In addition, the definition of a basis requires that the new basis vectors are linearly independent, which they are; otherwise, they could not span the whole three-dimensional space.

We note that there are different bases, but they all have the same number of elements.

Theorem 2.7: Number of elements of a basis

Every basis of a vector space contains the same number of vectors, this number being the largest set of linearly independent vectors in the set.

Proof

Suppose that a vector space V has two bases, one of which is $\{\vec{w}_1, ..., \vec{w}_p\}$ that contains p vectors and the other $\{\vec{z}_1, ..., \vec{z}_q\}$ that contains q vectors. We can assume that $p > q$. We shall show that the vectors $\{\vec{w}_1, ..., \vec{w}_p\}$ are linearly dependent, contradicting the fact that they form a basis.
To begin, since $V = \text{span}\,(\vec{z}_1, ..., \vec{z}_q)$ and each of the $\vec{w}$'s lies in V, they must be linear combinations of the $\vec{z}$'s. i.e.,

$$\begin{aligned} \vec{w}_1 &= c_{11}\vec{z}_1 + c_{21}\vec{z}_2 + \cdots + c_{q1}\vec{z}_q \\ &\vdots \\ \vec{w}_p &= c_{1p}\vec{z}_1 + c_{2p}\vec{z}_2 + \cdots + c_{qp}\vec{z}_q. \end{aligned}$$

Now consider a linear combination of the $\vec{w}$'s:

$$\begin{aligned}\sum_{j=1}^{p} a_j \vec{w}_j =& a_1 \vec{w}_1 + \cdots + a_p \vec{w}_p \\ =& a_1 \left(c_{11}\vec{z}_1 + c_{21}\vec{z}_2 + \cdots + c_{q1}\vec{z}_q\right) \\ &+ a_2 \left(c_{12}\vec{z}_1 + c_{22}\vec{z}_2 + \cdots + c_{q2}\vec{z}_q\right) \\ &\vdots \\ &+ a_p \left(c_{1p}\vec{z}_1 + c_{2p}\vec{z}_2 + \cdots + c_{qp}\vec{z}_q\right) \\ =& \left(c_{11}a_1 + \cdots + c_{1p}a_p\right)\vec{z}_1 + \cdots + \left(c_{q1}a_1 + \cdots + c_{qp}a_p\right)\vec{z}_q.\end{aligned}$$

Can we choose the a's so that the right-hand side is zero? Since the $\vec{z}$'s are l.i., this will require that each of the coefficients be zero:

$$\begin{aligned} c_{11}a_1 + \cdots + c_{1p}a_p &= 0 \\ &\vdots \\ c_{q1}a_1 + \cdots + c_{qp}a_p &= 0 \end{aligned}$$

but this is a system of q equations in p unknowns with $p > q$. Such a system always has non-zero solutions. Hence, there will be a choice if the coefficients $a_1, \ldots, a_p$ (not all of which are zero) so that $a_1\vec{w}_1 + \cdots + a_p\vec{w}_p = 0$. But this means that the $\vec{w}$'s are l.d.. We have a contradiction, since we assumed the $\vec{w}$'s form a basis. A similar argument can be applied if we assume that $q > p$. Thus, a contradiction can be avoided only when $p = q$. □

Definition 2.8: Dimension

The number of elements of a basis is called the dimension of the space.

Example 2.24. $\mathbb{R}^n$ has dimension n.

Example 2.25. The space of polynomials of degree $\leq n$ has basis $\{1, x, x^2, \ldots, x^n\}$, so its dimension is $n + 1$ (not n).

i Remark on dimension

Note that the dimension of V depends on the field K. Thus the complex numbers $\mathbb{C}$ can be considered as a space of dimension 1 over $\mathbb{C}$, or as a space of dimension 2 over $\mathbb{R}$, where $\{1, i\}$ is a basis for $\mathbb{C}$ over $\mathbb{R}$.

Theorem 2.9: Smaller spanning set

Any linearly dependent vectors of a spanning set can be omitted without making the span smaller.

$$\text{span}(\vec{v}_1, \vec{v}_2, ..., \vec{v}_n) = \text{span(largest l.i. subset of } \vec{v}_1, \vec{v}_2, ..., \vec{v}_n)$$

Proof

Without loss of generality (w.l.o.g.) we can assume that $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_r\}$ is the largest linearly independent subset ($r \leq n$). Then any remaining vectors $\vec{v}_{r+1}, ..., \vec{v}_n$ can be expressed as a linear combination of $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_r\}$ hence $\text{span}(\vec{v}_1, \vec{v}_2, ..., \vec{v}_r) = \text{span}(\vec{v}_1, \vec{v}_2, ..., \vec{v}_n)$ □

Corollary 2.10

n linearly independent vectors in $\mathbb{R}^n$ span $\mathbb{R}^n$. Vice versa a set of $n+1$ vectors in $\mathbb{R}^n$ must be linearly dependent.

Proof

We know that the standard basis of $\mathbb{R}^n$ consists of n elements, hence the dimension is n. Suppose the n vectors would not span $\mathbb{R}^n$. Then we could add vectors such that the new set spans the space and forms a basis. However, we would then have a basis with more than n vectors. That can't be, as any basis of $\mathbb{R}^n$ has to have n vectors. So the assumption that the vectors don't span of $\mathbb{R}^n$ is false and they do span the space. Similar, $n+1$ vectors in $\mathbb{R}^n$ must be linearly dependent. If they were linearly independent they could form a basis with more than n elements. Again this can't be true, hence they are linearly dependent. □

2.5 Subspaces

Suppose that W is a nonempty subset of the vector space V, which also satisfies all the vector space axioms itself, then W is called a subspace of V. It is important to realise that for W to be subspace of V, then W must satisfy the following conditions:

Definition 2.11: Subspace

A subset $W \subseteq V$ of vectors in a vector space V is called a subspace of V if the following conditions hold:
(i) $\vec{0} \in W$,
(ii) if $\vec{u} \in W$ and $\vec{v} \in W$ then $\vec{u} + \vec{v} \in W$;
(iii) if $\vec{u} \in W$ then $a\vec{u} \in W$ for all $a \in K$.

All other axioms are automatically satisfied because each element in W is already in V.

Example 2.26. For any vector space V, V is always a subspace of itself.

Example 2.27. We also always have a subspace $\{\vec{0}\}$ consisting of the zero vector alone. This is called the trivial subspace, and its dimension is 0, because it has no linearly independent sets of vectors at all.

Example 2.28. The subspaces of $\mathbb{R}^2$ are $\{\vec{0}\}$, lines through the origin, and $\mathbb{R}^2$ itself.

Example 2.29. The subspaces of $\mathbb{R}^3$ are $\{\vec{0}\}$, lines through the origin, planes through the origin, and $\mathbb{R}^3$ itself.

Example 2.30. Let $W = \{(x, y, z)^T : y = 2x, z = 0, x, y, z \in \mathbb{R}\}$. This is the set of points lying on the line $y = 2x$ in the plane $z = 0$. W is a subspace of $\mathbb{R}^3$.

Example 2.31. The set $W = \{(x, y, z, w)^T : x + y + z = 1\}$ is not a subspace of $\mathbb{R}^4$. The simplest way to see this is that $\vec{0} \notin W$. Another way is to choose suitable vectors $\vec{u}, \vec{v} \in W$ then show that $\vec{u} + \vec{v} \notin W$. A third way is to choose a suitable $\vec{u} \in W$ then show that $a\vec{u} \notin W$ for any $a \neq 1$. Any one of the reasons suffices.

Example 2.32. Consider the subset S of P_2 defined so as to contain all polynomial $\mathrm{p}(x) = a + bx + cx^2$ for which $a + b - 2c = 0$. Show that S is a subspace of P_2 and obtain a basis for S.

Theorem 2.12: Intersection of subspaces

If W_1 and W_2 are subspaces of V then so is $W_1 \cap W_2$.

Proof

Let $\vec{u}, \vec{v} \in W_1 \cap W_2$ and $a \in K$. Then $\vec{u} + \vec{v} \in W_1$ (because W_1 is a subspace) and $\vec{u} + \vec{v} \in W_2$ (because W_2 is a subspace). Hence $\vec{u} + \vec{v} \in W_1 \cap W_2$. Similarly, we get $a\vec{v} \in W_1 \cap W_2$ so $W_1 \cap W_2$ is a subspace of V. $\square$

Example 2.33. Let $V = \mathbb{R}^{2\times 2}$ and define W_1 to be the subspace of matrices of the form

$$\begin{bmatrix} a & 0 \\ b & c \end{bmatrix}$$

(a, b, c real) while W_2 is the subspace of matrices of the form $\begin{bmatrix} a & a \\ b & b \end{bmatrix}$. Then $W_1 \cap W_2$ consists of all matrices of the form $\begin{bmatrix} 0 & 0 \\ b & b \end{bmatrix}$, which is a subspace of V.

Remark

While $W_1 \cap W_2$ is a subpace, $W_1 \cup W_2$ is in general not a subspace. For exampe if W_1 and W_2 are the x- and y-axis of $\mathbb{R}^2$, the union is not a subspace.

Example 2.34. Let $V = \mathbb{R}^2$, let $W = \{(a, 0)^T : a \in \mathbb{R}\}$ and $W_2 = \{(0, b)^T : b \in \mathbb{R}\}$. Then W_1, W_2 are subspaces of V, but $W_1 \cup W_2$ is not a subspace, because $(1, 0)^T \in W_1 \cup W_2$ and $(0, 1)^T \in W_1 \cup W_2$, but $(1, 0)^T + (0, 1)^T = (1, 1)^T \notin W_1 \cup W_2$.

Example 2.35. The set $W = \{(x, y, z)^T : x^2 + y^2 = z\}$ is not a subspace of $\mathbb{R}^3$.

The vectors $\vec{u} = (0, 1, 1)^T$ and $\vec{v} = (1, 2, 5)^T$ are both in W but $\vec{u} + \vec{v}$ is not.

The three conditions in the definition of a subspace state that every linear combination of vectors of the subspace has to be an element of the subspace. This proves the following theorem:

Lemma 2.13: Span is a subspace

$\text{span}(\vec{v}_1, \vec{v}_2, ..., \vec{v}_n)$ is a subspace of V for $\vec{v}_1, \vec{v}_2, ..., \vec{v}_n \in V$.

Proof

$W = \text{span}(\vec{v}_1, \vec{v}_2, ..., \vec{v}_n)$ is by definition a subset of V. The $\vec{0}$ is in W, which is property (i) in the definition of a subspace, and (ii) and (iii) are also satisfied due to the span including all linear combinations. □

Example 2.36. The set $W = \{(x, y, z, w)^T : x + y + z = w\}$ is a subspace of $\mathbb{R}^4$. We can show this in two ways: (a) by using the definition of a subspace (2.11) or (b) by finding a spanning set and using the above theorem. Method (a) requires checking the three conditions for a subspace:

1. $\mathbf{0} = (0, 0, 0, 0)^T \in W$ for $x = y = z = 0$.
2. With $\mathbf{u} = (x, y, z, x + y + z)^T \in W$ and $\mathbf{v} = (x', y', z', x' + y' + z')^T \in W$ also $\mathbf{u} + \mathbf{v} = (x + x', y + y', z + z', (x + x') + (y + y') + (z + z'))^T \in W$
3. With $\mathbf{u} = (x, y, z, x + y + z)^T \in W$ and $\lambda\mathbf{u} = (\lambda x, \lambda y, \lambda z, \lambda(x + y + z))^T \in W$

For method (b), we have to find vectors which span W. Note that the set W has three free parameters, which can be used to represent an arbitrary element of W:

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

Hence $W = \text{span}((1, 0, 0, 1)^T, (0, 1, 0, 1)^T, (0, 0, 1, 1)^T)$. This second approach has the advantage that the spanning set can also be used to find a basis for the subspace. Indeed, the three vectors are also linearly independent (prove!), and we obtain the basis $\vec{v}_1 = (1, 0, 0, 1)^T, \vec{v}_2 = (0, 1, 0, 1)^T, \vec{v}_3 = (0, 0, 1, 1)^T$.

2.6 Scalar- and Inner Product

Definition 2.14: scalar product, Euclidean product, inner product, dot product

For any two vectors $\vec{v}, \vec{w} \in \mathbb{R}^n$ with components

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \quad \vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$$

we define the inner or scalar product as

$$\vec{v} \cdot \vec{w} = v_1 w_1 + v_2 w_2 + .. + v_n w_n \in \mathbb{R}.$$

Usually we use as a symbol for the scalar product the dot. Sometimes the notation $\langle \vec{v}, \vec{w} \rangle$ is used as well.

Corollary 2.15

If $\vec{u}$, $\vec{v}$, $\vec{w}$ are vectors in $\mathbb{R}^n$, and α is scalar, then

a)
$$\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$$

b)
$$\vec{u} \cdot (\vec{v} + \vec{w}) = \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{w}$$

c)
$$\alpha(\vec{u} \cdot \vec{v}) = \alpha\vec{u} \cdot \vec{v} = \vec{u} \cdot (\alpha\vec{v})$$

Proof

(a)

$$\begin{aligned}\vec{u} \cdot \vec{v} &= u_1 v_1 + u_2 v_2 + \cdots + u_n v_n \\ &= v_1 u_1 + v_2 u_2 + \cdots + v_n n_n \\ &= \vec{v} \cdot \vec{u}\end{aligned}$$

(b)

$$\begin{aligned}\vec{u} \cdot (\vec{v} + \vec{w}) &= u_1(v_1 + w_1) + \cdots + u_n(v_n + w_n) \\ &= (u_1 v_1 + \cdots + u_n v_n) + (u_1 w_1 + \cdots + u_n w_n) \\ &= \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{w}\end{aligned}$$

(c)

$$\begin{aligned}\alpha(\vec{u} \cdot \vec{v}) &= \alpha(u_1 v_1 + u_2 v_2 + \cdots + u_n v_n) \\ &= \alpha u_1 v_1 + \alpha u_2 v_2 + \cdots + \alpha u_n v_n \\ &= (\alpha\vec{u}) \cdot \vec{v}\end{aligned}$$

We also have

$$\begin{aligned}\alpha u_1 v_1 + \alpha u_2 v_2 + \cdots + \alpha u_n v_n &= u_1 \alpha v_1 + u_2 \alpha v_2 + \cdots + u_n \alpha v_n \\ &= \vec{u} \cdot (\alpha\vec{v})\end{aligned}$$

We note that if we use for both entries of the scalar product the same vector then we get

Definition 2.16

The norm (or length) of a vector $\vec{v} \in \mathbb{R}^n$ is defined as

$$\|\vec{v}\| = \sqrt{\vec{v} \cdot \vec{v}} = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2} \geq 0$$

Note that this definition matches with the length of vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ as calculated by the Pythagorean theorem.

Corollary 2.17

The scalar product is related to the angle between the vectors by

$$\vec{v} \cdot \vec{w} = \|\vec{v}\| \; \|\vec{w}\| \cos\alpha.$$

Proof

Let $\vec{v}$, $\vec{w}$ be non-zero vectors. Then three vectors $\vec{v}$, $\vec{w}$ and $\vec{v} - \vec{w}$ give three sides of a triangle, see figure **??**. Therefore by the Pythagorean theorem we have

$$\begin{aligned}\|\vec{v} - \vec{w}\|^2 = a^2 + b^2 &= ((\|\vec{v}\| - \|\vec{w}\| \cos\alpha)^2 + (\|\vec{w}\| \sin\alpha)^2) \\ &= \|\vec{v}\|^2 \; - 2\|\vec{v}\|\|\vec{w}\| \cos\alpha + \|\vec{w}\|^2 \cos^2\alpha + \|\vec{w}\|^2 \sin^2\alpha \\ &= \|\vec{v}\|^2 + \|\vec{w}\|^2 - 2\|\vec{v}\|\|\vec{w}\| \cos\alpha\end{aligned}$$

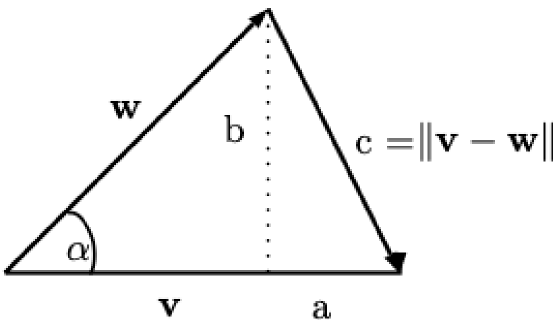


Figure 2.1: Pythagoras

On the other hand the definition of a length gives us

$$\begin{aligned}\|\vec{v} - \vec{w}\|^2 &= (\vec{v} - \vec{w}) \cdot (\vec{v} - \vec{w}) \\ &= \vec{v} \cdot \vec{v} - 2\vec{w} \cdot \vec{v} + \vec{w} \cdot \vec{w}\end{aligned}$$

Comparing the two results we find $\vec{v} \cdot \vec{w} = \|\vec{v}\|\|\vec{w}\| \cos\alpha$.

An immediate consequence is that for two non-vanishing vectors ($\|\vec{v}\|, \|\vec{w}\| \neq 0$) we have

$$\vec{v} \cdot \vec{w} = 0 \Leftrightarrow \vec{v} \perp \vec{w},$$

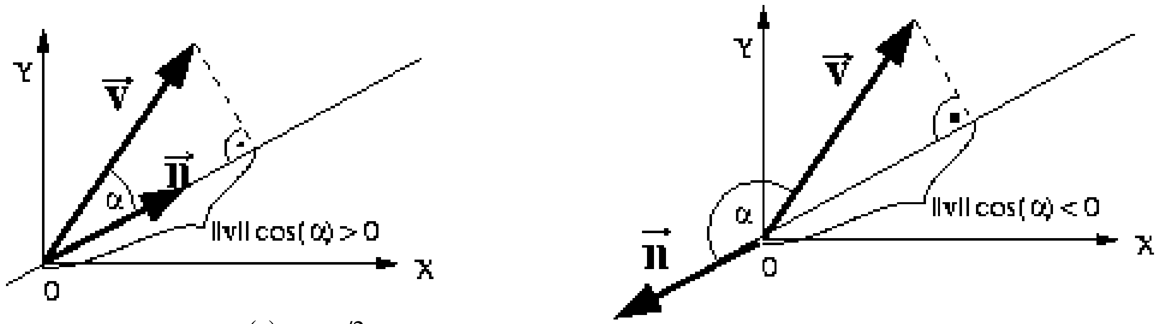
i.e.~the scalar product vanishes if and only if they are perpendicular to each other.

