LAG 1 Revision Notes

Francesco Chotuck

Contents

1 Foundational material

1.1 Sums

1.1.1 Relabelling indices

You can shift indices provided you also shift the bounds to match.

$$\sum_{k=1}^{n}(k-1)=\sum_{k=0}^{n-1}k$$

1.1.2 Double Sums

$$\left(\sum_{k=0}^{n}a_k\right)\left(\sum_{k=0}^{m}b_k\right)=\sum_{j=0}^{n}\sum_{k=0}^{m}a_jb_k$$

1.1.3 Standard Sums

- $$\sum_{n=1}^{n}1=\underbrace{1+1+\ldots+1}_{\text{n times}}=n$$
- $$\sum_{n=1}^{n}k=1+2+3+\ldots+n=\frac{1}{2}n(n+1)$$
- $$\sum_{n=1}^{n}r^k=\frac{r-r^{n+1}}{1-r}$$

1.2 Complex Numbers

1.2.1 The Fundamental Theorem of Algebra

Let P be a polynomial of degree n, i.e.

$$P(z)=a_nz^n+a_{n-1}z^{n-1}+\ldots+a_1z+a_0,$$

where $a_0,\ldots,a_n\in\mathbb{C}$. Then the equation $P(z)=0$ has n solutions $w_1,w_2,\ldots,w_n$ (some of the solutions may be repeated). This means that $P(z)$ can be factorised as

$$P(z)=a_n(z-w_1)(z-w_2)\ldots(z-w_n).$$

Remark. *A field that satisfies this property is called* ***algebraically closed****. Thus,* $\mathbb{C}$ *is algebraically closed whereas* $\mathbb{R}$ *is not.*

1.2.2 Complex conjugate and modulus

Properties of complex conjugation: Let $z = a + ib$ and $w = c + id$, where $a, b, c, d \in \mathbb{R}$.

1. $\overline{z + w} = \overline{z} + \overline{w}$.
2. $\overline{zw} = \overline{z}\,\overline{w}$.
3. $\text{Re}(z) = \frac{1}{2}(z + \overline{z})$.
4. $\text{Im}(z) = \frac{1}{2i}(z - \overline{z})$.

Properties of the modulus:

1. $|z| = \sqrt{z\overline{z}}$.
2. $|z|$ is always a non-negative real number.
3. $|z| = |\overline{z}|$.
4. $|\text{Re}(z)| \leq |z|$ and $|\text{Im}(z)| \leq |z|$ since $a^2 \leq a^2 + b^2$ and $b^2 \leq a^2 + b^2$.
5. $|zw| = |z||w|$.

Proposition 1. (The triangle inequality). For every $z, w \in \mathbb{C}$,

$$|z + w| \leq |z| + |w|.$$

1.2.3 Euler's Formula

Theorem 1.1. (Euler's Formula) For every $\theta \in \mathbb{R}$,

$$e^{i\theta} = \cos\theta + i\sin\theta.$$

Therefore, a complex number $z = re^{i\theta}$ where $r = |z|$ and $\arg(z) = \theta$.

Remark. *The argument of z i.e. θ lies in the interval $-\pi < \theta \leq \pi$.*

1.2.4 De Moivre's theorem

Theorem 1.2. (De Moivre's theorem). For every $\theta \in \mathbb{R}$ and $n \in \mathbb{N}$,

$$(\cos\theta + i\sin\theta)^n = \cos(n\theta) + i\sin(n\theta).$$

1.2.5 Roots of Unity

Proposition 2. Let n be a positive integer. The equation $z^n = 1$ has exactly n distinct roots in the complex numbers and they are

$$z = e^{\frac{i2\pi k}{n}}, \quad k = 0, 1, 2, \ldots, n-1.$$

Remark. *Sketching the roots of unity on an Argand diagram form a n-polygon.*

Proposition 3. Let n be an integer with $n \geq 2$. The sum of the n-th roots of unity is equal to zero, i.e.

$$\sum_{k=0}^{n-1} e^{\frac{i2\pi k}{n}} = 0.$$

1.2.6 Logarithms and complex powers

Definition 1.1. Let $z \in \mathbb{C}, z \neq 0$. Then the **principal logarithm** of z is denoted $\text{Log } z$ and defined by

$$\text{Log } z = \log|z| + i\text{Arg } z.$$

2 Vectors

A vector is an object contained in a vector space, usually denoted as a column or a row of entries/components in $\mathbb{R}$ or $\mathbb{C}$.