Corollary 2.18

If $\vec{n}$ is a unit vector, i.e.~$\|\vec{n}\| = 1$, then $\vec{n} \cdot \vec{v}$ is the length of the projection of $\vec{v}$ onto the direction of $\vec{n}$. If the angle between $\vec{n}$ and $\vec{v}$ is greater than $\pi/2$ it is minus the length of the projection.

Proof

$$\vec{n} \cdot \vec{v} = \underbrace{\|\vec{n}\|}_{=1} \; \|\vec{v}\| \cos\alpha = \; \|\vec{v}\| \cos\alpha$$



(a) $\alpha < \pi/2$

(b) $\alpha > \pi/2$

Figure 2.2: The projection $\vec{n} \cdot \vec{v}$ of a vector $\vec{v}$ onto a unit vector $\vec{n}$ for an angle $\alpha < \pi/2$ and $\alpha > \pi/2$

The distance between two points $\vec{u}$ and $\vec{v}$ is in general given by the norm of $\vec{u} - \vec{v}$:

$$||\vec{u} - \vec{v}|| = ||\vec{v} - \vec{u}||.$$

This is a meaningful definition since it is the shortest distance between the points with the position vectors $\vec{v}$ and $\vec{u}$ is the length of the vector between the two points. This is stated by the triangle inequality:

Corollary 2.19

For any two vectors $\vec{x}$ and $\vec{y}$ (of a normed vector space) we have the triangle inequality:

$$||\vec{x} + \vec{y}|| \leq ||\vec{x}|| + ||\vec{y}||.$$

Proof

$$\begin{aligned}
||\vec{x} + \vec{y}||^2 &= (\vec{x} + \vec{y}) \cdot (\vec{x} + \vec{y}) \\
&= \vec{x} \cdot \vec{x} + \vec{y} \cdot \vec{y} + 2\vec{x} \cdot \vec{y} \\
&= ||\vec{x}||^2 + ||\vec{y}||^2 + 2||\vec{x}||\,||\vec{y}|| \cos(\alpha) \\
&\leq ||\vec{x}||^2 + ||\vec{y}||^2 + 2||\vec{x}||\,||\vec{y}|| \\
&= (||\vec{x}|| + ||\vec{y}||)^2 \\
\Rightarrow ||\vec{x} + \vec{y}|| &\leq ||\vec{x}|| + ||\vec{y}||
\end{aligned}$$

One can use the key properties of the Scalar Product (2.15) as a basis to define a generalisation of a scalar product, called innner product.

Definition 2.20: Inner Product

The Inner product between two vectors in the vector space V over the field $\mathbb{C}$ (or $\mathbb{R}$) is a map

$$\begin{aligned}
\langle \cdot, \cdot \rangle : \quad & V \times V \to \mathbb{C} \\
& (\vec{u}, \vec{v}) \to \langle \vec{u}, \vec{v} \rangle
\end{aligned}$$

that satisfies the following requirements $\forall \vec{u}, \vec{v}, \vec{w} \in V$ and $a, b \in \mathbb{C}$:

(1) $\langle \vec{u}, \vec{v} \rangle = \overline{\langle \vec{v}, \vec{u} \rangle}$ (symmetric up to complex conjugation)

(2) $\langle a\vec{u} + b\vec{v}, \vec{w} \rangle = a\langle \vec{u}, \vec{w} \rangle + b\langle \vec{v}, \vec{w} \rangle$ (linearity)

(3) $\langle \vec{v}, \vec{v} \rangle \geq 0$. $\langle \vec{v}, \vec{v} \rangle = 0$ if and only if $\vec{v} = 0$. (positive definiteness)

Remark

The third property implies that the inner prodoct of a vector with itself is real and non-negative. This allows us to define the norm of a vector in the same way as before

$$\|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v} \rangle}$$

and the norm becomes a real non-negative number.

Example 2.37. The space of polynomials of degree n, $P_n(t)$, defined on the interval $[-1, 1] \subset \mathbb{R}$ can be given an inner product

$$\begin{aligned} \langle \cdot, \cdot \rangle : \quad & P_n(t) \times P_n(t) \to \mathbb{R} \\ & (p(t), q(t)) \longrightarrow \langle p, q \rangle = \int_{-1}^{1} p(t)q(t)\, dt. \end{aligned}$$

3 Linear equations

3.1 Matrices

Definition 3.1: Matrix

An $m \times n$ (read "m by n") matrix is an array of numbers or expressions with m rows and n columns. The set of all $m \times n$ matrices is denoted by $M_{m \times n}$.

Example 3.1. A 2×3 matrix:

$$\mathrm{A} = \begin{pmatrix} 1 & 2 & 5 \\ 3 & 4 & -7 \end{pmatrix} \in M_{2\times 3}, \quad A_{12} = 2.$$

Notation

Matrices are usually named by upper case roman letters, e.g. A. Individual entries are referred to by $A_{i,j}$, i.e. the entry in row i and column j of the array. Entries are also sometimes denoted by the corresponding lower-case letter, e.g. $a_{i,j}$. A matrix can be symbolically represented by its general entry: $\mathrm{A} = \left[a_{ij}\right]$.

Analogous to the rules of addition and scalar multiplication of vectors, we have the following rules for matrices:

Definition 3.2: Matrix sum and scalar multiplication

Addition of two matrices of the same type ($\mathrm{A}, \mathrm{B} \in M_{m\times n}$) is defined by

$$\mathrm{A} + \mathrm{B} = [a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}],$$

and the scalar multiplication by

$$c\mathrm{A} = c[a_{ij}] = [ca_{ij}].$$

Example:

$$\begin{pmatrix} 1 & 2 & 5 \\ 3 & 4 & -7 \end{pmatrix} + \begin{pmatrix} 2 & 1 & 0 \\ -1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 3 & 5 \\ 2 & 6 & -4 \end{pmatrix}$$

$$3\begin{pmatrix} 1 & 2 & 5 \\ 3 & 4 & -7 \end{pmatrix} = \begin{pmatrix} 3 & 6 & 15 \\ 9 & 12 & -21 \end{pmatrix}$$

The neutral element of the addition is the *zero matrix*, O, where all entries are zero. The inverse element of the addition is the negative matrix $-\mathrm{A} = [-a_{ij}]$. The two operations lead to the following properties:

Corollary 3.3: Basic Matrix Operations

$$\begin{aligned}
\mathrm{A}+\mathrm{B} &= \mathrm{B}+\mathrm{A} \quad \text{(addition is ccommutative)}\\
(\mathrm{A}+\mathrm{B})+\mathrm{C} &= \mathrm{A}+(\mathrm{B}+\mathrm{C}) \quad \text{(addition is associative)}\\
\mathrm{A}+\mathrm{O} &= \mathrm{A} \quad \text{(Existence of a neutral element for add.)}\\
\mathrm{A}+(-\mathrm{A}) &= \mathrm{O} \quad \text{(Existence of an inverse element for add.)}\\
a(\mathrm{A}+\mathrm{B}) &= a\mathrm{A}+a\mathrm{B} \quad \text{(Distributive law 1)}\\
(a+b)\mathrm{A} &= a\mathrm{A}+b\mathrm{A} \quad \text{(Distributive law 2)}\\
a(b\mathrm{A}) &= (ab)\mathrm{A} \quad \text{(scalar mult. is associative)}
\end{aligned}$$

In addition to the rules above, we have a matrix multiplication:

Definition 3.4: Matrix product

The matrix product of the $m{\times}r$ matrix G and the $r{\times}n$ matrix H is the $m{\times}n$ matrix P, where

$$p_{i,j} = g_{i,1}h_{1,j} + g_{i,2}h_{2,j} + \cdots + g_{i,r}h_{r,j} = \sum_{s=1}^{r} g_{is}h_{sj}$$

that is, the i,j-th entry of the product is the dot product of the i-th row and the j-th column.

$$GH = \begin{pmatrix} & \vdots & & \\ g_{i,1} & g_{i,2} & \cdots & g_{i,r} \\ & \vdots & & \end{pmatrix}\begin{pmatrix} & h_{1,j} & \\ \cdots & h_{2,j} & \cdots \\ & \vdots & \\ & h_{r,j} & \end{pmatrix} = \begin{pmatrix} & \vdots & \\ \cdots & p_{i,j} & \cdots \\ & \vdots & \end{pmatrix}$$

Example 3.2.

$$\begin{pmatrix} 2 & 0 \\ 4 & 6 \\ 8 & 2 \end{pmatrix}\begin{pmatrix} 1 & 3 \\ 5 & 7 \end{pmatrix} = \begin{pmatrix} 2\cdot 1+0\cdot 5 & 2\cdot 3+0\cdot 7 \\ 4\cdot 1+6\cdot 5 & 4\cdot 3+6\cdot 7 \\ 8\cdot 1+2\cdot 5 & 8\cdot 3+2\cdot 7 \end{pmatrix} = \begin{pmatrix} 2 & 6 \\ 34 & 54 \\ 18 & 38 \end{pmatrix}$$

Corollary 3.5: Matrix product rules

The matrix multiplication is distributive with respect to the addition and scalar multiplication, and associative but not commutative.

$$\begin{aligned}
(\mathrm{A}+\mathrm{B})\mathrm{C} &= \mathrm{AC}+\mathrm{BC},\\
\mathrm{C}(\mathrm{A}+\mathrm{B}) &= \mathrm{CA}+\mathrm{CB},\\
(a\mathrm{A})(b\mathrm{B}) &= (ab)\mathrm{AB},\\
(\mathrm{AB})\mathrm{C} &= \mathrm{A}(\mathrm{BC})
\end{aligned}$$

Warning

Matrix multiplication is not commutative in general: $\mathrm{AB} \neq \mathrm{BA}$. This is clear already for dimensional reasons. If $\mathrm{A} \in M_{m\times r}$ and $\mathrm{B} \in M_{r\times n}$ then $\mathrm{AB} \in M_{m\times n}$, but the product BA is not even defined, unless $m = n$. However, even if we have quadratic matrices ($m = r = n$), the product is not necessarily commutative as the following example shows:

$$\begin{pmatrix}0&0\\1&0\end{pmatrix}\begin{pmatrix}2&0\\0&0\end{pmatrix}=\begin{pmatrix}0&0\\2&0\end{pmatrix} \quad \text{but} \quad \begin{pmatrix}2&0\\0&0\end{pmatrix}\begin{pmatrix}0&0\\1&0\end{pmatrix}=\begin{pmatrix}0&0\\0&0\end{pmatrix}.$$

Example 3.3. Let

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

AB, CA, CB are not defined, but the other product are

$$BA=\begin{pmatrix}3&1&1\\6&4&3\end{pmatrix},\quad BC=\begin{pmatrix}2\\6\end{pmatrix},\quad AC=\begin{pmatrix}3\\5\\3\end{pmatrix},\quad A^2=\begin{pmatrix}3&4&-3\\1&4&4\\4&1&3\end{pmatrix}$$

.

The matrix C in the above example shows that vectors $\in \mathbb{R}^n$ can be considered as $n\times 1$ matrices. Similarly, a $1\times n$ matrix is often called a "row-vector". We can use the operation of transposition to convert a vector to a "row vector" and vice versa. This operation is defined for arbitrary matrices.

Definition 3.6: Transpose of a Matrix

If $A=[a_{ij}]$ is an $m\times n$ matrix, then the transpose of A, denoted by A^T, is defined as

$$A^T=[a_{ji}]\in M_{n\times m}.$$

This means that, for example, the first row of A^T is the first column of A and so on. Note that unless we have a square matrix ($m=n$), the transpose of a matrix is of a different type than the original matrix. In particular, the transpose of a column matrix (i.e. a $n\times 1$ matrix) is a row matrix, and vice versa.

Example 3.4.

$$\begin{pmatrix}2&0&1\\1&0&3\end{pmatrix}^T=\begin{pmatrix}2&1\\0&0\\1&3\end{pmatrix} \qquad (2,0,1)^T=\begin{pmatrix}2\\0\\1\end{pmatrix}$$

Corollary 3.7: Properties of the Transpose

$$\begin{aligned}&(i)\quad (\mathrm{A}^T)^T=\mathrm{A}\\&(ii)\quad (\mathrm{A}+\mathrm{B})^T=\mathrm{A}^T+\mathrm{B}^T\\&(iii)\quad (\mathrm{AB})^T=\mathrm{B}^T\mathrm{A}^T\end{aligned}$$

Proof

(i) follows directly from the definition.

(ii)

$$\left((\mathrm{A}+\mathrm{B})^T\right)_{ij}=(\mathrm{A}+\mathrm{B})_{ji}=a_{ji}+b_{ji}=(\mathrm{A}^T)_{ij}+(\mathrm{B}^T)_{ij}$$

(iii) This requires, of course, that the product is defined, that is $\mathrm{A} \in M_{m\times r}$ and $\mathrm{B} \in M_{r\times n}$.

$$
\begin{aligned}
(\mathrm{AB})_{ij} &= \sum_{s=1}^{r} a_{is} b_{sj} \\
\left((\mathrm{AB})^T\right)_{ij} &= \sum_{s=1}^{r} a_{js} b_{si} \\
&= \sum_{s=1}^{r} (\mathrm{A}^T)_{sj} (\mathrm{B}^T)_{is} \\
&= \sum_{s=1}^{r} (\mathrm{B}^T)_{is} (\mathrm{A}^T)_{sj} = \mathrm{B}^T \mathrm{A}^T
\end{aligned}
$$

□

3.2 Symmetric and Skew-symmetric Matrices

Definition 3.8: Symmetric and skew-symmetric matrices

Let A be a square matrix. If $\mathrm{A} = \mathrm{A}^T$ then A is said to be symmetric. If $\mathrm{A} = -\mathrm{A}^T$ then A is said to be skew-symmetric or anti-symmetric.

Some matrices are neither symmetric nor skew-symmetric. However, every square matrix can be written uniquely as the sum of a symmetric and skew-symmetric matrix because of the identity

$$
\mathrm{A} = \frac{\mathrm{A}^T + \mathrm{A}}{2} + \frac{\mathrm{A} - \mathrm{A}^T}{2}, \tag{3.1}
$$

noting that $(A^T + A)/2$ is symmetric, and $(\mathrm{A} - \mathrm{A}^T)/2$ is skew-symmetric.

Example 3.5. Example for a symmetric and a skew-symmetric matrix.

$$
\mathrm{A} = \begin{pmatrix} 1 & 0 & -4 \\ 0 & 2 & 5 \\ -4 & 5 & 3 \end{pmatrix} = \mathrm{A}^{\mathrm{T}}; \quad \mathrm{A} = \begin{pmatrix} 0 & 1 & -4 \\ -1 & 0 & 5 \\ 4 & -5 & 0 \end{pmatrix} = (-\mathrm{A}^T).
$$

Note that a skew symmetric matrix has to have zeros on the diagonal.

Example 3.6. Example for the decomposition Equation 3.1.

$$
\begin{pmatrix} 2 & 0 & -4 \\ -2 & 2 & 0 \\ 2 & 4 & 6 \end{pmatrix} = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & 2 \\ -1 & 2 & 6 \end{pmatrix} + \begin{pmatrix} 0 & 1 & -3 \\ -1 & 0 & -2 \\ 3 & 2 & 0 \end{pmatrix}.
$$

3.3 Inverse Matrix

There exists a neutral element, $\mathrm{I} \in M_{n\times n}$, w.r.t.~the matrix multiplication, called the identity matrix for the multiplication, which consists of 1's down the main diagonal and 0's everywhere else:

$$\mathrm{I} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix}$$

The existence of a neutral element with respect to matrix multiplication means:

$$\mathrm{AI} = \mathrm{A} \quad \text{and} \quad \mathrm{IA} = \mathrm{A} \quad \text{for } \mathrm{A} \in M_{m\times r}.$$

Note that if $\mathrm{A} \in M_{m\times r}$ then in the first equation $\mathrm{I} \in M_{r\times r}$ while in the second equation $\mathrm{I} \in M_{m\times m}$.

Example 3.7.

$$\mathrm{IA} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} = \mathrm{A}.$$

$$\mathrm{AI} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} = \mathrm{A}.$$

Definition 3.9: Inverse of a matrix

If two square matrices A and B satisfy

$$\mathrm{AB} = \mathrm{BA} = \mathrm{I}$$

then $\mathrm{B} = \mathrm{A}^{-1}$ is called the inverse of A (and vice versa A is the inverse of B). If A has an inverse, it is said to be non-singular. Any matrix which does not have an inverse is said to be singular.

Remark

- Note that we have have to require both $\mathrm{AB} = \mathrm{I}$ and $\mathrm{BA} = \mathrm{I}$ since the matrix product is not commutative.
- Non-square matrices can never have an inverse (see remark 1).

Warning

A division of matrices is **not** defined.

That we don't have a division of matrices is due to the fact that many matrices don't have an inverse contrary to the multiplication of real numbers where only 0 doesn't have an inverse. And even if they do have an inverse it matters whether we multiply from the left or the right, as the multiplication is not commutative. For instance the matrix equation ($\mathrm{A}, \mathrm{B}, \mathrm{C}$ are $n \times n$ matrices)

$$\mathrm{AB} = \mathrm{C}$$

can be solved for B using the inverse of A, provided it exists. We multiply from the left by A^{-1}:

$$\mathrm{A}^{-1}\mathrm{AB} = \mathrm{A}^{-1}\mathrm{C}$$

$$\Rightarrow \mathrm{IB} = \mathrm{B} = \mathrm{A}^{-1}\mathrm{C}$$

However, we can't write C/A as that would not indicate whether we have to multiply with the inverse from the left or right.

Example 3.8. Check that $\mathrm{AA}^{-1} = \mathrm{I}$ and $\mathrm{A}^{-1}\mathrm{A} = \mathrm{I}$.

$$A = \begin{bmatrix} 2 & 0 & 1 \\ 1 & -2 & 1 \\ 3 & 1 & 1 \end{bmatrix}, \quad A^{-1} = \begin{bmatrix} -3 & 1 & 2 \\ 2 & -1 & -1 \\ 7 & -2 & -4 \end{bmatrix}$$

Corollary 3.10: Inverse of Transpose and Product

If A and B are non-singular $n \times n$ matrices then

$$(\mathrm{A}^{-1})^T = (\mathrm{A}^T)^{-1}$$

and

$$(\mathrm{AB})^{-1} = \mathrm{B}^{-1}\mathrm{A}^{-1}$$

Proof

$$\mathrm{I} = \mathrm{I}^T = (\mathrm{A}^{-1}\mathrm{A})^T = \mathrm{A}^T \left(\mathrm{A}^{-1}\right)^T$$

and

$$\mathrm{I} = \mathrm{I}^T = (\mathrm{AA}^{-1})^T = \left(\mathrm{A}^{-1}\right)^T \mathrm{A}^T$$

imply that $\left(\mathrm{A}^{-1}\right)^T = \left(\mathrm{A}^T\right)^{-1}$.
The second identity is proved by

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_nA^{-1} = AA^{-1} = I_n.$$

and the corresponding reverse sequence:

$$(B^{-1}A^{-1}(AB)) = B^{-1}A^{-1}AB = I_n.$$

□

3.4 Gaussian Elimination

In the previous sections, we encountered systems of equations of the type

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = d.$$

In the following, we will use a method called Gaussian elimination to determine solutions.

We start with a simple example. Consider the system

$$\begin{aligned} 2x + 3y &= 1, \\ x + y &= 2. \end{aligned}$$

Usually we would solve the system by solving e.g. the second equation for x and substituting the result in the first equation:

$$\begin{aligned} 2(2-y) + 3y &= 1 \\ \Leftrightarrow (3 - 2\cdot 1)y &= 1 - 2\cdot 2 \\ \Rightarrow y &= -3 \end{aligned}$$

The result $y = -3$ is then substituted back in either of the two equations, and we find $x = 5$. When it comes to systems with large numbers of variables and equations, this approach becomes increasingly hard to follow, and we need a more systematic approach. Note that the second line above can be understood as the result of subtracting twice the second equation from the first equation in the original system. That is, we converted the system

$$\begin{aligned} 2x + 3y &= 1, \\ 2x + 2y &= 4\ . \end{aligned}$$

into the equivalent system

$$\begin{aligned} 2x + 3y &= 1, \\ -y &= 3\ , \end{aligned}$$

by subtracting equations. "Equivalent" here means that both systems have the same solutions. This system is now trivial to solve. Operations which lead to equivalent systems are, e.g., multiplying an equation by a (non-zero) number and, as we have seen above, adding multiples of other equations.

Example 3.9. Let us try this method with a more complicated system. We indicate on the right-hand side of the system the operation (multiplication and subtraction) which leads to the new system.

$$\begin{aligned} x + y - z + 2t &= 12, \\ 2x - y + z - t &= -5, \qquad Eq.2 - 2Eq.1 \\ x - 2y + 3z + 4t &= 10, \qquad Eq.3 - Eq.1 \\ 3x + 3y + z + t &= 12. \qquad Eq.4 - 3Eq.1 \end{aligned}$$

$$\Leftrightarrow \qquad \begin{aligned} x + y - z + 2t &= 12, \\ 0 - 3y + 3z - 5t &= -29, \\ 0 - 3y + 4z + 2t &= -2, \qquad Eq.3 - Eq.2 \\ 0 + 0 + 4z - 5t &= -24. \end{aligned}$$

$$\Leftrightarrow \qquad \begin{aligned} x + y - z + 2t &= 12, \\ 0 - 3y + 3z - 5t &= -29, \\ 0 + 0 + z + 7t &= 27, \\ 0 + 0 + 4z - 5t &= -24. \qquad Eq.4 - 4Eq.3 \end{aligned}$$

$$
\begin{aligned}
& x + y - z + 2t = 12, \\
\Leftrightarrow \quad & 0 - 3y + 3z - 5t = -29, \\
& 0 + 0 + z + 7t = 27, \\
& 0 + 0 + 0 - 33t = -132.
\end{aligned}
$$

Now the system can be solved recursively: The last equation implies $t = 4$, which can be used in the third equation to find $z = -1$. This, in turn, leads to $y = 2$ in the second equation, and eventually, we get $x = 1$ from the first equation.

Note that it was now possible to easily solve the system because the non-zero entries on the left-hand side form an upper triangle, so that we can successively solve for all the variables (back-substitution).

To make the notation more compact, we can suppress the variables and write the system as a matrix, called *augmented matrix*. Instead of refering to equations we refer now to rows, and $2R_2 - R_1$ for instance means take twice row 2 and subtract row 1. For the example from above, we have e.g.

$$
\begin{array}{rrrr|rl}
1 & 1 & -1 & 2 & 12 & \\
2 & -1 & 1 & -1 & -5 & (R_2 - 2R_1) \\
1 & -2 & 3 & 4 & 10 & (R_3 - R_1) \\
3 & 3 & 1 & 1 & 12 & (R_4 - 3R_1)
\end{array}
\Rightarrow
\begin{array}{rrrr|rl}
1 & 1 & -1 & 2 & 12 & \\
0 & -3 & 3 & -5 & -29 & \\
0 & -3 & 4 & 2 & -2 & (R_3 - R_2) \\
0 & 0 & 4 & -5 & -24 &
\end{array}
$$

$$
\Rightarrow
\begin{array}{rrrr|rl}
1 & 1 & -1 & 2 & 12 & \\
0 & -3 & 3 & -5 & -29 & \\
0 & 0 & 1 & 7 & 27 & \\
0 & 0 & 4 & -5 & -24 & (R_4 - 4R_3)
\end{array}
\Rightarrow
\begin{array}{rrrr|r}
1 & 1 & -1 & 2 & 12 \\
0 & -3 & 3 & -5 & -29 \\
0 & 0 & 1 & 7 & 27 \\
0 & 0 & 0 & -33 & -132
\end{array}
$$

This method is called Gauss elimination, named after Carl Friedrich Gauss (see Appendix [Gauss]). To bring the system into upper triangular form, we can use three rules which do not change the solutions of the system of equations:

1. multiply an equation by an arbitrary non-zero number,
2. adding (or subtracting) multiples of equations to other equations (not itself!),
3. change the sequence of equations.

There are two cases where the system can fail to give a unique solution. The first case is that the system is *underdetermined* (or has infinitely many solutions), that is, we either have fewer equations than variables or the equations are not linearly independent. In this case, we are left with one (or more) free variables, and the best we can do is express all the other variables in terms of this free variable.

Example 3.10.

$$
\begin{array}{rrrr|r}
1 & 1 & -1 & 2 & 12 \\
2 & -1 & 1 & -1 & -5 \\
1 & -2 & 3 & 4 & 10 \\
2 & 2 & -3 & -3 & -3
\end{array}
\Rightarrow
\begin{array}{rrrr|r}
1 & 1 & -1 & 2 & 12 \\
0 & -3 & 3 & -5 & -29 \\
0 & -3 & 4 & 2 & -2 \\
0 & 0 & -1 & -7 & -27
\end{array}
$$

$$
\Rightarrow
\begin{array}{rrrr|r}
1 & 1 & -1 & 2 & 12 \\
0 & -3 & 3 & -5 & -29 \\
0 & 0 & 1 & 7 & 27 \\
0 & 0 & -1 & -7 & -27
\end{array}
\Rightarrow
\begin{array}{rrrr|r}
1 & 1 & -1 & 2 & 12 \\
0 & -3 & 3 & -5 & -29 \\
0 & 0 & 1 & 7 & 27 \\
0 & 0 & 0 & 0 & 0
\end{array}
$$

Here, the last equation does not determine the variable t. We can, however, express all the other variables in terms of t. The third step yields $z = 27 - 7t$, the second step: $y = 110/3 - 26/3t$ and eventually we find $x = 7/3 - t/3$. Note that we can write this as the parametric form of a line

$$\vec{r} = \begin{pmatrix} 7/3 \\ 110/3 \\ 27 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1/3 \\ -26/3 \\ -7 \\ 1 \end{pmatrix}.$$

Indeed, we would have the same situation if we had ignored the last equation in the first place. The equation is superfluous; unfortunately, it is often not trivial to recognise which equation can be omitted if a system is underdetermined.

The other case where the method can fail to produce a unique solution is when the system is *overdetermined* or *inconsistent*.

Example 3.11.

$$\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \\ 1 & 2 & 3 & -3 \end{array} \Rightarrow \begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 2 & 2 & -4 \end{array} \Rightarrow \begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & -8 \end{array}$$

In this case, the last equation states: $0 \cdot z = -8$, which is impossible to satisfy for any z. This case often occurs if there are more equations than variables.

All three cases —the case of a unique solution, the overdetermined case, and the underdetermined case — can be understood in geometric terms. We recall that the solution to each equation is a hyperplane in $\mathbb{R}^n$. Hence, the solution to the whole system is the intersection of all these hyperplanes. Two basic situations can be distinguished.

1. The set of normal vectors of the planes is linearly independent. The dimension of the space of solutions is $m = n - k$, where n is the number of unknowns and k is the number of equations, that is, all variables can be expressed in terms of m free variables or parameters. For the case $k = n$, there are no free parameters, and we obtain a unique solution.
2. The set of normal vectors of the hyperplanes is linearly dependent. This is always the case if $k > n$.

    1. One (or more) equations are linear combinations of the others. These equations can be removed from the system, until we are left with a set of linearly independent equations. If the normal vectors are linearly independent, case 1 applies; otherwise, we have case 2b).
    2. The equations are linearly independent and the system has no solution.

In Example 3.9 we had the situation of case 1 with $k = n$ and hence a unique solution.
In Example 3.10, we had the situation that one equation is a linear combination of the others: $(Eq.\ 4) = (Eq.\ 1) + (Eq.\ 2) - (Eq.\ 3)$. This is case 2a) in the above scheme. If we remove this equation, we have situation 1 with $n = 4$, $k = 3$ and therefore $m = 1$, which corresponds to a line as a solution.

In Example 3.11, we had case 2b). The normal vectors of the three planes are

$$\vec{n}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \vec{n}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad \vec{n}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

They are not linearly independent $\vec{n}_3 = \vec{n}_1 + 2\vec{n}_2$. But the equations are linearly independent, i.e. $Eq.3 \neq Eq_1 + 2Eq.2$.

The left figure is the situation in Example 3. The three hyperplanes intersect in three parallel lines. The right figure shows the situation of three linearly independent hyperplanes in $\mathbb{R}3$ which have one point in common. This is situation 1.

3.5 Elementary Matrices

Each of the three transformations we performed on the augmented matrix can be achieved by multiplying the matrix on the left by an *elementary matrix*. The corresponding elementary matrix can be found by applying one of the three elementary row transformations to the identity matrix.

Definition 3.11

An elementary matrix is an $n \times n$ matrix which can be obtained from the identity matrix I_n by performing on I_n a single elementary row transformation.

Example 3.12.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is an elementary matrix. It can be obtained by multiplying row 2 of the identity matrix by 3. In other words, we are performing the row operation $3R_2 \to R_2$.

Example 3.13.

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

is an elementary matrix. It can be obtained by replacing row 3 of the identity matrix by row 3 plus -2 times row 1. In other words, we are performing the row operation $R_3 - 2R_1 \to R_1$.

Example 3.14.

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

is an elementary matrix. It can be obtained by switching rows 1 and 2 of the identity matrix. In other words, we are performing on the row operation $R_1 \leftrightarrow R_2$.

Suppose we want to perform an elementary row operation on a matrix A. In that case, it is equivalent to multiplying the matrix A on the left by the elementary matrix obtained from the identity matrix by the same transformation.

Interchanging Rows: $R_i \leftrightarrow R_j$

To interchange rows i and j of matrix A ($R_i \leftrightarrow R_j$), we multiply A on the left by the elementary matrix obtained from the identity matrix in which rows i and j have been interchanged.

Multiplying a Row by a Constant: $aR_i \to R_i$

To multiply row i of matrix A by a number a ($aR_i \to R_i$), we multiply A on the left by the elementary matrix obtained from the identity matrix in which row i has been multiplied by a.

Replacing a Row by Itself Plus a Multiple of Another: $R_i + aR_j \to R_i$

To replace a row i by itself plus a multiple of another row j ($R_i + aR_j \to R_i$), we multiply A on the left by the elementary matrix obtained from the identity matrix in which row i has been replaced by itself plus row j multiplied by a.

Example 3.15. $R_1 \leftrightarrow R_3$:

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 7 & 8 & 9 \\ 4 & 5 & 6 \\ 1 & 2 & 3 \end{pmatrix}.$$

Example 3.16. $3R_1 \to R_1$:

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 3 & 6 & 9 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Example 3.17. $R_2 - R_1 \to R_2$:

$$\begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 3 & 3 \\ 7 & 8 & 9 \end{pmatrix}.$$

Corollary 3.12: Inverse of elementary matrices

The elementary matrices are nonsingular. Furthermore, their inverse is also an elementary matrix.

- The inverse of the elementary matrix which interchanges two rows is itself. For example

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

- The inverse of the elementary matrix which multiply a row i by a constant a, i.e. $aR_i \to R_i$ is the elementary matrix which multiply a row i by $\frac{1}{a}$, i.e. $\frac{1}{a}R_i \to R_i$. For example,

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

- The inverse of the elementary matrix which replaces a row i by itself plus a multiple of a row j, i.e. $R_i + aR_j \to R_i$ is the elementary matrix which replaces a row i by itself minus a multiple of a row j. For example,

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

3.6 Calculating the inverse

Now we can prove that a matrix is invertible if we can convert it to the identity matrix with elementary row operations. Assume that there exists a sequence of row operations, with corresponding elementary matrices E_i, that converts a matrix A into an identity matrix.

$$\mathrm{I} = \mathrm{E}_k ... \mathrm{E}_2 \mathrm{E}_1 \mathrm{A} \tag{3.2}$$

If we call the product of these elementary matrice **B** then

$$\mathrm{I} = \mathrm{BA} \quad \text{with} \quad \mathrm{B} = \mathrm{E}_k ... \mathrm{E}_2 \mathrm{E}_1 \mathrm{I}$$

To show that B is the inverse, we have to show in addition AB = I. From Eq. 3.2 it follows by multiplying successively with the inverse of the elementary matrices E_j from the left:

$$\begin{aligned} \mathrm{E}_k^{-1}\mathrm{I} &= \underbrace{\mathrm{E}_k^{-1}\mathrm{E}_k}_{=\mathrm{I}} \mathrm{E}_{k-1} ... \mathrm{E}_2 \mathrm{E}_1 \mathrm{A} \\ &\vdots \\ \mathrm{E}_1^{-1} ... \mathrm{E}_k^{-1}\mathrm{I} &= \mathrm{A} \end{aligned}$$

Hence,

$$\begin{aligned} \mathrm{AB} &= \mathrm{E}_1^{-1} ... \mathrm{E}_k^{-1} \mathrm{I} \mathrm{E}_k ... \mathrm{E}_2 \mathrm{E}_1 \mathrm{I} \\ &= \mathrm{E}_1^{-1} ... \underbrace{\mathrm{E}_k^{-1}\mathrm{E}_k}_{=\mathrm{I}} ... \mathrm{E}_2 \mathrm{E}_1 \\ &\vdots \\ &= \mathrm{I} \end{aligned}$$

Hence, we can obtain the inverse of a matrix A (if it exists) by applying the same row operations which convert A into an identity matrix. The method is usually applied to the augmented matrix A|I where any row operations are executed simultaneously on both sides, to reduce it to $\mathrm{I}|\mathrm{A}^{-1}$.

Example 3.18. To calculate A^{-1} when

$$\mathrm{A} = \begin{pmatrix} 1 & 3 & 3 \\ 1 & 4 & 3 \\ 1 & 3 & 4 \end{pmatrix},$$

we reduce the augmented matrix to an identity matrix using elementary row operations.

$$\begin{array}{ccc|ccc} 1 & 3 & 3 & 1 & 0 & 0 \\ 1 & 4 & 3 & 0 & 1 & 0 \\ 1 & 3 & 4 & 0 & 0 & 1 \end{array} \qquad \begin{array}{l} \\ R_2 - R_1 \\ R_3 - R_1 \end{array}$$

$$\Rightarrow \begin{array}{ccc|ccc} 1 & 3 & 3 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{array} \qquad \begin{array}{l} R_1 - 3R_2 - 3R_3 \\ \\ \\ \end{array}$$

$$\Rightarrow \begin{array}{ccc|ccc} 1 & 0 & 0 & 7 & -3 & -3 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{array}$$

Example 3.19. The solution of the equations

$$\begin{pmatrix} 2 & 0 & 1 \\ 1 & -2 & 1 \\ 3 & 1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix}$$

is given by

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -3 & 1 & 2 \\ 2 & -1 & -1 \\ 7 & -2 & -4 \end{pmatrix}\begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} -2 \\ 2 \\ 6 \end{pmatrix}$$

3.7 Factorising Matrices: The LU factorisation

Definition 3.13: Upper/Lower Triangular Matrix

A square matrix is upper (lower) triangular if all its entries below (above)the main diagonal are zero.