2.1 Linear combinations and span

Definition 2.1. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ be vectors in $\mathbb{R}^n$. A **linear combination** of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ is an expression of the form

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \ldots + \alpha_m\mathbf{v}_m = \sum_{k=1}^{m} \alpha_m\mathbf{v}_m,$$

where $\alpha_1, \ldots, \alpha_n$ are scalars (in $\mathbb{R}^n$).

Definition 2.2. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ be vectors in $\mathbb{R}^n$. The **linear span** of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ is the set of all vectors which are linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$. We denote the linear span of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ by $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$, i.e.

$$\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\} = \left\{ \sum_{k=1}^{m} \alpha_k\mathbf{v}_k : \alpha_1, \ldots, \alpha_m \in \mathbb{R} \right\}$$

Some examples:

- Let $\mathbf{v} \in \mathbb{R}^n$ be a non-zero vector. Then

$$\text{span}\{\mathbf{v}\} = \{\alpha\mathbf{v} : \alpha \in \mathbb{R}^n\}.$$

  Geometrically, this is a line. In fact it is the line through the origin in the direction of the vector $\mathbf{v}$. Moreover, every line through the origin is of this form.

- Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ be a non-zero vectors. Then

$$\text{span}\{\mathbf{v}, \mathbf{w}\} = \{\alpha\mathbf{v} + \beta\mathbf{w} : \alpha, \beta \in \mathbb{R}^n\}.$$

  Geometrically, this is a plane containing $\mathbf{v}, \mathbf{w}$ and the origin. However, if $\mathbf{w}$ is a multiple of $\mathbf{v}$, then $\text{span}\{\mathbf{v}, \mathbf{w}\} = \text{span}\{\mathbf{v}\}$ and so we a get a line again and vice versa.

2.2 Lengths and Dot Product

Definition 2.3. Let $\mathbf{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$. The **length** or **norm** of $\mathbf{v}$ is denoted $||\mathbf{v}||$ and defined by

$$\sqrt{\sum_{k=1}^{n} v_k^2} = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}.$$

We can define the distance between two points $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ to be $||\mathbf{v} - \mathbf{u}||$.

Proposition 4. (Properties of the norm). For every $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and every scalar α we have

- $||\mathbf{v}|| \geq 0$ and $||\mathbf{v}|| = 0$ if and only if $\mathbf{v} = \mathbf{0}$,
- $||\alpha\mathbf{v}|| = |\alpha|\,||\mathbf{v}||$,
- $||\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||$.

Definition 2.4. Let $\mathbf{u} = (u_1, \ldots, u_n), \mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$. Then the **dot product** (or **scalar product** or **inner product**) of $\mathbf{u}$ and $\mathbf{v}$ is denoted $\mathbf{u}\cdot\mathbf{v}$ and defined by

$$\mathbf{u} \cdot \mathbf{v} = \sum_{k=1}^{n} u_k v_k = v_1 u_1 + u_2 v_2 + \ldots + u_n v_n.$$

Proposition 5. (Properties of the dot product). For every $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and every scalar α we have

- $||\mathbf{v}|| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$,
- $\mathbf{v} \cdot \mathbf{v} \geq 0$ and $\mathbf{v} \cdot \mathbf{v} = 0 \iff \mathbf{v} = 0$,
- $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$,
- $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$,
- $(\alpha \mathbf{u}) \cdot \mathbf{v} = \alpha(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (\alpha \mathbf{v})$.

Proposition 6. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ with $n = 2, 3$ and let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$. Then

$$\mathbf{u} \cdot \mathbf{v} = ||\mathbf{u}|| \, ||\mathbf{v}|| \cos \theta.$$

Remark.

1. *In particular, we have that two non-zero vectors* $\mathbf{u}$ *and* $\mathbf{v}$ *are perpendicular (or orthogonal for* $\mathbb{R}^3$*) if and only if* $\mathbf{u} \cdot \mathbf{v} = 0$.
2. *Since* $|\cos \theta| \leq 1$, *we have that*

$$|\mathbf{u} \cdot \mathbf{v}| \leq ||\mathbf{u}|| \, ||\mathbf{v}||.$$

This is known as the Cauchy-Schwartz inequality (check LAG 2).