So far, we know two methods to solve a system of linear equations

$$A\vec{r} = \vec{q}, \quad \vec{r}, \vec{q} \in \mathbb{R}^n, \mathrm{A} \in M_{n\times n},$$

Gaussian elimination and the inversion of the matrix A. The former uses the augmented matrix $A|\vec{q}$ and elementary row operations to convert the system into a form where the left-hand side is an upper triangular matrix. In many applications, e.g., in algorithms to solve partial differential equations, A is a large matrix (1000 $times$1000 is not unusual), and the system has to be solved repeatedly for various right-hand sides $vecq$. Here, the Gaussian elimination is very inefficient, as for every new $\vec{q}$, the system has to be solved again. Inverting the matrix A seems to be much more efficient since we invert the matrix only once and then only apply A^{-1} to every new $\vec{q}$.

$$\mathrm{A}\vec{r} = \vec{q} \quad \Rightarrow \vec{r} = \mathrm{A}^{-1}\vec{q}.$$

However, the matrix inversion of such large matrices itself is often numerically difficult, “numerically unstable” that is, it tends to produce very large (or very small) numbers, which in turn produce significant numerical errors.

Here, another method, the so-called LU decomposition, has proven to be very efficient. The name is derived from the representation of the matrix A as a product

$$\mathrm{A} = \mathrm{LU}; \quad \mathrm{A}, \mathrm{L}, \mathrm{U} \in M_{n\times n}$$

of two triangular matrices L and U, where L is a lower triangular matrix and U is an upper triangular matrix.

The advantage of this method is that determining L and U is faster (i.e., it uses fewer steps) than inverting A, and we can still solve the system comparatively easily. The methods consist of three steps:

First step: Determine L and U.

Second step: Solve $\mathrm{L}\underbrace{(\mathrm{U}\vec{r})}_{\vec{s}} = \vec{q}$ for $\vec{s}$.

Third step: Solve $\mathrm{U}\vec{r} = \vec{s}$.

Step 1: We know already (see Gauss elimination) that we can use elementary row operations to convert a matrix A to an upper triangular matrix. We have also seen that each of these row operations can be expressed by a matrix E_j, so that

$$\mathrm{U} = \mathrm{E}_k \dots \mathrm{E}_2\mathrm{E}_1\mathrm{A},$$

where k is the number of row operations we need to bring A in an upper triangular form. For the following, we assume that we do not need to exchange rows in this process. (The most general case with the exchange of rows requires a more general representation of A as A = PLU where P includes all permutations of rows.) Then each of the elementary matrices E_j can be chosen as a lower triangular matrix and has an inverse E_j^{-1} which is again a lower triangular matrix, hence in

$$\underbrace{\mathrm{E}_1^{-1}\mathrm{E}_2^{-1} \dots \mathrm{E}_k^{-1}}_{:=\mathrm{L}}\mathrm{U} = \mathrm{A}$$

the product $\mathrm{E}_1^{-1}\mathrm{E}_2^{-1} \dots \mathrm{E}_k^{-1}$ is also a lower triangular matrix.

Example 3.20.

$$\begin{aligned} A = &\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & k \end{pmatrix} \begin{matrix} \\ R_2 - \lambda R_1 \to R_2 \\ R_3 - \mu R_1 \to R_3 \end{matrix} \\ \to &\begin{pmatrix} a & b & c \\ 0 & e' & f' \\ 0 & h' & k' \end{pmatrix} \begin{matrix} \\ \\ R_3 - \omega R_1 \to R_3 \end{matrix} \\ \to &\begin{pmatrix} a & b & c \\ 0 & e' & f' \\ 0 & 0 & k'' \end{pmatrix} = \mathrm{U}. \end{aligned}$$

where $\lambda = d/a$, $\mu = g/a$, $\omega = h'/e'$. Note that this can be accomplished by multiplying A from the left by the elementary matrice

$$\mathrm{E}_1 = \begin{pmatrix} 1 & 0 & 0 \\ -\lambda & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathrm{E}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\mu & 0 & 1 \end{pmatrix}, \quad \mathrm{E}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\omega & 1 \end{pmatrix},$$

which means:

$R_2 - \lambda R_1 \to R_2$:

$$\begin{pmatrix} 1 & 0 & 0 \\ -\lambda & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & k \end{pmatrix} = \begin{pmatrix} a & b & c \\ d-\lambda a & e-\lambda b & f-\lambda c \\ g & h & k \end{pmatrix} = \begin{pmatrix} a & b & c \\ 0 & e' & f' \\ g & h & k \end{pmatrix},$$

$R_3 - \mu R_1 \to R_3$:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\mu & 0 & 1 \end{pmatrix}\begin{pmatrix} a & b & c \\ 0 & e' & f' \\ g & h & k \end{pmatrix} = \begin{pmatrix} a & b & c \\ 0 & e' & f' \\ g-\mu a & h-\mu b & k-\mu c \end{pmatrix} = \begin{pmatrix} a & b & c \\ 0 & e' & f' \\ 0 & h' & k' \end{pmatrix},$$

$R_3 - \omega R_1 \to R_3$:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\omega & 1 \end{pmatrix}\begin{pmatrix} a & b & c \\ 0 & e' & f' \\ 0 & h' & k' \end{pmatrix} = \begin{pmatrix} a & b & c \\ 0 & e' & f' \\ 0 & h'-\omega e' & k'-\omega f' \end{pmatrix} = \begin{pmatrix} a & b & c \\ 0 & e' & f' \\ 0 & 0 & k'' \end{pmatrix}.$$

All three elementary matrices are of the lower triangular type and have an inverse

$$\mathrm{E}_1^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ \lambda & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathrm{E}_2^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & 0 & 1 \end{pmatrix}, \quad \mathrm{E}_3^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \omega & 1 \end{pmatrix}.$$

When we perform the matrix multiplication to obtain L:

$$\mathrm{L} = \mathrm{E}_1^{-1}\mathrm{E}_2^{-1}\mathrm{E}_3^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ \lambda & 1 & 0 \\ \mu & \omega & 1 \end{pmatrix},$$

We notice that this matrix multiplication can be performed by just putting the nonzero off-diagonal entries of the inverses of the elementary matrices into the appropriate positions in the matrix rmL. This means that the entries of L, which are the multiplying factors in the Gaussian elimination process, can be easily stored during the process of Gaussian elimination.

Thus, the matrix A was factorised into the product of the lower triangular matrix L and the upper triangular matrix U as follows

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & k \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \lambda & 1 & 0 \\ \mu & \omega & 1 \end{pmatrix} \begin{pmatrix} a & b & c \\ 0 & e' & f' \\ 0 & 0 & k'' \end{pmatrix}$$

We illustrate the method by a simple example.

Example 3.21. Let

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

We reduce A to upper triangular form in the usual way, but record the multiplying factors in a lower triangular matrix L:

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & -1 & 3 \\ -1 & 2 & 2 \end{pmatrix} \begin{matrix} \\ R_2 - 2R_1 \\ R_3 + R_1 \end{matrix}$$

$$\to \begin{pmatrix} 1 & 2 & 1 \\ 0 & -5 & 1 \\ 0 & 4 & 3 \end{pmatrix} \begin{matrix} \\ \\ R_3 + \frac{4}{5}R_2 \end{matrix}$$

$$\to \begin{pmatrix} 1 & 2 & 1 \\ 0 & -5 & 1 \\ 0 & 0 & 19/5 \end{pmatrix}$$

so L becomes

$$\mathrm{L} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -4/5 & 1 \end{pmatrix}.$$

Now check that

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -4/5 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 0 & -5 & 1 \\ 0 & 0 & 19/5 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & -1 & 3 \\ -1 & 2 & 2 \end{pmatrix}.$$

Once we have this factorisation, we can make use of it to solve $\mathrm{A}\vec{r} = \vec{q}$ as follows:

$$\mathrm{A}\vec{r} = \mathrm{LU}\vec{r} = \vec{q}$$

is equivalent to solving

$$\mathrm{L}\vec{s} = \vec{q} \quad \text{for} \quad \vec{s} = \mathrm{U}\vec{r},$$

using forward substitution, followed by

$$\mathrm{U}\vec{r} = \vec{s}$$

using backwards substitution.

Example 3.22. *Example 3.9.* To solve $\mathrm{A}\vec{r} = \begin{pmatrix} 2 \\ 9 \\ 0 \end{pmatrix}$, first we solve $\mathrm{L}\vec{s} = \vec{q}$ for $\vec{s} = (s_x, s_y, s_z)^T$:

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -4/5 & 1 \end{pmatrix} \begin{pmatrix} s_x \\ s_y \\ s_z \end{pmatrix} = \begin{pmatrix} 2 \\ 9 \\ 0 \end{pmatrix}$$

$$\Rightarrow \begin{cases} s_x = 2, \\ s_y = 9 - 2s_x = 5, \\ s_z = s_x + \frac{4}{5}s_y = 6, \end{cases}$$

$$\Rightarrow \vec{s} = \begin{pmatrix} 2 \\ 5 \\ 6 \end{pmatrix}.$$

Then we solve $\mathrm{U}\vec{r} = \vec{s}$ for $\vec{r} = (x, y, z)^T$:

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & -5 & 1 \\ 0 & 0 & 19/5 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \\ 6 \end{pmatrix}$$

$$\Rightarrow \begin{cases} z = 30/19, \\ y = -\frac{1}{5}(5 - z) = -13/19, \\ x = 2 - 2y - z = 34/19, \end{cases}$$

$$\Rightarrow \vec{r} = \begin{pmatrix} 34/19 \\ -13/19 \\ 30/19 \end{pmatrix}.$$

3.8 Range and Nullspace of a Matrix

Consider a system of linear equations, $A\mathbf{x} = \mathbf{y}$ where A is an $m \times n$ matrix. We can consider the matrix A as a mapping from the space of all $\mathbf{x}$ vectors in $\mathbb{R}^n$ onto the space of all possible $\mathbf{y}$ vectors, which is a subset of $\mathbb{R}^m$.

$$A : \mathbf{x} \in \mathbb{R}^n \longrightarrow \mathbf{y} = A\mathbf{x} \in \mathbb{R}^m$$

.

Definition 3.14: Range and Rank

Let A be an $m \times n$ matrix. The image of $\mathbb{R}^n$ under A is a subspace of $\mathbb{R}^m$ called the range (or column space) of A, denoted by range(A), and

$$\text{range}(A) = \text{span}(\text{columns of} A).$$

The dimension of the column space is called the rank and is denoted by rank(A).

Example 3.23. Calculate the range and rank of

$$\text{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix}.$$

Solution

$$\text{range(A)} = \text{span}\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix}\right)$$

However, the set of vectors (columns of A) is not linearly independent. A test for linear independence yields:

$$c_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + c_2 \begin{pmatrix} 2 \\ 3 \end{pmatrix} + c_3 \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

leads to

$$\begin{cases} c_1 + 2c_2 + 3c_3 = 0 \\ 2c_1 + 3c_2 + 4c_3 = 0 \end{cases} \Leftrightarrow \begin{cases} c_1 = -2c_2 - 3c_3 \\ 2(-2c_2 - 3c_3) + 3c_2 + 4c_3 = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} c_1 = -2c_2 - 3c_3 \\ -c_2 + 2c_3 = 0 \end{cases} \Leftrightarrow \begin{cases} c_1 = c_3 \\ c_2 = -2c_3 \end{cases}$$

That means we have infinitely many solutions. For every choice of c_3 there is a c_2 and c_1, e.g. $(c_1, c_2, c_3) = (1, -2, 1)$ is a solution. We can express for instance the third vector as

$$\begin{pmatrix} 3 \\ 4 \end{pmatrix} = 2 \begin{pmatrix} 2 \\ 3 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

so

$$\text{span}\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix}\right) = \text{span}\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix}\right).$$

This means the two vectors form a basis of the column space (range) and the rank is the dimension of the column space, i.e. the number of elements of the basis: 2.

In order for $A\mathbf{x} = \mathbf{b}$ to have a solution, $\mathbf{b}$ has to be in the range of A: $\mathbf{b} \in \text{range}(A)$.

Example 3.24. Let

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

Check whether the system $A\mathbf{x} = \mathbf{b}$ consistent when

(i) $\mathbf{b} = [1, 1, 1]^T$,

(ii) $\mathbf{b} = [1, 2, 1]^T$.

Under what general conditions on $\mathbf{b} = \left[\mathbf{b_1}, \mathbf{b_2}, \mathbf{b_3}\right]^T$ is the system consistent?

Solution

We reduce A to a row echelon form:

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 3 \\ 0 & 1 & 2 \end{bmatrix} \xrightarrow{R_2 - R_1} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 1 & 2 \end{bmatrix} \xrightarrow{R_3 - R_2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

Hence, for example, the first two columns span range(A) and hence $\mathbf{b} = \lambda [1, 1, 0]^T + \mu [0, 1, 1]^T = \left[\mathbf{b_1}, \mathbf{b_1} + \mathbf{b_3}, \mathbf{b_3}\right]^T$ is necessary for the system to be consistent. This is satisfied for $\mathbf{b} = [1, 2, 1]^T$, but not for $\mathbf{b} = [1, 1, 1]^T$. Note that the first and the third columns of A also span range(A). Even the last two columns can be used.

Definition 3.15: Nullspace and Nullity

Let A be an $m \times n$ matrix. The nullspace of A, is the set of vectors $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$. The nullspace is also called kernel and denoted by Null(A).
The dimension of the nullspace is called the nullity and is denoted by nullity(A).

Corollary 3.16: Nullspace is a subspace

Let A be an $m \times n$ matrix. The nullspace of A is a subspace of $\mathbb{R}^n$.

Proof

We have to check the three condition in the definition for a subspace (2.11):

(i) Since $A\mathbf{0_n} = \mathbf{0_m}, \mathbf{0_n} \in \text{Null}(A)$.

(ii) Let $\mathbf{u}$ and $\mathbf{v}$ be in Null(A). Therefore $A\mathbf{u} = \mathbf{0}$ and $A\mathbf{v} = \mathbf{0}$. It follows that $A(\mathbf{u}+\mathbf{v}) = A\mathbf{u}+A\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0}$ Hence, $\mathbf{u} + \mathbf{v} \in \text{Null}(A)$.

(iii) Finally, for any scalar λ, $A(\lambda\mathbf{u}) = \lambda(A\mathbf{u}) = \lambda\mathbf{0} = \mathbf{0}$ and therefore $\lambda\mathbf{u} \in \text{Null}(A)$. It follows that Null($A$) is a subspace of $\mathbb{R}^n$. $\square$

Example 3.25. Calculate the nullspace and nullity of

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

Solution

$$\text{A}\mathbf{x} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 + 3x_3 \\ 2x_1 + 3x_2 + 4x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

The vector equation is equivalent to two scalar equations for the components:

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 0 \\ 2x_1 + 3x_2 + 4x_3 = 0 \end{cases} \Leftrightarrow \begin{cases} x_1 = -2x_2 - 3x_3 \\ 2(-2x_2 - 3x_3) + 3x_2 + 4x_3 = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} x_1 = -2x_2 - 3x_3 \\ -x_2 + 2x_3 = 0 \end{cases} \Leftrightarrow \begin{cases} x_1 = x_3 \\ x_2 = -2x_3 \end{cases}$$

This indicates that there are infinitely many solutions. We can choose one of the variables, e.g.~x_3 as a free parameter.

$$\mathbb{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_3 \\ -2x_3 \\ x_3 \end{pmatrix} = x_3 \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \quad \lambda \in \mathbb{R}$$

In the last step we replaced x_3 by λ, as it is convention to use Greek letters for parameters. Thus

$$\mathrm{Null(A)} = \mathrm{span}\left(\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}\right)$$

and the nullity of A is 1.

Definition 3.17: Row and column space

The span of the columns of a $m \times n$ matrix A is called the column space. It is a subspace of $\mathbb{R}^n$. The span of the rows of A is called the row space and it is a subspace of $\mathbb{R}^m$.

Theorem 3.18: Dimension of row and column spaces

The row and column spaces of a matrix A have the same dimension.

Proof

Let R be a row echelon form of A, $row(A) = row(R)$, as we used only row operations to convert A to R.

$$\begin{aligned} \dim(\mathrm{row}(A)) &= \dim(\mathrm{row}(R)) \\ &= \text{number of nonzero rows of } R \\ &= \text{number of pivots of } R \end{aligned}$$

Let this number be called γ.
Now $rmcol(A) \neq$ col(R), but the columns of A and R have the same dependence relationships. Therefore, dim(col(A)) = dim(col(R)). Since there are γ pivots, R has γ columns that are linearly independent, and the remaining columns of R are linear combinations of them. Thus, dim(col(R)) = γ. It follows that dim(row(A)) = γ = dim(col(A)), as we wished to prove. □

Theorem 3.19: Rank Theorem

If A is an $m \times n$ matrix, then $\mathrm{rank}(A) + \mathrm{nullity}(A) = n$, where n is the number of columns of A.

Proof

Let R be a row echelon form of A, and suppose that $rank(A) = \gamma$. Then R has γ pivots, so there are γ variables corresponding to the leading entries and $n - \gamma$ free variables in the solution to $A\mathbf{x} = \mathbf{0}$. Since $\dim(\text{Null}(A)) = n - \gamma$, we have $\text{rank}(A) + \text{nullity}(A) = \gamma + (n - \gamma) = n$.