2.2.1 Length and dot product with complex numbers

Definition 2.5. Let $\mathbf{z} = (z_1, z_2, \ldots, z_n) \in \mathbb{C}^n$. The **length** or **norm** of $\mathbf{z}$ is denoted $||\mathbf{z}||$ and defined by

$$\sqrt{\sum_{k=1}^{n} ||z_k||^2} = \sqrt{|z_1|^2 + |z_2|^2 + \ldots + |z_n|^2}.$$

We can define the distance between two point $\mathbf{z}$ and $\mathbf{w}$ in $\mathbb{C}^n$ to be $||\mathbf{z} - \mathbf{w}||$.

Definition 2.6. Let $\mathbf{z} = (z_1, \ldots, z_n), \mathbf{w} = (w_1, \ldots, w_n) \in \mathbb{C}^n$. Then the **dot product** (or **scalar product** or **inner product**) of $\mathbf{z}$ and $\mathbf{w}$ is denoted $\mathbf{z} \cdot \mathbf{w}$ and defined by

$$\mathbf{z} \cdot \mathbf{w} = \sum_{k=1}^{n} z_k \overline{w_k} = z_1 \overline{w_1} + z_2 \overline{w_2} + \ldots + z_n \overline{w_n}.$$

Remark. *If the entries of* $\mathbf{z}$ *and* $\mathbf{w}$ *are real, the definition of the dot product in* $\mathbb{C}^n$ *and* $\mathbb{R}^n$ *coincide. This is because in this case* $\overline{w_k} = w_k$ *and so*

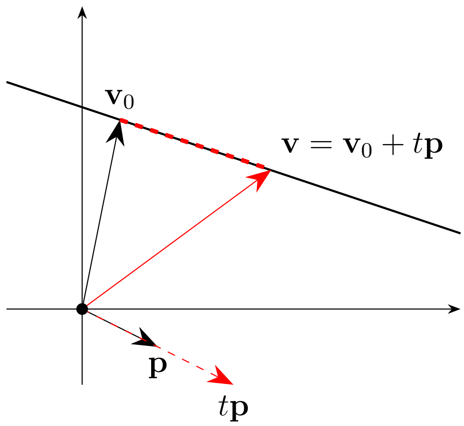
$$\sum_{k=1}^{n} z_k \overline{w_k} = \sum_{k=1}^{n} z_k w_k.$$

2.3 Lines and Planes

Let L be a line through the point $\mathbf{v}_0$ and parallel to $\mathbf{p}$. Then every point $\mathbf{v}$ on L is of the form

$$\mathbf{v} = \mathbf{v}_0 + t\mathbf{p}$$

for some $t \in \mathbb{R}$. As the real parameter t varies, we get every point on the line.



Any point $\mathbf{v}$ on the line can be expressed in the form $\mathbf{v} = \mathbf{v}_0 + t\mathbf{p}$ for some choice of $t \in \mathbb{R}$.

A plane can also be expressed parametrically as

$$\mathbf{v} = \mathbf{v}_0 + t\mathbf{p} + s\mathbf{q}$$

for some $t, s \in \mathbb{R}$, if and only if $\mathbf{p}$ is not a multiple of $\mathbf{q}$.

Remark. *The equation of a line and plane is not unique.*

Proposition 7. A plane in $\mathbb{R}^3$ is the set of points $\mathbf{v} = (v_1, v_2, v_3)$ satisfying an equation

$$av_1 + bv_2 + cv_3 = d,$$

where $a, b, c, d \in \mathbb{R}$ are constants and a, b, c not all zero.

Note. *The Cartesian equation of a plane is obtained by evaluating the dot product the vector equation of a plane* $\mathbf{v} = \mathbf{v}_0 + t\mathbf{p} + s\mathbf{q}$ *by a vector* $\mathbf{u}$ *which is perpendicular to the plane. This is because* $\mathbf{u} \cdot \mathbf{p} = \mathbf{u} \cdot \mathbf{q} = 0$ *therefore,* $\mathbf{u} \cdot \mathbf{v} = \mathbf{u} \cdot \mathbf{v_0}$. *Setting* $d = \mathbf{u} \cdot \mathbf{v_0}$ *we obtain* $av_1 + bv_2 + cv_3 = d$ *since,* $\mathbf{u} = (a, b, c)$.

3 Vector Spaces

Definition 3.1. A **vector space** is any collection of objects V (called **vectors**) for which two operations can be performed:

- **Vector addition**, which takes two vectors $\mathbf{v}, \mathbf{w} \in V$ and returns another vector $\mathbf{v} + \mathbf{w} \in V$ (V is **closed under addition**).
- **Scalar multiplication**, which takes a vector $\mathbf{v} \in V$ and a scaler $\alpha \in \mathbb{F}$ and returns a vector $\alpha\mathbf{v} \in V$ (V is **closed under scalar multiplication**).

Furthermore, the following properties (or **axioms**) must be satisfied.

1. Commutativity: $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$ for all $\mathbf{v}, \mathbf{w} \in V$;
2. Associativity: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w}$;
3. Zero vector: there exists a vector, denoted $\mathbf{0}$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$;
4. Additive inverse: For every vector $\mathbf{v} \in V$, there is a vector $-\mathbf{v} \in V$ such that $-\mathbf{v} + \mathbf{v} = \mathbf{0}$;
5. Multiplicative identity: $1\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in V$;
6. Multiplicative associativity: $\alpha(\beta\mathbf{v}) = (\alpha\beta)\mathbf{v}$ for all $\mathbf{v} \in V$ and all $\alpha, \beta \in \mathbb{F}$;
7. Distributivity: $\alpha(\mathbf{v} + \mathbf{w}) = \alpha\mathbf{v} + \alpha\mathbf{w}$ for all $\mathbf{v}, \mathbf{w} \in V$ and all $\alpha \in \mathbb{F}$;
   $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$ for all $\mathbf{v} \in V$ and all $\alpha, \beta \in \mathbb{F}$.

Examples of vector spaces

- Column Vectors;
- Zero vector space;
- Polynomials (usually denoted $\mathbb{P}_n$);
- Functions
- Solutions to differential equations (this applies only to linear homogenous ODEs.

3.1 Subspace

Definition 3.2. Let V be a vector space. A non-empty set $W \subseteq V$ is a **subspace** of V if

1. for every $\mathbf{u}, \mathbf{v} \in W, \mathbf{u} + \mathbf{v} \in W$ (W is closed under addition);
2. for every $\alpha \in \mathbb{F}$ and $\mathbf{v} \in W, \alpha\mathbf{v} \in W$ (W is closed under scalar multiplication).

Remark. *W is a subspace of V if W is a vector space and a subset of V.*

Proposition 8. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be vectors in V. Then $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a subspace of V. Moreover, if W is any subspace of V containing $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, then W must contain $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$

3.2 Bases

Definition 3.3. A collection of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in V$ is a **basis** for V if every vector $\mathbf{v} \in V$ admits a **unique** representation as a linear combination

$$\mathbf{v} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \ldots + \alpha_n\mathbf{v}_n = \sum_{k=1}^{n} \alpha_k\mathbf{v}_k.$$

The coefficients $\alpha_1, \alpha_2, \ldots, \alpha_n$ are called **coordinates** of $\mathbf{v}$ with respect to the basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.

Example of bases

- Let $V = \mathbb{F}^n$ the vectors $\mathbf{e}_1 = (1, 0, 0, \ldots, 0), \mathbf{e}_2 = (0, 1, 0, \ldots, 0), \ldots, \mathbf{e}_n = (0, 0, 0, \ldots, 1)$ form a basis for $\mathbb{F}^n$.
- Let $V = \mathbb{P}_n$ the vectors $\mathbf{e}_0 = 1, \mathbf{e}_1 = t, \mathbf{e}_2 = t^2, \ldots, \mathbf{e}_n = t^n$ form a basis for $\mathbb{F}^n$.

3.3 Coordinate vector

Definition 3.4. Let V be a vector space with basis $\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ and let $\mathbf{v} = \alpha_1\mathbf{e}_1 + \alpha_2\mathbf{e}_2 + \ldots + \alpha_n\mathbf{e}_n \in V$. Then the **coordinate vector** of $\mathbf{v}$ with respect to $\mathcal{E}$ is the column vector

$$[\mathbf{v}]_{\mathcal{E}} := \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} \in \mathbb{F}^n.$$

Thus we have a one-to-one correspondence (isomorphism) between V and $\mathbb{F}^n$:

$$\mathbf{v} \longleftrightarrow [\mathbf{v}]_{\mathcal{E}}.$$

Remark. *Addition and Scalar multiplication;*

$$[\mathbf{v} + \mathbf{w}]_{\mathcal{E}} = [\mathbf{v}]_{\mathcal{E}} + [\mathbf{v}]_{\mathcal{E}} \quad \textit{and} \quad [\beta \mathbf{v}]_{\mathcal{E}} = \beta [\mathbf{v}]_{\mathcal{E}}.$$