Example 3.26. Find the rank and nullity, and then verify the rank theorem for the matrix:

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

Solution

In order to find the rank we have to check whether the columns are linearly independent. For this we solve

$$c_1 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} + c_3 \begin{pmatrix} 0 \\ -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

This is a system of three equations for the three unkowns c_1, c_2, c_3 and we can either try to directly solve the equations:

$$\Leftrightarrow \begin{cases} c_1 = -c_2 \\ c_1 - c_3 = 0 \\ 2c_2 + 2c_3 = 0 \end{cases} \quad \Leftrightarrow \begin{cases} c_1 = -c_2 \\ c_1 = c_3 \end{cases}$$

or use Gaussian elimination to solve this:

$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 2 & 2 & 0 \end{array}\right] \xrightarrow{R_2 - R_1} \left[\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & -1 & -1 & 0 \\ 0 & 2 & 2 & 0 \end{array}\right] \xrightarrow{R_3 + 2R_2} \left[\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

In either case we see that there are non-trivial solutions, e.g.~$(c_1, c_2, c_3) = (1, -1, 1)$. So the system is linearly dependent. We can remove , e.g.~ the third column vector to obtain a linearly independent set (the first two columns are obviously linearly independent). Hence these can form a basis of the column space and the rank is 2.
The nullity, that is the dimension of the nullspace, is found by an almost identical calculation:

$$\mathrm{A}\mathbf{x} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ x_1 - x_3 \\ 2x_2 + 2x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

as this is the same system that we solved for the linear independence only the c_i are now x_i:

$$\Leftrightarrow \begin{cases} x_1 = -x_2 \\ x_1 - x_3 = 0 \\ 2x_2 + 2x_3 = 0 \end{cases} \quad \Rightarrow \begin{cases} x_1 = -x_2 \\ x_1 = x_3 \end{cases}$$

$$\Rightarrow \mathbb{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ -x_1 \\ x_1 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \quad \lambda \in \mathbb{R}$$

Hence,

$$\text{Null(A)} = \text{span}\left(\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}\right),$$

and the nullity of A is 1.

Example 3.27. Find the nullity of each of the following matrices:

$$M = \begin{bmatrix} 2 & 3 \\ 1 & 5 \\ 4 & 7 \\ 3 & 6 \end{bmatrix}; \qquad N = \begin{bmatrix} 2 & 1 & -2 & -1 \\ 4 & 4 & -3 & 1 \\ 2 & 7 & 1 & 8 \end{bmatrix}$$

Solution

Since the two columns of M are linearly independent, $\text{rank}(M) = 2$. Thus, by the Rank Theorem, $\text{nullity}(M) = 2 - \text{rank}(M) = 2 - 2 = 0$.
To find the nullity of N we first determine the dimension of the row space and the use Theorem 3.18 for find the rank. To find the dimension of the row space we can apply row operations to reduce N to a row echelon form, as row operations don't change the span of the row vectors:

$$\begin{bmatrix} 2 & 1 & -2 & -1 \\ 4 & 4 & -3 & 1 \\ 2 & 7 & 1 & 8 \end{bmatrix} \xrightarrow[R_3-R_1]{R_2-2R_1} \begin{bmatrix} 2 & 1 & -2 & -1 \\ 0 & 2 & 1 & 3 \\ 0 & 6 & 3 & 9 \end{bmatrix} \xrightarrow{R_3-2R_2} \begin{bmatrix} 2 & 1 & -2 & -1 \\ 0 & 2 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We see that there are only two nonzero rows, so $\dim(\text{row space}) = \text{rank}(N) = 2$. Hence, $\text{nullity}(N) = 4 - \text{rank}(N) = 4 - 2 = 2$.

Example 3.28. Let

$$A = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 2 & 1 & 2 & 5 \\ 1 & 1 & 0 & 2 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

Verify the Rank Theorem.

Solution

A row echelon form is given by

$$\begin{bmatrix} 1 & 0 & 1 & 2 \\ 2 & 1 & 2 & 5 \\ 1 & 1 & 0 & 2 \\ 0 & 1 & 0 & 1 \end{bmatrix} \xrightarrow[R_3-R_1]{R_2-2R_1} \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \xrightarrow[R_4-R_2]{R_3-R_2} \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Notice that columns $1, 2$ and 3 have pivots, but not column 4, so columns $1, 2$ and 3 are linearly independent and the $\text{rank}(A) = 3$.
Solving the equation $A\mathbf{x} = \mathbf{0}$, we see that all solutions are of the form $\mathbf{x} = c\,[1, 1, 1, -1]^T$, so that

the dimension of the nullspace is 1. Thus, rank(A) + nullity$(A) = 3 + 1 = 4$, verifying the Rank Theorem.

Example 3.29. Find the rank and nullity of

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

Identify bases for both the range and nullspace of A.

Solution

A row echelon form is given by

$$\begin{bmatrix} 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 & 0 \\ 1 & 2 & 1 & 3 & 1 \\ 2 & 3 & 1 & 4 & 2 \end{bmatrix} \xrightarrow[R_4-2R_1]{R_3-R_1} \begin{bmatrix} 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 & 0 \\ 0 & 1 & 1 & 2 & 0 \\ 0 & 1 & 1 & 2 & 0 \end{bmatrix} \xrightarrow[R_4-R_2]{R_3-R_2} \begin{bmatrix} 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Hence rank$(A) = 2$ and nullity$(A) = 5-$rank$(A) = 5-2 = 3$. Columns 1 and 2 have pivots, therefore a basis for range(A) is $\left\{[1,0,1,2]^T, [1,1,2,3]^T\right\}$.
Let $x_3 = r, x_4 = s$ and $x_5 = t$, then $x_1 = r + s - t, x_2 = -r - 2s$, so that

$$\mathbf{x} = r\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + s\begin{bmatrix} 1 \\ -2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

So a basis for the nullspace is

$$\left\{[1,-1,1,0,0]^T, [1,-2,0,1,0]^T, [-1,0,0,0,1]^T\right\}$$

Summary of the procedure to find a basis for the nullspace of A.

1. Find a row echelon form R of A.
2. Solve for the leading variables (x_1 and x_2 above) of $R\mathbf{x} = \mathbf{0}$ in terms of the free variables (x_3, x_4 and x_5 in the example).
3. Set the free variables equal to parameters, substitute back into $\mathbf{x}$, and write the results as a linear combination of f vectors (where f is the number of free variables). These f vectors form a basis for Null(A).

Note 1: Remark: Nullspace of non-square matrices

A non-square matrix, $m \times n$, with $n > m$ must always have a non-trivial nullspace, i.e. a nullity > 0. The reason is that the rank can be at most m as the dimension of column space and row space are the same (Theorem 3.18). Then the rank theorem implies that the nullity is at least 1. This is the reason why a non-square matrix can't have an inverse.

4 Determinants

4.1 Introduction

Determinants occur in many situations, geometrically they can beinterpreted as the volume of the parallelepiped spanned by a set of n vectors. This is easily visualised in two dimensions (x-y-plane). The area (2-dimensional volume) of a parallelepiped spanned by the two vectors $\mathbf{b} = (b_1, b_2)^T$ and $\mathbf{c} = (c_1, c_2)^t$ is up to a sign given by

$$area = \pm(b_1 c_2 - b_2 c_1)$$

We call this the determinant of the matrix A = [**b**, **c**] and write

$$\det(\mathrm{A}) = |\mathrm{A}| = \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix} = b_1 c_2 - b_2 c_1.$$

The determinant of a square matrix A, denoted by $\det(A)$, or $|A|$, is a number (scalar). The now-standard notation was first introduced by Cayley in 1841.

In three dimensions, the volume of the cell spanned by the vectors **a**, **b**, and **c** is given by

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = a_1(b_2 c_3 - b_3 c_2) + a_2(b_3 c_1 - b_1 c_3) + a_3(b_1 c_2 - b_2 c_1).$$

Note that we can write this in terms of the determinants of 2x2 matrices:

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix},$$

where each of the 2×2 matrices is obtained by deleting the first column and the corresponding row of the coefficient a_j from the matrix A = [**a**, **b**, **c**]. This suggests a recursive scheme where determinants of arbitrary $n \times n$ matrices are defined in terms of determinants of $(n-1) \times (n-1)$ matrices, which again are expanded in determinants of $(n-2) \times (n-2)$ and so on.

Example 4.1. Evaluate

$$\begin{vmatrix} 1 & 2 & 3 \\ 2 & 3 & 2 \\ 3 & 3 & 4 \end{vmatrix}.$$

Definition 4.1: Minor and Cofactor

Corresponding to each entry $a_{i,j}$ in an $n \times n$ matrix A, we define a number M_{ij}, called the *minor* of a_{ij}, which is the $(n-1)$th order determinant obtained by deleting the i-th row and j-th column from A.
The cofactor C_{ij} corresponding to an entry a_{ij} in an $n \times n$ matrix A is the product of its minor and the sign $(-1)^{i+j}$: $C_{ij} = (-1)^{i+j} M_{ij}$. The n^2 cofactors form a matrix of cofactors.

Definition 4.2: Determinants

A determinant of an $n \times n$ matrix can be expanded in terms of $(n-1)\times(n-1)$ determinants using either a column or a row. Expansion along the i-th column:

$$|A| = a_{1,i}C_{1,i} + a_{2,i}C_{2,i} + a_{3,i}C_{3,i} + \cdots + a_{n,i}C_{n,i}.$$

Expansion along the i-th row:

$$|A| = a_{i,1}C_{i,1} + a_{i,2}C_{i,2} + a_{i,3}C_{i,3} + \cdots + a_{i,n}C_{i,n}.$$

The $(n-1)\times(n-1)$ determinants can then be recursively further reduced down to 2×2 matrices.

Example 4.2. Find the expansion for the matrix A of the previous Example. Expansion along the first row:

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

Expansion along the first column:

$$\begin{vmatrix} 1 & 2 & 3 \\ 2 & 3 & 2 \\ 3 & 3 & 4 \end{vmatrix} = 1\begin{vmatrix} 3 & 2 \\ 3 & 4 \end{vmatrix} - 2\begin{vmatrix} 2 & 3 \\ 3 & 4 \end{vmatrix} + 3\begin{vmatrix} 2 & 3 \\ 3 & 2 \end{vmatrix} = 6 + 2 - 15 = -7$$

Expansion along the second column:

$$\begin{vmatrix} 1 & 2 & 3 \\ 2 & 3 & 2 \\ 3 & 3 & 4 \end{vmatrix} = -2\begin{vmatrix} 2 & 2 \\ 3 & 4 \end{vmatrix} + 3\begin{vmatrix} 1 & 3 \\ 3 & 4 \end{vmatrix} - 3\begin{vmatrix} 1 & 3 \\ 2 & 2 \end{vmatrix} = -4 - 15 + 12 = -7$$

Example 4.3. Evaluate

$$\begin{vmatrix} 1 & 2 & 1 & 3 \\ 1 & 1 & 2 & 1 \\ 3 & 2 & 0 & 2 \\ 2 & 2 & 1 & 0 \end{vmatrix} = 1\begin{vmatrix} 1 & 2 & 1 \\ 2 & 0 & 2 \\ 2 & 1 & 0 \end{vmatrix} - 1\begin{vmatrix} 2 & 1 & 3 \\ 2 & 0 & 2 \\ 2 & 1 & 0 \end{vmatrix} + 3\begin{vmatrix} 2 & 1 & 3 \\ 1 & 2 & 1 \\ 2 & 1 & 0 \end{vmatrix} - 2\begin{vmatrix} 2 & 1 & 3 \\ 1 & 2 & 1 \\ 2 & 0 & 2 \end{vmatrix}$$

$$= 1\begin{vmatrix} 0 & 2 \\ 1 & 0 \end{vmatrix} - 2\begin{vmatrix} 2 & 2 \\ 2 & 0 \end{vmatrix} - \ldots 10 \text{ more terms}$$

The previous example should illustrate the large amount of computation needed by using a "brute force" approach to evaluating determinants. We will now look at smarter ways, using two fundamental theorems to develop a simple numerical procedure closely related to Gaussian elimination. The following example shows that it is much more efficient to calculate determinants with many zeros in columns or rows:

Example 4.4. Evaluate

$$\begin{vmatrix} 1 & 2 & 1 & 3 \\ 0 & 1 & 1 & -2 \\ 0 & 0 & 7 & -1 \\ 0 & 0 & 3 & 2 \end{vmatrix}$$

Solution

Expanding along the first column, twice:

$$\begin{vmatrix} 1 & 2 & 1 & 3 \\ 0 & 1 & 1 & -2 \\ 0 & 0 & 7 & -1 \\ 0 & 0 & 3 & 2 \end{vmatrix} = \begin{vmatrix} 1 & 1 & -2 \\ 0 & 7 & -1 \\ 0 & 3 & 2 \end{vmatrix} = \begin{vmatrix} 7 & -1 \\ 3 & 2 \end{vmatrix} = 17.$$

4.2 Simplification of Determinants

The following rules help to simplify determinants efficiently:

Corollary 4.3: Calculation of Determinants

1. If two rows (columns) of A are interchanged to give a matrix B then $|B| = -|A|$.
2. If two rows (columns) of A are equal then $|A| = 0$
3. If all the entries in any row (column) of A are multiplied by a scalar k, the determinant is also multiplied by k
4. If one row (column) of A is a multiple of another row (column), then $|A| = 0$
5. If the matrix B is obtained from A by taking multiple of one row (column) and adding it to another row (column), then $|B| = |A|$.
6. $|A| = \left|A^T\right|$.

Proof

The proof of (1) and (6) is omitted.
(2) is proved by using (1): If we interchange the two equal columns of A, then we obtain A again, hence |A| remains the same, but according to (1) it should change its sign. This is only possible if |A| = 0.
(3) If B is obtained by multiplying the ith column of A by k :

$$|A| = \begin{vmatrix} a_1 & b_1 & c_1 & \cdots \\ a_2 & b_2 & c_2 & \\ \vdots & & & \\ a_i & b_i & c_i & \\ \vdots & & & \end{vmatrix}, \quad |B| = \begin{vmatrix} a_1 & b_1 & c_1 & \cdots \\ a_2 & b_2 & c_2 & \\ \vdots & & & \\ ka_i & kb_i & kc_i & \\ \vdots & & & \end{vmatrix}$$

and expand both determinants about their ith rows:

$$\begin{aligned} |A| &= a_i C_{i,1} + b_i C_{i,2} + \cdots \\ |B| &= \left(ka_i\right) C_{i,1} + \left(kb_i\right) C_{i,2} + \cdots \\ &= k\left(a_i C_{i,1} + b_i C_{i,2} + \cdots\right) = k\,|A|. \end{aligned}$$

(4) We use (3) and (2).

(5) Suppose we add $k\times$ row j to row i of $|A|$ to give $|B|$:

$$|A| = \begin{vmatrix} a_1 & b_1 & c_1 & \cdots \\ a_2 & b_2 & c_2 & \\ \vdots & & & \\ a_i & b_i & c_i & \\ \vdots & & & \end{vmatrix}, \quad |B| = \begin{vmatrix} a_1 & b_1 & c_1 & \cdots \\ a_2 & b_2 & c_2 & \\ \vdots & & & \\ a_i + ka_j & b_i + kb_j & c_i + kc_j & \\ \vdots & & & \end{vmatrix}$$

and we now expand $|B|$ by its ith row

$$\begin{aligned} |A| &= a_i C_{i,1} + b_i C_{i,2} + \cdots \\ |B| &= \left(a_i + ka_j\right) C_{i,1} + \left(b_i + kb_j\right) C_{i,2} + \cdots \\ &= \left(a_i C_{i,1} + b_i C_{i,2} + \cdots\right) + k\left(a_j C_{i,1} + b_i C_{j,2} + \cdots\right) \\ &= \begin{vmatrix} a_1 & b_1 & c_1 & \cdots \\ a_2 & b_2 & c_2 & \\ \vdots & & & \\ a_i & b_i & c_i & \\ \vdots & & & \end{vmatrix} + k \begin{vmatrix} a_1 & b_1 & c_1 & \cdots \\ a_2 & b_2 & c_2 & \\ \vdots & & & \\ a_j & b_j & c_j & \\ \vdots & & & \end{vmatrix} = |A| \end{aligned}$$

since rows i and j in the determinant $\begin{vmatrix} a_1 & b_1 & c_1 & \cdots \\ a_2 & b_2 & c_2 & \\ \vdots & & & \\ a_j & b_j & c_j & \\ \vdots & & & \end{vmatrix}$ are identical, so it is zero (Corollary 4.3, item 4). □

Remember

The rules are similar to, **but not the same as,** Gaussian elimination; adding k times row 2 to row 1 is OK if the result goes in row 1, and row 2 is left unchanged, but if you replace row 2 by k times row 2 plus row 1, you will change the value of the determinant by a factor of k (item 3 of Corollary 4.3 tells us this). A further difference is that a multiple of a column can be added to another column (not allowed in Gaussian elimination).

Example 4.5. Evaluate the determinant $\begin{vmatrix} 1 & 2 & -1 & 3 \\ 2 & 0 & 1 & 1 \\ 1 & 3 & 2 & 1 \\ 2 & 0 & 0 & 1 \end{vmatrix}$.

Solution

First expand by the 4th row:

$$= -2 \times \begin{vmatrix} 2 & -1 & 3 \\ 0 & 1 & 1 \\ 3 & 2 & 1 \end{vmatrix} + 1 \times \begin{vmatrix} 1 & 2 & -1 \\ 2 & 0 & 1 \\ 1 & 3 & 2 \end{vmatrix},$$

now subtract c_3 from c_2 in the 1st , subtract $2 \times c_3$ from c_1 in the 2nd and expand both by their 2nd rows.

$$= -2 \times \begin{vmatrix} 2 & -4 & 3 \\ 0 & 0 & 1 \\ 3 & 1 & 1 \end{vmatrix} + 1 \times \begin{vmatrix} 3 & 2 & -1 \\ 0 & 0 & 1 \\ -3 & 3 & 2 \end{vmatrix}$$

$$= -2 \times (-1) \begin{vmatrix} 2 & -4 \\ 3 & 1 \end{vmatrix} + 1 \times (-1) \begin{vmatrix} 3 & 2 \\ -3 & 3 \end{vmatrix}$$

$$= -2 \times (-1) \times 14 + 1 \times (-1) \times 15 = 13.$$

Example 4.6. Evaluate the determinant

$$\begin{vmatrix} 4 & 1 & 3 & -1 \\ 2 & 0 & 1 & 2 \\ 1 & -1 & 2 & 5 \\ 2 & 1 & 3 & 1 \end{vmatrix}$$

.