3.4 Spanning and linearly independent sets

Definition 3.5. We say that a collection of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in V$ **spans** V, or that is a **spanning** set (or **generating** set), if every vector $\mathbf{v} \in V$ admits a representation as a linear combination

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \ldots + \alpha_n \mathbf{v}_n = \sum_{k=1}^{n} \alpha_k \mathbf{v}_k.$$

Equivalently, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \,\mathrm{span}\, V$ if

$$\mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} = V.$$

Remark. *One can always add additional vectors to a spanning set and still get a spanning set. However, removing a vector from a spanning set may cause it to stop spanning.*

Definition 3.6. A collection of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in V$ are **linearly dependent** if there exists $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}^n$, not all zero, such that

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \ldots + \alpha_n \mathbf{v}_n = 0.$$

The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in V$ are **linearly independent** if they are not linearly dependent.

Remark. *A collection of vectors is said to be linearly independent if there is precisely one representation of* $\mathbf{0}$ *as a linear combination i.e. it is unique implying* $\alpha_1 = \alpha_2 = \ldots = \alpha_n = 0$.

Proposition 9. The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in V$ are linearly dependent if and only if one of the vectors can be represented as a linear combination of the others.

Remark. *Suppose* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ *are linearly dependent, then by Proposition 3.2 one of the vectors, say* $\mathbf{v}_1$*, is a linear combination of the others. Then any of the vectors can be written as a linear combination of* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ *can also be written as a linear combination of* $\mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$ *i.e.*

$$\mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} = \mathrm{span}\{\mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n\}.$$

Therefore, if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly dependent, then one of the $\mathbf{v}_k$ is a linear combination of the others and we can remove this vector without changing their linear span.

Proposition 10. A collection of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in V$ is a basis if and only if it is linearly independent and spans V.

Proposition 11. Any finite spanning set contains a basis.

3.5 Dimension

Definition 3.7. The dimension of a vector space V is denoted $\dim V$ and defined to be the number of elements in a basis for V. If V consists of only the zero vector we set $\dim V = 0$ and if V does not have a finite basis we set $\dim V = \infty$.

Theorem 3.1. (Dimension theorem). Let V be a vector space. Then every basis of V has the same number of elements. Moreover, if V has a basis of n elements then

- any set of vectors in V with less than n elements doesn't span V;
- any set of vectors in V with more than n elements is linearly dependent.

Proposition 12. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m \in V$ be linearly independent vectors in a finite dimensional vector space V. If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ do not span V, there exist vectors $\mathbf{v}_{m+1}, \mathbf{v}_{m+2}, \ldots, \mathbf{v}_n$ such that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is a basis of V.

Proposition 13. A collection of n vectors in an n dimensional vector space is spanning if and only if it is linearly independent.

Remark. *If we want to check if a collection of n vectors in $\mathbb{F}^n$ is a basis, we only need to check that it is linearly independent (or that it is spanning).*

4 Linear Maps

Definition 4.1. Let V and W be vector spaces over $\mathbb{F}$ (either both real or both complex). Then a map (or transformation) $T : V \to W$ is linear if it satisfies the following two conditions:

- for every $\mathbf{u}, \mathbf{v} \in V, T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$;
- for every $\mathbf{v} \in V$ and every $\alpha \in \mathbb{F}, T(\alpha\mathbf{v}) = \alpha T(\mathbf{v})$.

Remark. *Observe that we have two different addition rules in the above definition:*

$$T(\underbrace{\mathbf{u} + \mathbf{v}}_{+ \text{ in } V}) = \underbrace{T(\mathbf{u}) + T(\mathbf{v})}_{+ \text{ in } W}$$

Examples of linear maps

- Identity map;
- Zero map
- Reflection
- Rotation
- Differentiation

4.1 Operations on linear maps

Addition

Let V and W be vectors spaces (over $\mathbb{F}$) and let $S : V \to W$ and $T : V \to W$ be linear maps (note that they have the same domain and range). We can add S and T by the rule

$$(S+T)(\mathbf{v}) = S(\mathbf{v}) + T(\mathbf{v}) \quad \text{for all} \quad \mathbf{v} \in V.$$

Remark. *The map $(S+V)$ is a linear map.*

Scalar multiplication

Multiplying a linear map $T : V \to W$ by a scalar $\alpha \in \mathbb{F}$ according to the rule

$$(\alpha T)(\mathbf{v}) = \alpha T(\mathbf{v}).$$

Remark. *The map αT is a linear map.*

Composition

Let U, V and W be vector spaces and let $S : U \to V$ and $T : V \to W$ be linear maps (note that the range of S is the domain of T). We can consider the map we obtain by first applying S and then applying T, i.e. the map $T \circ S : U \to W$ defined by

$$(T \circ S)(\mathbf{u}) = T(S(\mathbf{u})) \quad \text{for all} \quad \mathbf{u} \in U$$

Remark. *The map $T \circ S$ is a linear map.*

5 Matrices

5.1 Matrix of a linear map between general vector spaces

Let V and W be vector spaces with $\dim V = n$ and $\dim W = m$. Suppose that we fix bases $\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n\}$ and $\mathcal{F} = \{\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_m\}$ of V and W respectively. Then we can get a matrix for a linear map $T : V \to W$ with respect to the bases $\mathcal{E}$ and $\mathcal{F}$.

Recall that if $\mathbf{v} = \alpha_1\mathbf{e}_1 + \alpha_2\mathbf{e}_2 + \ldots + \alpha_n\mathbf{e}_n$, the coordinate vector of $\mathbf{v}$ with respect to $\mathcal{E}$ is the column vector

$$[\mathbf{v}] = [\mathbf{v}]_{\mathcal{E}} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} \in \mathbb{F}^n.$$

Since $\mathcal{F}$ is a basis of W, we can write each of the vectors $T(\mathbf{e}_1), T(\mathbf{e}_2), \ldots, T(\mathbf{e}_n)$ as a linear combination of $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m$. Let

$$\begin{aligned} T(\mathbf{e}_1) &= a_{11}\mathbf{f}_1 + a_{21}\mathbf{f}_2 \ldots + a_{m1}\mathbf{f}_m \\ T(\mathbf{e}_2) &= a_{12}\mathbf{f}_1 + a_{22}\mathbf{f}_2 \ldots + a_{m1}\mathbf{f}_m \\ \vdots \quad & \quad \vdots \qquad\qquad\qquad \vdots \\ T(\mathbf{e}_n) &= a_{1n}\mathbf{f}_1 + a_{2n}\mathbf{f}_2 \ldots + a_{mn}\mathbf{f}_m. \end{aligned}$$

We can out the coefficients a_{jk} into a matrix A, so that the k^{th} column of A is the coordinate vector of $T(\mathbf{e}_k)$ with respect to the basis $\mathcal{F}$:

$$A = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ [T(\mathbf{e}_1)]_{\mathcal{F}} & [T(\mathbf{e}_2)]_{\mathcal{F}} & \ldots & [T(\mathbf{e}_1)]_{\mathcal{F}} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \in M_{m,n}(\mathbb{F}).$$

Then with $\mathbf{v} = \alpha_1\mathbf{e}_1 + \alpha_2\mathbf{e}_2 + \ldots + \alpha_n\mathbf{e}_n$, we have that

$$[T(\mathbf{v})]_{\mathcal{F}} = \alpha_1[T(\mathbf{e}_1)]_{\mathcal{F}} + \alpha_2[T(\mathbf{e}_2)]_{\mathcal{F}} + \ldots + \alpha_n[T(\mathbf{e}_n)]_{\mathcal{F}}$$

and thus

$$[T(\mathbf{v})]_{\mathcal{F}} = A[\mathbf{v}]_{\mathcal{E}}.$$

We say that A is the matrix of T with respect to the bases $\mathcal{E}$ and $\mathcal{F}$.

To summarise:

To get the matrix for a linear map T with respect to the bases $\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}2, ..., \mathbf{e}_n\}$ and $\mathcal{F} = \{\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_m\}$, we have to

1. express each of $T(\mathbf{e}_1), T(\mathbf{e}_2), ..., T(\mathbf{e}_n)$ as a linear combination of $\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_m$;

2. write the coefficients in a matrix so that the k^{th} column consists of the coefficients of $T(\mathbf{e}_k)$.

Remark. *The notation* $[T]_{\mathcal{F}}^{\mathcal{E}}$ *denotes the matrix of* T *from the basis* $\mathcal{E}$ *to* $\mathcal{F}$.