Solution

Subtract r_4 from r_1; then add c_1 to c_4, then expand about the 1st row:

$$r_1 - r_4 \to r_1 : \quad \begin{vmatrix} 2 & 0 & 0 & -2 \\ 2 & 0 & 1 & 2 \\ 1 & -1 & 2 & 5 \\ 2 & 1 & 3 & 1 \end{vmatrix},$$

$$c_4 + c_1 \to c_4 : \quad \begin{vmatrix} 2 & 0 & 0 & 0 \\ 2 & 0 & 1 & 4 \\ 1 & -1 & 2 & 6 \\ 2 & 1 & 3 & 3 \end{vmatrix} = 2 \times \begin{vmatrix} 0 & 1 & 4 \\ -1 & 2 & 6 \\ 1 & 3 & 3 \end{vmatrix}$$

add r_3 to r_2 :

$$r_2 + r_3 \to r_2 : \quad 2 \times \begin{vmatrix} 0 & 1 & 4 \\ 0 & 5 & 9 \\ 1 & 3 & 3 \end{vmatrix} = 2 \times 1 \times \begin{vmatrix} 1 & 4 \\ 5 & 9 \end{vmatrix} = 2 \times 1 \times (-11) = -22.$$

Show that

$$D = \begin{vmatrix} 1 & a & a^2 & b+c+d \\ 1 & b & b^2 & c+d+a \\ 1 & c & c^2 & d+a+b \\ 1 & d & d^2 & a+b+c \end{vmatrix} = 0$$

for all values of a, b, c, d.

Solve for x

$$\begin{vmatrix} 1 & 2 & x \\ x & 0 & 3 \\ 1 & x & 2 \end{vmatrix} = 0$$

Corollary 4.4

If A is either lower or upper triangular, then $\det(A) = a_{1,1}a_{2,2}\cdots a_{n,n}$, is the product of its diagonal entries.

Example 4.7.

$$\begin{vmatrix} 1 & 7 & -2 \\ 0 & 14 & 32 \\ 0 & 0 & -2 \end{vmatrix} = 1 \times 14 \times (-2) = -28$$

.

We can see this by expanding down the first column:

$$1\begin{vmatrix} 14 & 32 \\ 0 & -2 \end{vmatrix} = 1(14 \times (-2)) = -28.$$

Theorem 4.5: Product Rule

For any two matrices A and B, $\det(AB) = \det(A)\det(B)$.

The general proof of this result uses elementary matrices. Although it is a key result, a proof will not be given (See Poole, p. 268).

Note that $\det(A + B) \neq \det(A) + \det(B)$.

Corollary 4.6

If A is $n \times n$, we have, by Corollary 4.3

$$\begin{aligned} \det(kA) &= \det(kI\ A) \\ &= \det(kI)\det(A) \\ &= k^n \det(A). \end{aligned}$$

Corollary 4.7

When A is non-singular (i.e., invertible)

$$\det\left(A^{-1}\right) = \frac{1}{\det(A)}.$$

Proof

$1 = \det(\mathrm{I}) = \det(AA^{-1}) = \det(A)\det(A^{-1})$. □

Example 4.8. Let A be an $n \times n$ matrix with $\det(A) = 5$. Then

$$\det\left(A^4\right) = 5^4$$

$$\det\left(A^{-1}\right) = \frac{1}{5},$$

$$\det\left(3A^2\right) = 3^n \times 25$$

.

Corollary 4.8

The matrix A has an inverse if, and only if, $|A| \neq 0$.

Proof

First, we show that if the matrix A has an inverse, then $|A| \neq 0$. If A has an inverse then $|\mathrm{AA}^{-1}| = |\mathrm{A}||\mathrm{A}^{-1}| = |I| = 1$ and hence $|\mathrm{A}| \neq 0$.
Next, we show the reverse. Assume that $|A| \neq 0$. We want to show that A has an inverse.
By using elementary row operations, that is by either (i) adding a multiple of one row to another or (ii) interchanging rows we can reduce A to upper triangular form—call this U. Operations of type i) do not affect the value of $|A|$ while operations of type (ii) may cause a change of sign. Hence,

$$|A| = \pm |U|,$$

and so, in particular $|A| \neq 0$.

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & & u_{2n} \\ 0 & 0 & \ddots & \\ 0 & 0 & & u_{nn} \end{bmatrix},$$

$$|U| = u_{11}u_{22}\cdots u_{nn} \neq 0,$$

so $u_{ij} \neq 0$ for all $i, j = 1, 2, \ldots, n$. Hence A has n pivots and is invertible. □

Corollary 4.9

If A is a square matrix, the system $A\mathbf{x} = \mathbf{0}$ has non-trivial solutions if, and only if, $|A| = 0$ (so A is a singular matrix).

Note:

Various authors define non-singular matrices to be either

- invertible matrices
- matrices A for which $\det(A) \neq 0$.
- matrices for which $A\mathbf{x} = \mathbf{0}$ has only the trivial (zero) solution.

The above theorem shows that they are all equivalent definitions.

4.3 The Adjoint Matrix

Definition 4.10

The transpose of the matrix of cofactors of A is called the *adjugate* or *adjoint* of A and denoted by $\mathrm{adj(A)} = (\mathrm{Cofactors})^{\mathrm{T}}$.

Theorem 4.11: Inverse and Adjoint

If $\det(\mathrm{A}) \neq 0$ then

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

Proof

Let C be the matrix of cofactors of A.

$$\begin{aligned}
(\mathrm{A\ adj(A)})_{\mathrm{ij}} &= \sum_{k=1}^{n} a_{ik}(\mathrm{C}^T)_{kj} = \sum_{k=1}^{n} a_{ik}\mathrm{C}_{jk} \\
&= \begin{cases} 0 & i \neq j \\ \det(\mathrm{A}) & i = j \end{cases} \\
\Rightarrow \mathrm{A\ adj(A)} &= \det(\mathrm{A})\mathrm{I}
\end{aligned}$$

Here, the result for $i \neq j$ is obtained from the fact that $\sum_{k=1}^{n} a_{ik}\mathrm{C}_{jk}$ is the expansion of the determinant of a matrix with two identical rows. For the case $i = j$, however, this is exactly the definition of the expansion of a determinant along row i of A. So

$$\begin{aligned}
& A\mathrm{adj}(A) = \det(A)I \\
\Leftrightarrow \quad & \frac{A\mathrm{adj}(A)}{\det(A)} = I.
\end{aligned}$$

Multiplying A^{-1} from the left gives

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

□

Example 4.9. Find the adjoint of the matrix

$$\mathrm{A} = \begin{bmatrix} 1 & -1 & 2 \\ 3 & 2 & -1 \\ 2 & -1 & 2 \end{bmatrix}$$

Solution

(i) Matrix of minors is

$$\begin{bmatrix} 3 & 8 & -7 \\ 0 & -2 & 1 \\ -3 & -7 & 5 \end{bmatrix}.$$

(ii) The matrix of cofactors is

$$\begin{bmatrix} 3 & -8 & -7 \\ 0 & -2 & -1 \\ -3 & 7 & 5 \end{bmatrix}.$$

(iii) Adjoint is

$$\text{adj(A)} = \begin{bmatrix} 3 & 0 & -3 \\ -8 & -2 & 7 \\ -7 & -1 & 5 \end{bmatrix}.$$

Notice that

$$\text{A adj(A)} = \begin{bmatrix} 1 & -1 & 2 \\ 3 & 2 & -1 \\ 2 & -1 & 2 \end{bmatrix} \begin{bmatrix} 3 & 0 & -3 \\ -8 & -2 & 7 \\ -7 & -1 & 5 \end{bmatrix}$$

$$= \begin{bmatrix} -3 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -3 \end{bmatrix} = -3I = det(A)I.$$

Example 4.10. Use the adjoint matrix to find the inverse of the matrix $A = \begin{pmatrix} 1 & -1 & 2 \\ 3 & 2 & -1 \\ 2 & -1 & 2 \end{pmatrix}$ in the previous example.

Solution

The determinant of the matrix A is

$$1 \begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} + 1 \begin{vmatrix} 3 & -1 \\ 2 & 2 \end{vmatrix} - 2 \begin{vmatrix} 3 & 2 \\ 2 & -1 \end{vmatrix}$$

$$= (4 - 1) + (6 + 2) + 2(-3 - 4)$$
$$= 3 + 8 - 14$$
$$= -3.$$

Thus, the inverse is

$$A^{-1} = \frac{-1}{3} \begin{pmatrix} 3 & 0 & -3 \\ -8 & -2 & 7 \\ -7 & -1 & 5 \end{pmatrix}.$$

Example 4.11. Find the inverse of

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

Solution

$$
\begin{aligned}
\det(A) &= 2\begin{vmatrix} 6 & 7 \\ 9 & 1 \end{vmatrix} - 3\begin{vmatrix} 5 & 7 \\ 8 & 1 \end{vmatrix} + 4\begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} \\
&= 2(6-63) - 3(5-56) + 4(45-48) \\
&= 2(-57) - 3(-51) + 4(-3) \\
&= -114 + 153 - 12 \\
&= 27.
\end{aligned}
$$

$$
\begin{aligned}
A^{-1} &= \frac{1}{27}\begin{pmatrix} -57 & 33 & -3 \\ 51 & -30 & 6 \\ -3 & 6 & -3 \end{pmatrix} \\
&= \frac{1}{9}\begin{pmatrix} -19 & 11 & -1 \\ 17 & -10 & 2 \\ -1 & 2 & -1 \end{pmatrix}.
\end{aligned}
$$

Example 4.12. In the 2×2 case, if

$$
\begin{aligned}
A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} &\Rightarrow \text{adj(A)} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \\
\Rightarrow A^{-1} &= \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.
\end{aligned}
$$

4.4 Inverse Matrix theorem

The following theorem combines a number of topics, inverse matrices, determinants, range, and nullspace.

Theorem 4.12: Inverse Matrix theorem

The following statements are equivalent for a $n \times n$ matrix A:

1. the matrix A has an inverse
2. the determinant of A is non-zero
3. the columns of A are linearly independent
4. the range of A is $\mathbb{R}^n$
5. the rank of A is n
6. the nullspace of A is $\{\vec{0}\}$
7. the equation $A\vec{x} = \vec{0}$ has only the trivial solution $\vec{x} = \vec{0}$
8. the nullity of A is 0
9. the equation $A\vec{x} = \vec{b}$ has a unique solution for any $\vec{b} \in \mathbb{R}^n$
10. the mapping $\vec{x} \to \vec{y} = A\vec{x}$ is one-to-one
11. the transpose of A is invertible
12. the rows of A are linearly independent

Proof

- 1) $\Leftrightarrow$ 2) : Corollary 4.8.
- 2) $\Leftrightarrow$ 3) : Definition of a determinant as the volume spanned by n (column vectors).
- 3) $\Leftrightarrow$ 4) : n linearly independent vectors in $\mathbb{R}^n$ span $\mathbb{R}^n$ (Corollary 2.10)
- 4) $\Leftrightarrow$ 5) : Definition of rank as the dimension of the range ($\Rightarrow$). For the reverse ($\Leftarrow$) note that rank = n means ther exists a basis for the space with n elements.
- 5) $\Leftrightarrow$ 6) : The rank theorem 3.19.
- 6) $\Leftrightarrow$ 7) : Definition of the nullspace
- 7) $\Leftrightarrow$ 8) : Definition of nullity as the dimension of the nullspace. Vice versa if the dimension of the nullspace is zero, then the nullspace must consist of $\{0\}$, as the zero-vector is always in the nullspace.
- 8) $\Leftrightarrow$ 9) : We know from 4) that the range is the whole of $\mathbb{R}^n$ and if the solutions were not unique then there would be non-trivial solutions in the nullspace. (assume two different $\vec{x}$ map to the same $\vec{b}$: $A\vec{x}_1 = \vec{b}$ and $A\vec{x}_2 = \vec{b}$ then $A(\vec{x}_1 - \vec{x}_2) = \vec{0}$ and $(\vec{x}_1 - \vec{x}_2)$ would be a non-trivial vector in the nullspace.
- 9) $\Leftrightarrow$ 10) : same statement.
- 10) $\Leftrightarrow$ 11) : The transpose has the same determinant as A and we can then use the equivalence of 1) and 2).
- 11) $\Leftrightarrow$ 12) : Statement 3) for the transpose, together with 11).

5 Eigenvalues and Eigenvectors

5.1 Introduction to Eigenvalues and Eigenvectors

Consider an elastic body, e.g., a rubber ball. If we compress the ball in one direction, it will be elongated in the plane perpendicular to that direction. The figure below shows, in a cross-section, how individual points on the surface move under the compression.

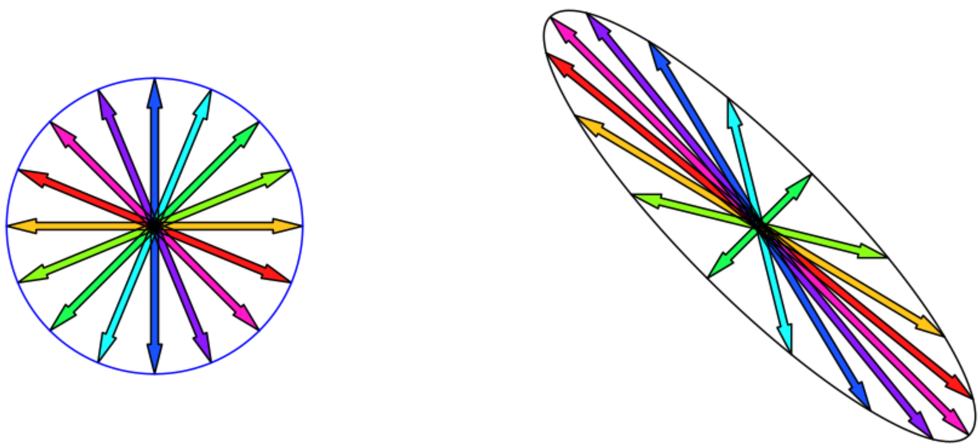


Figure 5.1: Effect of a deformation on the points (vectors) in an elastic sphere.

There are some vectors in these plots that don't change direction, but only their length under the deformation. We can find those by superimposing the two plots:

These directions are characteristic of the deformation, as is the factor of elongation/compression along these directions. These directions can be found mathematically as the eigenvectors of a matrix, the deformation matrix, which maps the original vectors to the vectors of the deformed ball. The eigenvalues are the factors of compression/elongation along these directions. The word "eigen" is German for "own" and was likely introduced by David Hilbert (also known for Hilbert spaces).

Eigenvalues and eigenvectors play an essential role in mathematics, physics and engineering. The stability of an equilibrium of a dynamical system is determined by the eigenvalues of a matrix that describes the linearised system at the equilibrium point. The values of a measurement in quantum mechanics are the eigenvalues of an operator. The principal axes of a rigid body are the eigenvectors of the moment of inertia tensor.

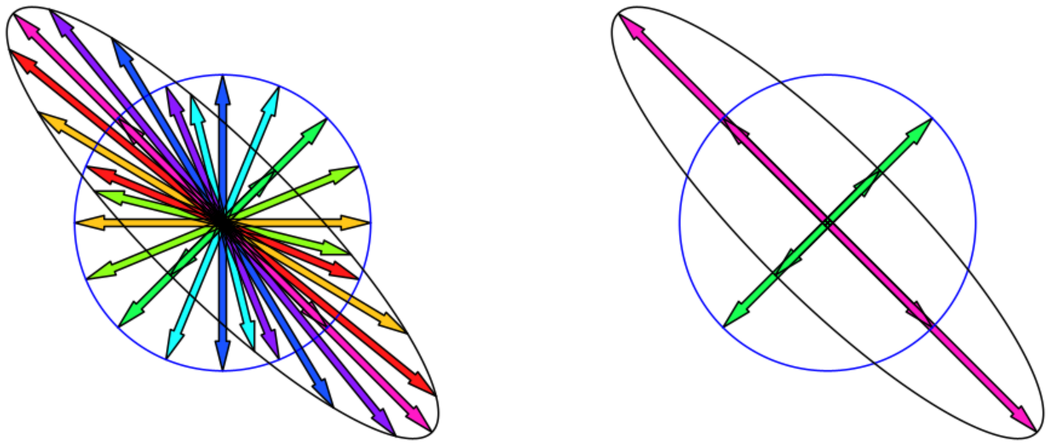


Figure 5.2: There are directions in which the vectors only change their length.

Definition 5.1: Eigenvalue and Eigenvector

Let A be an $n \times n$ matrix. A scalar λ is called an eigenvalue of A if there is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$. Such a vector $\mathbf{x}$ is called an eigenvector of A corresponding to λ.

Example 5.1. Suppose $A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2x \\ 0 \end{bmatrix}$. Then

$$A \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

so 2 is an eigenvalue and $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ a corresponding eigenvector. Also,

$$A \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0 \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

so 0 is an eigenvalue and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ a corresponding eigenvector.

Notice that $\begin{bmatrix} \alpha \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ \alpha \end{bmatrix}$ are eigenvectors for any $\alpha \neq 0$.
In general, if $\mathbf{v}$ is an eigenvector of A, then so is $\alpha\mathbf{v}$ for any nonzero scalar α.

Example 5.2. Show that 5 is an eigenvalue of $A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$ and determine all eigenvectors corresponding to this eigenvalue.

Solution

We must show that there is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = 5\mathbf{x}$, which is equivalent to the equation $(A - 5I)\mathbf{x} = \mathbf{0}$. We compute the nullspace by:

$$[A - 5I|\mathbf{0}] = \left[\begin{array}{cc|c} -4 & 2 & 0 \\ 4 & -2 & 0 \end{array}\right] \xrightarrow{R_2+R_1} \left[\begin{array}{cc|c} -4 & 2 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

Thus $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \text{Null}(A - 5I)$ satisfies $-4x_1 + 2x_2 = 0$, or $x_2 = 2x_1$.

Thus, $A\mathbf{x} = 5\mathbf{x}$ has a nontrivial solution of the form $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ 2x_1 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, so 5 is an eigenvalue of A and the corresponding eigenvectors are the nonzero multiples of $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

Definition 5.2: Eigenspace

Let A be an $n \times n$ matrix and let λ be an eigenvalue of A. The collection of all eigenvectors corresponding to λ, together with the zero vector, is called the eigenspace of λ and is denoted by E_λ.

Therefore, in the above Example, $E_5 = \left\{ t \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$, where $t \in \mathbb{R}$.

Theorem 5.3: Characteristic Equation

Let A be an $n \times n$ matrix. Then λ is an eigenvalue of A if and only if $|A - \lambda I_n| = 0$ (or $\det(A - \lambda I_n) = 0$).

Proof

Suppose that λ is an eigenvalue of A. Then $A\mathbf{v} = \lambda\mathbf{v}$ for some nonzero $\mathbf{v} \in \mathbb{R}^n$. This is equivalent to $A\mathbf{v} = \lambda I_n \mathbf{v}$ or $(A - \lambda I_n)\mathbf{v} = \mathbf{0}$. But this means that $\mathbf{v}$ is a nonzero solution to the homogeneous system of equations defined by the matrix $A - \lambda I_n$. This means $A - \lambda I_n$ is singular, and so $|A - \lambda I_n| = 0$. Conversely, if $|A - \lambda I_n| = 0$ then $A - \lambda I_n$ is singular, and so the system of equations defined by $A - \lambda I_n$ has nonzero solutions. Hence there exists a nonzero $\mathbf{v} \in \mathbb{R}^n$ with $(A - \lambda I_n)\mathbf{v} = \mathbf{0}$, which is equivalent to $A\mathbf{v} = \lambda\mathbf{v}$, and so λ is an eigenvalue of A. □

Definition 5.4: Characteristic Equation/Polynomial

For an $n \times n$ matrix A, the equation $|A - \lambda I_n| = 0$ is called the characteristic equation of A, and $|A - \lambda I_n|$ is called the characteristic polynomial of A.

Example 5.3. Find the eigenvalues and the corresponding eigenvectors of $A = \begin{bmatrix} 1 & 2 \\ 4 & -1 \end{bmatrix}$.

Solution

The characteristic polynomial is

$$\begin{aligned}|A-\lambda I_2| &= \begin{vmatrix} 1-\lambda & 2 \\ 4 & -1-\lambda \end{vmatrix} \\ &= (1-\lambda)(-1-\lambda)-8 \\ &= \lambda^2-9 \\ &= (\lambda+3)(\lambda-3).\end{aligned}$$

Hence the eigenvalues of A are the roots of $(\lambda+3)(\lambda-3)=0$; that is $\lambda_1=-3$ and $\lambda_2=3$.
To find the eigenvectors corresponding to $\lambda_1=-3$, we find the nullspace of

$$A-(-3)I_2 = \begin{bmatrix} 4 & 2 \\ 4 & 2 \end{bmatrix}$$

Row reduction produces

$$\begin{bmatrix} A+3I_2 | \mathbf{0} \end{bmatrix} = \left[\begin{array}{cc|c} 4 & 2 & 0 \\ 4 & 2 & 0 \end{array}\right] \xrightarrow{R_2-R_1} \left[\begin{array}{cc|c} 4 & 2 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

Thus $\mathbf{x} \in \text{Null}(A+3I_2)$ if and only if $4x_1+2x_2=0$.
Setting the free variable $x_2=t$, we see that $x_1=-\frac{1}{2}t$. We take $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ be our eigenvector; or indeed any nonzero multiple of $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$.
To find the eigenvectors corresponding to $\lambda_2=3$, we find the nullspace of $A-3I_2$ by row reduction:

$$\begin{bmatrix} A-3I_2 | \mathbf{0} \end{bmatrix} = \left[\begin{array}{cc|c} -2 & 2 & 0 \\ 4 & -4 & 0 \end{array}\right] \xrightarrow{R_2+2R_1} \left[\begin{array}{cc|c} -2 & 2 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

So $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \text{Null}(A-3I_2)$ if and only if $-2x_1+2x_2=0$. Setting the free variable $x_2=t$, we find $x_1=t$. We take $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ to be our eigenvector.

Example 5.4. Find the eigenvalues and the corresponding eigenvectors of

$$A = \begin{bmatrix} 3 & 2 & 2 \\ 1 & 4 & 1 \\ -2 & -4 & -1 \end{bmatrix}$$

.

Solution

The characteristic equation is

$$\begin{aligned}
0 = |A-\lambda I| &= \begin{vmatrix} 3-\lambda & 2 & 2 \\ 1 & 4-\lambda & 1 \\ -2 & -4 & -1-\lambda \end{vmatrix} \\
&= (3-\lambda)\begin{vmatrix} 4-\lambda & 1 \\ -4 & -1-\lambda \end{vmatrix} - 2\begin{vmatrix} 1 & 1 \\ -2 & -1-\lambda \end{vmatrix} + 2\begin{vmatrix} 1 & 4-\lambda \\ -2 & -4 \end{vmatrix} \\
&= (3-\lambda)\{(4-\lambda)(-1-\lambda)+4\} - 2\{(-1-\lambda)+2\} + 2\{-4+2(4-\lambda)\} \\
&= -\lambda^3 + 6\lambda^2 - 11\lambda + 6 \\
&= (\lambda-1)(-\lambda^2+5\lambda-6) \\
&= (\lambda-1)\{-(\lambda^2-5\lambda+6)\} \\
&= (\lambda-1)\{-(\lambda-2)(\lambda-3)\}
\end{aligned}$$

Hence, the eigenvalues are $\lambda_1 = 1, \lambda_2 = 2$ and $\lambda_3 = 3$.
For $\lambda_1 = 1$, we compute

$$\left[A - I|\mathbf{0}\right] = \left[\begin{array}{ccc|c} 2 & 2 & 2 & 0 \\ 1 & 3 & 1 & 0 \\ -2 & -4 & -2 & 0 \end{array}\right] \xrightarrow[R_3+R_1]{R_2-\frac{1}{2}R_1} \left[\begin{array}{ccc|c} 2 & 2 & 2 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & -2 & 0 & 0 \end{array}\right]$$

$$\xrightarrow{R_3+R_2} \left[\begin{array}{ccc|c} 2 & 2 & 2 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

from which it follows that an eigenvector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

satisfies $x_2 = 0$ and $x_3 = -x_1$. We take $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$ to be our eigenvector .
For $\lambda_2 = 2$, we compute

$$\left[A - 2I|\mathbf{0}\right] = \left[\begin{array}{ccc|c} 1 & 2 & 2 & 0 \\ 1 & 2 & 1 & 0 \\ -2 & -4 & -3 & 0 \end{array}\right] \xrightarrow[R_3+2R_1]{R_2-R_1} \left[\begin{array}{ccc|c} 1 & 2 & 2 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 3 & 0 \end{array}\right]$$

$$\xrightarrow{R_3+3R_2} \left[\begin{array}{ccc|c} 1 & 2 & 2 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

which gives an eigenvector

$$\mathbf{x} = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$$

.

For $\lambda_3 = 3$, we compute

$$\left[A - 3I \,|\, \mathbf{0}\right] = \left[\begin{array}{ccc|c} 0 & 2 & 2 & 0 \\ 1 & 1 & 1 & 0 \\ -2 & -4 & -4 & 0 \end{array}\right] \xrightarrow{R_2 \leftrightarrow R_1} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 2 & 2 & 0 \\ -2 & -4 & -4 & 0 \end{array}\right]$$

$$\xrightarrow{R_3+2R_1} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 2 & 2 & 0 \\ 0 & -2 & -2 & 0 \end{array}\right] \xrightarrow{R_3+R_2} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

which gives an eigenvector

$$\mathbf{x} = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

.

Example 5.5. Find the eigenvalues and the corresponding eigenspaces of

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -5 & 4 \end{bmatrix}$$

Solution

The characteristic equations is

$$\begin{aligned} 0 = |A - \lambda I| &= \begin{vmatrix} -\lambda & 1 & 0 \\ 0 & -\lambda & 1 \\ 2 & -5 & 4-\lambda \end{vmatrix} \\ &= -\lambda \begin{vmatrix} -\lambda & 1 \\ -5 & 4-\lambda \end{vmatrix} - \begin{vmatrix} 0 & 1 \\ 2 & 4-\lambda \end{vmatrix} \\ &= -\lambda(\lambda^2 - 4\lambda + 5) - (-2) \\ &= -\lambda^3 + 4\lambda^2 - 5\lambda + 2 \\ &= (\lambda - 1)(-\lambda^2 + 3\lambda - 2) \\ &= -(\lambda - 1)^2(\lambda - 2) \end{aligned}$$

Hence, the eigenvalues are $\lambda_1 = \lambda_2 = 1$ and $\lambda_3 = 2$.
To find the eigenvectors corresponding to $\lambda_1 = \lambda_2 = 1$, we compute

$$\left[A - I \,|\, \mathbf{0}\right] = \left[\begin{array}{ccc|c} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 2 & -5 & 3 & 0 \end{array}\right] \xrightarrow{R_3+2R_1} \left[\begin{array}{ccc|c} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -3 & 3 & 0 \end{array}\right] \xrightarrow{R_3-3R_2} \left[\begin{array}{ccc|c} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Thus,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

is in the eigenspace E_1 if and only if $-x_1 + x_2 = 0$ and $-x_2 + x_3 = 0$. Setting the free variable $x_3 = t$, we see that $x_1 = t$ and $x_2 = t$, from which it follows that

$$E_1 = \left\{ \begin{bmatrix} t \\ t \\ t \end{bmatrix} \right\} = \left\{ t \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\} = \operatorname{span}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right)$$

To find the eigenvectors correspond to $\lambda_3 = 2$, we find the nullspace of $A - 2I$ by row reduction:

$$\left[A - 2I \,|\, \mathbf{0}\right] = \left[\begin{array}{ccc|c} -2 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ 2 & -5 & 2 & 0 \end{array}\right] \xrightarrow{R_3 + R_1} \left[\begin{array}{ccc|c} -2 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & -4 & 2 & 0 \end{array}\right]$$

$$\xrightarrow{R_3 - 2R_2} \left[\begin{array}{ccc|c} -2 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

So

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

is in the eigenspace E_2 if and only if $-2x_1 + x_2 = 0$ and $-2x_2 + x_3 = 0$. Setting the free variable $x_3 = t$, we have

$$E_2 = \left\{ \begin{bmatrix} \frac{1}{4}t \\ \frac{1}{2}t \\ t \end{bmatrix} \right\} = \left\{ t \begin{bmatrix} \frac{1}{4} \\ \frac{1}{2} \\ 1 \end{bmatrix} \right\} = \operatorname{span}\left(\begin{bmatrix} \frac{1}{4} \\ \frac{1}{2} \\ 1 \end{bmatrix} \right) = \operatorname{span}\left(\begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} \right)$$

Definition 5.5: Algebraic and Geometric Multiplicity

The algebraic multiplicity of an eigenvalue is its multiplicity as a root of the characteristic equation. The geometric multiplicity of an eigenvalue λ is $\dim(E_\lambda)$, the dimension of its corresponding eigenspace.

In the above Example, $\lambda = 1$ has algebraic multiplicity 2 and geometric multiplicity 1. $\lambda = 2$ has algebraic multiplicity 1 and geometric multiplicity 1.

Corollary 5.6

The eigenvalues of a triangular matrix are the entries on its main diagonal.

Example 5.6. Let

$$A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 5 & 7 & 4 & -2 \end{bmatrix}$$

The characteristic polynomial is:

$$
\begin{aligned}
|A-\lambda I| &= \begin{vmatrix} 2-\lambda & 0 & 0 & 0 \\ -1 & 1-\lambda & 0 & 0 \\ 3 & 0 & 3-\lambda & 0 \\ 5 & 7 & 4 & -2-\lambda \end{vmatrix} \\
&= (2-\lambda)\begin{vmatrix} 1-\lambda & 0 & 0 \\ 0 & 3-\lambda & 0 \\ 7 & 4 & -2-\lambda \end{vmatrix} \\
&= (2-\lambda)(1-\lambda)\begin{vmatrix} 3-\lambda & 0 \\ 4 & -2-\lambda \end{vmatrix} \\
&= (2-\lambda)(1-\lambda)(3-\lambda)(-2-\lambda)
\end{aligned}
$$

Hence, the eigenvalues are $\lambda_1 = 2, \lambda_2 = 1, \lambda_3 = 3, \lambda_4 = -2$.

Note that diagonal matrices are a special case of Corollary 5.6.

Theorem 5.7

Let A be an $n \times n$ matrix and let $\lambda_1, \lambda_2, ..., \lambda_m$ be distinct eigenvalues of A with corresponding eigenvectors $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_m}$. Then $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_m}$ are linearly independent.

Proof

We prove this by contradiction.
Suppose $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_m}$ are linearly dependent. Let $\mathbf{v_{k+1}}$ be the first of the vectors $\mathbf{v_i}$ that can be expressed as a linear combination of the previous ones. In other words, $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_k}$ are linearly independent, but there are $c_1, c_2, ..., c_k$ such that

$$\mathbf{v_{k+1}} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + ... + c_k\mathbf{v_k} \qquad (1)$$

Multiplying both sides of Equation (1) by A from left and using the fact that $A\mathbf{v_i} = \lambda_i\mathbf{v_i}$ for each i, we have

$$
\begin{aligned}
\lambda_{k+1}\mathbf{v_{k+1}} = A\mathbf{v_{k+1}} &= A(c_1\mathbf{v_1} + c_2\mathbf{v_2} + ... + c_k\mathbf{v_k}) \\
&= c_1A\mathbf{v_1} + c_2A\mathbf{v_2} + ... + c_kA\mathbf{v_k} \\
&= c_1\lambda_1\mathbf{v_1} + c_2\lambda_2\mathbf{v_2} + ... + c_k\lambda_k\mathbf{v_k} \qquad (2)
\end{aligned}
$$

Now we multiply both sides of Equation (1) by λ_{k+1} to obtain

$$\lambda_{k+1}\mathbf{v_{k+1}} = c_1\lambda_{k+1}\mathbf{v_1} + c_2\lambda_{k+1}\mathbf{v_2} + ... + c_k\lambda_{k+1}\mathbf{v_k} \qquad (3)$$

When we subtract Equation (3) from Equation (2), we obtain

$$\mathbf{0} = c_1(\lambda_1 - \lambda_{k+1})\mathbf{v_1} + c_2(\lambda_2 - \lambda_{k+1})\mathbf{v_2} + ... + c_k(\lambda_k - \lambda_{k+1})\mathbf{v_k}$$

The linear independence of

$$\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_k}$$

implies that

$$c_1(\lambda_1 - \lambda_{k+1}) = c_2(\lambda_2 - \lambda_{k+1}) = ... = c_k(\lambda_k - \lambda_{k+1}) = 0$$

Since the eigenvalues λ_i are all distinct, $\lambda_i - \lambda_{k+1} \neq 0$ for all $i = 1, ..., k$. Hence $c_1 = c_2 = ... = c_k = 0$. This implies that

$$\mathbf{v_{k+1}} = 0\mathbf{v_1} + 0\mathbf{v_2} + ... + 0\mathbf{v_k} = \mathbf{0}$$

which is impossible since the eigenvector $\mathbf{v_{k+1}}$ cannot be zero.
Thus, our assumption that $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_m}$ are linearly dependent is false. It follows that $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_m}$ must be linearly independent. □

Example 5.7 (Coupled oscillators, normal modes)**.** Consider a system of two coupled oscillators connected by springs as shown in the figure.

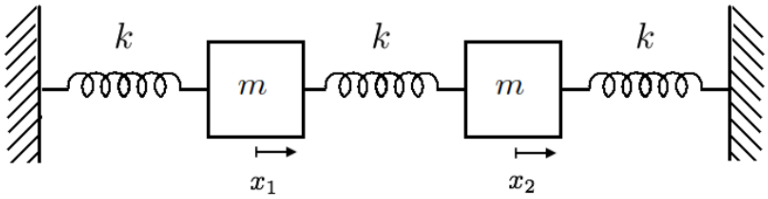


Figure 5.3: Coupled Oscillators

We can write down a system of equations for the dynamics of these two oscillators:

$$m\ddot{x_1} = -kx_1 + k(x_2 - x_1)$$

$$m\ddot{x_x} = -k(x_2 - x_1) - kx_2$$

We can write this as a vector equation for the vector

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$m\ddot{\vec{x}} = \begin{pmatrix} -2k & k \\ k & -2k \end{pmatrix} \vec{x}$$

If we are looking for the **normal modes** of the system, we are looking for harmonic oscillations with a single frequency. So we assume

$$\vec{x} = \begin{pmatrix} x_{10} \\ x_{20} \end{pmatrix} e^{i\omega t} \quad \ddot{\vec{x}} = -\omega^2 \begin{pmatrix} x_{10} \\ x_{20} \end{pmatrix} e^{i\omega t}$$

Substituting in the vector equation above leads to an eigenvalue problem

$$\Rightarrow -m\omega^2 \vec{x_0} = \begin{pmatrix} -2k & k \\ k & -2k \end{pmatrix} \vec{x_0}$$

$$\Rightarrow \begin{pmatrix} 2k/m & -k/m \\ -k/m & 2k/m \end{pmatrix} \vec{x_0} = \omega^2 \vec{x_0}$$

for the frequency ω. The eigenvalues and eigenvectors of the matrix are

$$\lambda_1 = 3k/m \quad \vec{x}_0 = \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

and

$$\lambda_1 = k/m \quad \vec{x}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

corresponding to an oscillation with frequency $\omega = \sqrt{3k/m}$ where the two oscillators are 180 degree out of phase, and an oscillation with frequency $\omega = \sqrt{k/m}$ where the oscillators are in phase.