5.1.1 Composition

Let $S : \mathbb{F}^m \to \mathbb{F}^k$ and $T : \mathbb{F}^n \to \mathbb{F}^m$ be linear maps. Then the composition $(S \circ T) : \mathbb{F}^n \to \mathbb{F}^k$ is well-defined (since the range of T is the domain of S) and linear. Let A be the $k \times m$ matrix for S and B be the $m \times n$ matrix for T. Observe that for $\mathbf{v} \in \mathbb{F}^n$

$$(S \circ T)(\mathbf{v}) = S(T(\mathbf{v})) = S(B\mathbf{v}) = A(B\mathbf{v}).$$

5.2 Invertible linear maps

Let V be a vector space. Recall that the identity map $I : V \to V$ is given by $I(\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v} \in V$ (we will write I_V when we want to emphasize the space on which it acts). Then for linear maps $A : U \to V$ and $B : V \to U$ we have

$$IA = A \quad \text{and} \quad BI = B.$$

Definition 5.1. Let$A : V \to W$ be a linear map. Then A is **invertible** or **non-singular** if there exist a linear map $A^{-1} : W \to V$ such that

$$AA^{-1} = I_W \quad \text{and} \quad A^{-1}A = I_V.$$

In this case, we call A^{-1} the **inverse** of A.

Proposition 14. (Properties of the inverse). Let A and B be linear maps (matrices) such that the product AB is defined. Then the following properties hold:

- if A invertible then A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$;
- if A and B are invertible then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.

Theorem 5.1. Let $A : V \to W$ be an invertible linear map and let $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ be a basis of V. Then $A\mathbf{e}_1, A\mathbf{e}_2, \ldots, A\mathbf{e}_n$ is a basis of W.

6 System of Linear Equations

Recall that a *linear equation* in the unknowns $x_1, x_2, \ldots, x_n$ is an equation of the form

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b,$$

and a *system* of linear equations (or sometimes just *linear system*) is a finite collection of linear equations.

6.1 Geometry of linear equations

In this section, all the equations are real and so we consider only real solutions (the discussion below extends to $\mathbb{C}^n$ but it is difficult to visualise and so we avoid it here).

Consider two linear equations in two unknowns:

$$\begin{aligned} a_1x + b_1y &= c_1 \\ a_2x + b_2y &= c_2 \end{aligned}$$

Suppose that the coefficients a_k and b_k on each row are not both zero. Then the set of solutions of each row is a line in $\mathbb{R}^2$. The set of solutions of this linear system is the intersection of these two lines. There are three possible arrangements of three lines in $\mathbb{R}^2$:

1. **Lines intersect at a point.** The point of intersection is the unique solution.

2. **Lines are parallel.** Since the lines do not intersect, the system has no solutions.

3. **Lines are equal.** Every point on the line is a solution and so there are infinitely many solutions (sometimes called a 1-*parameter family* of solutions).

Next, let us consider a system with two equations in three unknowns:

$$\begin{aligned} a_1x + b_1y + c_1z &= d_1 \\ a_2x + b_2y + c_2z &= d_2. \end{aligned}$$

In this case, assuming the coefficients on each row are not all zero, each equation determines a plane in $\mathbb{R}^3$. As before, we can determine the number of possible solutions by considering the possible arrangements of two planes in $\mathbb{R}^3$. Now we can have the following arrangements:

1. **Planes intersect at a line.** There are infinitely many solutions (1-parameter family of solutions).

2. **Planes are parallel.** The system has no solutions.
3. **Planes are equal.** There are infinitely many solutions (a 2-*parameter* family of solutions).

In particular, this shows that two equations in three unknowns can never have a unique solution. With three equations however, we can have a unique solution, but there is also more variety of things that could go wrong. To be precise, the possible arrangements of three planes in $\mathbb{R}^3$ are:

1. **Planes intersect at a point.** There is a unique solution.
2. **Two parallel planes.** The system has no solutions. (This includes the case where exactly two planes are equal.)
3. **No parallel planes but no intersection.** The intersection of any two planes will be a line but the intersection of all three will be empty so there are no solutions. (This is the most common case when we don't have a unique solution.)
4. **All three planes intersect at a line.** There are infinitely many solutions (1-parameter family).
5. **All three planes are parallel.** The system has no solutions.
6. **All three planes are equal.** There are infinitely many solutions (2-parameter family).

6.2 Elimination

6.2.1 The idea of elimination

The idea of elimination, to perform operations on the rows of the augmented matrix (or equivalently on the equations themselves) until we reduce it to a simple form where one can write down the solutions.

In general, we simplify our system by performing elementary row operations of the following types:

I. interchange two rows of the matrix;

II. multiply a row by a non-zero scalar;

III. add a multiple of one row to a different row.

These operations do not change the set of solutions – that is, the solutions of the linear system obtained after applying an elementary row operation are the same as the solutions of the original system.