Other solutions for the system can now be obtained as a linear combination of these two modes. The modes are a basis of the solution space.

5.2 Similarity and Diagonalisation

5.2.1 Introduction

In many applications, matrices represent linear mappings of vectors in a physical space, as in the example given at the start of the Eigenvector section. The choice of a coordinate system (in particular, its orientation) in this space is arbitrary, and this choice determines what the matrix looks like. In this section, we show that under certain conditions, there exists a choice of a coordinate system in which the matrix becomes a diagonal matrix. In this case, the coordinate axes have the direction of the eigenvectors of the matrix, and the diagonal elements of the matrix are the eigenvalues.

5.2.2 Similar Matrices

Definition 5.8: Similar Matrices

Let A and B be $n \times n$ matrices. We say that A is similar to B if there is an invertible $n \times n$ matrix P such that $P^{-1}AP = B$. If A is similar to B, we write $A \sim B$.

Remark

If $A \sim B$, we can write, equivalently, that $A = PBP^{-1}$ or $AP = PB$. The matrix P depends on A and B. It is not unique for a given pair of similar matrices A and B.

Theorem 5.9: Properties of Similar Matrices

Let A and B be $n \times n$ matrices with $A \sim B$. Then
(a) $\det(A) = \det(B)$
(b) A and B have the same rank.
(c) A and B have the same characteristic polynomial.
(d) A and B have the same eigenvalues.

Proof

If $A \sim B$, then $\mathrm{P}^{-1}\mathrm{AP} = \mathrm{B}$ for some invertible matrix P.
(a)

$$\begin{aligned} \det(B) &= \det(P^{-1}AP) \\ &= \det(\mathrm{P}^{-1})\det(\mathrm{A})\det(\mathrm{P}) \\ &= \frac{1}{\det(P)}\det(\mathrm{A})\det(\mathrm{P}) \\ &= \det(A). \end{aligned}$$

(b)
We first show that nullity(A)=nullity(B). One can then use the rank theorem 3.19 to show that this also implies that rank(A)=rank(B).
Let $\{\vec{x}_1, ..., \vec{x}_k\}$ be a basis of the nullspace of A. Hence $\vec{y}_i = P^{-1}\vec{x}_i$, $i = 1..k$, is a linearly independent set (P is one-to-one) in the nullspace of B, as

$$B\vec{y}_i = P^{-1}APP^{-1}\vec{x}_i = P^{-1}A\vec{x}_i = P^{-1}\vec{0} = \vec{0}.$$

To show that the $\vec{y}_i$ indeed are a basis of the nullspace of B, we observe that for any $\vec{y}$ in the nullspace of B, $P\vec{y} = \vec{x}$ is in the nullspace of A. So $\vec{y}_1, ...\vec{y}_k$ are a basis of the nullspace of B and nullity(A)=nullity(B).
(c) The characteristic polynomial of B is

$$\begin{aligned} \det(B - \lambda I) &= \det(P^{-1}AP - \lambda I) \\ &= \det(P^{-1}AP - \lambda P^{-1}IP) \\ &= \det(P^{-1}AP - P^{-1}(\lambda I)P) \\ &= \det(P^{-1}(A - \lambda I)P) \\ &= \det(P^{-1})\det(A - \lambda I)\det(P) \\ &= \frac{1}{\det(P)}\det(A - \lambda I)\det(P) \\ &= \det(A - \lambda I) \end{aligned}$$

□

Theorem 5.9 is helpful in showing that two matrices are not similar, since A and B cannot be similar if any of the properties fail.

Example 5.8.

(a) The two matrices

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},$$

are not similar since $\det(A) = -3$ but $\det(B) = 3$.

(b) The two matrices

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ 3 & -1 \end{bmatrix}$$

are not similar, since $|A - \lambda I| = \lambda^2 - 3\lambda - 4$ while $|B - \lambda I| = \lambda^2 - 4$. Note that A and B have the same determinant and rank, however.

5.2.3 Diagonalisation

Definition 5.10: Diagonalisable Matrices

An $n \times n$ matrix A is diagonalisable if there is a diagonal matrix D such that A is similar to D - that is, if there is an invertible $n \times n$ matrix P such that $P^{-1}AP = D$.

Theorem 5.11: Condition for Diagonalisability

Let A be an $n \times n$ matrix. Then A is diagonalisable if and only if A has n linearly independent eigenvectors.
More precisely, there exists an invertible matrix P and a diagonal matrix D such that $P^{-1}AP = D$ if and only if the columns of P are n linearly independent eigenvectors of A and the diagonal entries of D are the eigenvalues of A corresponding to the eigenvectors in P in the same order.*

Proof

Suppose first that A is similar to the diagonal matrix D by $P^{-1}AP = D$ or, equivalently, $AP = PD$. Let the columns of P be $\mathbf{p_1}, \mathbf{p_2}, ..., \mathbf{p_n}$ and let the diagonal entries of D be $\lambda_1, \lambda_2, ..., \lambda_n$. Then

$$A\left[\mathbf{p_1}, \mathbf{p_2}, \cdots, \mathbf{p_n}\right] = \left[\mathbf{p_1}, \mathbf{p_2}, \cdots, \mathbf{p_n}\right]\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \qquad (1)$$

or

$$\left[A\mathbf{p_1}, A\mathbf{p_2}, \cdots, A\mathbf{p_n}\right] = \left[\lambda_1\mathbf{p_1}, \lambda_2\mathbf{p_2}, \cdots, \lambda_n\mathbf{p_n}\right] \qquad (2)$$

Equating columns, we have

$$A\mathbf{p_1} = \lambda_1\mathbf{p_1}, A\mathbf{p_2} = \lambda_2\mathbf{p_2}, \cdots, A\mathbf{p_n} = \lambda_n\mathbf{p_n}$$

which proves that the column vectors of P are eigenvectors of A whose corresponding eigenvalues are the diagonal entries of D in the same order. Since P is invertible, its columns are linearly independent.
Conversely, if A has n linearly independent eigenvectors $\mathbf{p_1}, \mathbf{p_2}, \cdots, \mathbf{p_n}$ with corresponding eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$, respectively, then

$$A\mathbf{p_1} = \lambda_1\mathbf{p_1}, A\mathbf{p_2} = \lambda_2\mathbf{p_2}, \cdots, A\mathbf{p_n} = \lambda_n\mathbf{p_n}$$

This implies Eq. (2), which is equivalent to Eq. (1), that is $AP = PD$. Since the columns $\mathbf{p_1}, \mathbf{p_2}, \cdots, \mathbf{p_n}$ of P are linearly independent, P is invertible, so $P^{-1}AP = D$, that is, A is diagonalisable. □

Example 5.9. If possible, find a matrix P that diagonalises

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -5 & 4 \end{bmatrix}$$

Solution

We studied this matrix previously and found that it has eigenvalues $\lambda_1 = \lambda_2 = 1$ and $\lambda_3 = 2$. The eigenspaces have the following bases:

For $\lambda_1 = \lambda_2 = 1$, E_1 has basis $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

For $\lambda_3 = 2$, E_2 has basis $\begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$.

Since all other eigenvectors are just multiples of one of these two basis vectors, there cannot be three linearly independent eigenvectors. By Theorem 4.6, A is not diagonalisable.

Example 5.10. If possible, find a matrix P that diagonalises

$$A = \begin{bmatrix} 2 & 2 & 0 \\ 0 & 1 & 0 \\ -4 & -8 & 1 \end{bmatrix}$$

Solution

This is the matrix of Question 3, Worksheet 6. There we found that the eigenvalues of A are $\lambda_1 = 2$ and $\lambda_2 = \lambda_3 = 1$, with the following bases for the eigenspaces:

For $\lambda_1 = 2$, E_2 has basis $\mathbf{p_1} = \begin{bmatrix} 1 \\ 0 \\ -4 \end{bmatrix}$.

For $\lambda_2 = \lambda_3 = 1$, E_1 has basis $\mathbf{p_2} = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{p_3} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

Now we check whether $\{\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_3}\}$ is linearly independent.

$$\begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix} \xrightarrow{R_3+4R_1} \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & -8 & 1 \end{bmatrix} \xrightarrow{R_3+8R_2} \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Since $rank = 3$, $\{\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_3}\}$ is linearly independent. Thus, if we take

$$P = [\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_3}] = \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}$$

then P is invertible. Furthermore,

$$P^{-1}AP = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = D$$

(Note: It is much easier to check the equivalent equation $AP = PD$).

Remark

Eigenvectors can be placed into the columns of P in any order. However, the eigenvalues will come up on the diagonal of D in the same order as their corresponding eigenvectors in P. For example, if we had chosen

$$P = [\mathbf{p_2}, \mathbf{p_3}, \mathbf{p_1}] = \begin{bmatrix} -2 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & -4 \end{bmatrix}$$

Then we would have found

$$P^{-1}AP = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

We checked that the eigenvectors $\mathbf{p_1}, \mathbf{p_2}$ and $\mathbf{p_3}$ were linearly independent. However, the following Theorem guarantees that linear independence is preserved when the bases of different subspaces are combined.

Theorem 5.12

Let A be an $n \times n$ matrix and let $\lambda_1, \lambda_2, \cdots, \lambda_k$ be distinct eigenvalues of A. If B_i is a basis for the eigenspace E_i, then $B = B_1 \cup B_2 \cup \cdots \cup B_k$ (i.e. the total collection of basis vectors for all of the eigenspaces) is linearly independent.

Theorem 5.13

If A is an $n \times n$ matrix with n distinct eigenvalues, then A is diagonalisable.

Proof

Let $\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_n}$ be eigenvectors corresponding to the n distinct eigenvalues of A. By theorem 5.7, $\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_n}$ are linearly independent, so, by Theorem 5.11, A is diagonalisable. □

Example 5.11. The matrix

$$A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 5 & 7 & 4 & -2 \end{bmatrix}$$

has eigenvalues $\lambda_1 = 2, \lambda_2 = 1, \lambda_3 = 3$ and $\lambda_4 = -2$, by Corollary 5.6. Since these are four distinct eigenvalues for a 4×4 matrix, A is diagonalisable, by Theorem 5.13.

Corollary 5.14

If A is an $n \times n$ matrix, then the geometric multiplicity of each eigenvalue is less than or equal to its algebraic multiplicity.

Theorem 5.15: Diagonalisation Theorem

Let A be an $n \times n$ matrix whose distinct eigenvalues are $\lambda_1, \lambda_2, \cdots, \lambda_k$, where $1 \leq k \leq n$. The following statements are equivalent:

(a) A is diagonalisable.

(b) The union B of the bases of the eigenspaces of A contains n vectors.

(c) The algebraic multiplicity of each eigenvalue equals its geometric multiplicity.

Example 5.12.

(a) The matrix

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -5 & 4 \end{bmatrix}$$

has two distinct eigenvalues $\lambda_1 = \lambda_2 = 1$ and $\lambda_3 = 2$. Since the eigenvalue $\lambda_1 = \lambda_2 = 1$ with $E_1 = \text{span}((1,1,1)^T)$ has algebraic multiplicity 2 but geometric multiplicity 1, A is not diagonalisable, by the Diagonalisation Theorem.

(b) The matrix

$$A = \begin{bmatrix} 2 & 2 & 0 \\ 0 & 1 & 0 \\ -4 & -8 & 1 \end{bmatrix}$$

has two distinct eigenvalues $\lambda_1 = 2$ and $\lambda_2 = \lambda_3 = 1$. We found: for $\lambda_1 = 2$, E_1 has basis $\mathbf{p_1} = (1,0,-4)^T$, and for $\lambda_2 = \lambda_3 = 1$, E_2 has basis $\mathbf{p_2} = (-2,1,0)^T$ and $\mathbf{p_3} = (0,0,1)^T$. Thus, the eigenvalue 2 has algebraic and geometric multiplicity 1, and the eigenvalue 1 has algebraic and geometric multiplicity 2. Thus, A is diagonalisable, by the Diagonalisation Theorem.

Theorem 5.16: Joint eigenbasis

If two matrices commute and one of them is diagonalisable then they share a basis of eigenvectors.

Proof

Let A be diagonalisable. Then A has an eigenbasis $\vec{x}_i, i = 1..n$. For simplicity we first assume that there are n distinct eigenvalues, but one also prove the theorem for repeated eigenvalues.
Let B be a matrix that commutes with A:

$$\mathrm{AB} = \mathrm{BA} \quad \Rightarrow \mathrm{AB}\vec{x}_i = \mathrm{BA}\vec{x}_i \quad \Rightarrow \mathrm{A}(\mathrm{B}\vec{x}_i) = \lambda_i(\mathrm{B}\vec{x}_i)$$

That means that $\mathrm{B}\vec{x}_i$ is also an eigenvector of A to the eigenvalue λ_i and since the eigenspace to λ_i is 1-dimensional and spanned by $\vec{x}_i$ we have $\mathrm{B}\vec{x}_i = \mu_i\vec{x}_i$, for some real number μ_i. This means that $\vec{x}_i$ is also an eigenvector of B. A similar argumentation works for repeated eigenvalues.

i Remark

Joint eigenbases play an important role in quantum mechanics, where the eigenstates of commuting operators can be expressed w.r.t.~a common basis. It also means that measurements of these operator can be carried out simultaneously with (theoretically) infinite precision, while non-commuting operators have to satisfy a Heisenberg Uncertainty Principle.

5.3 Adjoint Operator

If our vector space is equipped with an inner product $\langle .,.\rangle$, and we are given a linear operator (matrix) A then there exists a further linear operator closely related to A, the Hermitian adjoint.

Definition 5.17: (Hermitian) Adjoint Operator

Given a linear operator A the hermitian adjoint of A is defined by

$$\langle \mathrm{A}\vec{x}, \vec{y}\rangle = \langle \vec{x}, \mathrm{A}^\dagger \vec{y}\rangle.$$

The operator A is called self-adjoint if $\mathrm{A} = \mathrm{A}^\dagger$.

For the simplest case of a real, finite-dimensional vector space with an inner product given by the standard scalar product the hermitian adjoint of A is just the transpose of A. This is because the scalar product $\vec{x} \cdot \vec{y}$ can be written also as a matrix product $\vec{x}^T \vec{y}$ and hence

$$\langle \mathrm{A}\vec{x}, \vec{y}\rangle = (\mathrm{A}\vec{x})^T \vec{y} = \vec{x}^T \mathrm{A}^T \vec{y} = \langle \vec{x}, \mathrm{A}^T \vec{y}\rangle.$$

Corollary 5.18: Properties of the hermitian adjoint

- $(\mathrm{A}^\dagger)^\dagger = \mathrm{A}$
- $\mathrm{A}^\dagger + \mathrm{B}^\dagger = (\mathrm{A} + \mathrm{B})^\dagger$
- $(\mathrm{AB})^\dagger = \mathrm{B}^\dagger \mathrm{A}^\dagger$

Corollary 5.19: orthogonal eigenbasis

For a self-adjoint operator eigenvectors to different eigenvalues are orthogonal.

Proof

$$\langle \mathrm{A}\vec{x}, \vec{y}\rangle = \langle \vec{x}, \mathrm{A}^\dagger \vec{y}\rangle = \langle \vec{x}, \mathrm{A}\vec{y}\rangle$$

$$\Rightarrow \langle \lambda_x \vec{x}, \vec{y}\rangle = \langle \vec{x}, \lambda_y \vec{y}\rangle \quad \Rightarrow \underbrace{(\lambda_x - \lambda_y)}_{\neq 0}\langle \vec{x}, \vec{y}\rangle = 0 \quad \Rightarrow \langle \vec{x}, \vec{y}\rangle = 0$$

6 Appendix

6.1 Greek Letters

Table 6.1: **Lower case Greek alphabet**

alpha	α	iota	ι	rho	ρ
beta	β	kappa	κ	sigma	σ
gamma	γ	lambda	λ	tau	τ
delta	δ	mu	μ	upsilon	υ
epsilon	ϵ	nu	ν	phi	ϕ
zeta	ζ	xi	ξ	chi	χ
eta	η	omicron	o	psi	ψ
theta	θ	pi	π	omega	ω

6.2 Carl Friedrich Gauss

Carl Friedrich Gauss (Gauß) (30 April 1777 - 23 February 1855) was a German mathematician and scientist of profound genius who contributed significantly to many fields, including number theory, analysis, differential geometry, geodesy, magnetism, astronomy and optics. Sometimes known as ”the prince of mathematicians” and ”greatest mathematician since antiquity”, Gauss had a remarkable influence in many fields of mathematics and science and is ranked as one of history’s most influential mathematicians.

Gauss was a child prodigy, of whom there are many anecdotes pertaining to his astounding precocity while a mere toddler, and made his first ground-breaking mathematical discoveries while still a teenager. He completed Disquisitiones Arithmeticae, his magnum opus, at the age of twenty-one (1798), though it would not be published until 1801. This work was fundamental in consolidating number theory as a discipline and has shaped the field to the present day. [1]

[1] From Wikipedia, the free encyclopaedia

Figure 6.1: CF Gauss