6.2.2 Echelon Form

Definition 6.1. A matrix is in **echelon** form if

1. all zero rows, if there are any, are below all non-zero rows;

2. for each non-zero row, its left-most non-zero entry is strictly to the right of the left-most non-zero entry of the row above.

The left-most non-zero entry in each row in echelon form is called a **pivot** entry or simply *pivot*.

A matrix is in **reduced echelon form** (or REF for short) if it is in echelon form and

1. all pivot entries are equal to 1;

2. all entries above the pivots are zero.

Remark. *If a column does not have a pivot the associated variable is said to be* ***free*** *hence, it can vary and acts as a parameter.*

6.3 Elementary matrices

Definition 6.2. A square matrix is **elementary** if it differs from the identity matrix by one elementary row operation.

Every elementary row operation is equivalent to multiplication on the left by an elementary matrix. Suppose we want to apply a row operation R to a matrix A. Let E be the matrix obtained by applying R to the identity matrix. Then the matrix obtained by applying R to A is EA. For example, if A has 4 rows we have the following:

1. Interchanging row 2 and row 4 is equivalent to left multiplication by

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

2. Multiplying row 3 by α is equivalent to left multiplication by

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

3. Adding α times row 1 to row 2 is equivalent to left multiplication by

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

Remark. *Since row operations are reversible, elementary matrices are invertible. To conclude we have the following*

> Given a matrix A, we can find an invertible matrix E such that EA is in echelon form (or reduced echelon form).

6.4 Analysing the pivots

Proposition 15. A linear system is inconsistent (i.e. has no solutions) if and only if the echelon form of the augmented matrix has a pivot in the last column.

Proposition 16. Let $A : \mathbb{F}^n \to \mathbb{F}^m$be a linear map (matrix).

- The linear system $A\mathbf{x} = \mathbf{b}$ is consistent for all right sides $\mathbf{b} \in \mathbb{F}^m$ if and only if the echelon form of the coefficient matrix has a pivot in every row.
- The linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution for every right side $\mathbf{b} \in \mathbb{F}^m$ if and only if the echelon form of the coefficient matrix has a pivot in every column and every row.

Corollary 6.0.1. *A matrix A is invertible if and only if its echelon form has a pivot in every column and every row.*

Remark. *Any row or column of a matrix in echelon form can have at most one pivot in it. Therefore if the echelon form of a matrix has a pivot in every row and every column, it must have the same number of rows and columns. In particular, this shows us that an invertible matrix must be square.*

Proposition 17. Let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ be vectors in $\mathbb{F}^n$. Let A be the $n \times n$ matrix with columns $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$, i.e.

$$A = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_m \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}.$$

Then the following hold:

- the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are linearly independent if and only if the echelon form of A has a pivot in every column;

- the vector $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ span $\mathbb{F}^n$ if and only if the echelon form of A has a pivot in every row;

- the vector $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ form a basis of $\mathbb{F}^n$ if and only if the echelon form of A has a pivot in every column and every row.

6.4.1 Consequences for square matrices

Proposition 18. Let A be a (square) $n \times n$ matrix. Then A has a left inverse if and only if it has a right inverse.

Proposition 19. Let A be a (square) $n \times n$ matrix. Then A is invertible if and only if the equation $A\mathbf{x} = \mathbf{0}$ has a unique solution.

6.5 Computing the inverse by elimination

There exists a matrix E such that

$$EA = I.$$

Recall that E is the product of the elementary matrices corresponding to the row operations required to put A into REF. Thus, E can be computed by applying the same row operations to the identity matrix. This gives us the following algorithm for computing $E = A^{-1}$ for an $n \times n$ matrix A:

1. Form an *augmented* $n \times 2n$ matrix $(A|I)$.

2. Perform row operations on the augmented matrix to transform A to the identity matrix.

3. The matrix I that was added will be transformed to A^{-1}.

6.6 Principle of linearity

Definition 6.3. A linear equation (linear system) $A\mathbf{x} = \mathbf{b}$ is **homogeneous** if $\mathbf{b} = \mathbf{0}$ – that is, a homogeneous linear equation is an equation of the form $A\mathbf{x} = \mathbf{0}$. Otherwise we say it is **inhomogeneous**.

Theorem 6.1. (Principle of linearity). Let $A : V \to W$ be a linear map and let $\mathbf{b} \in W$. Let $\mathbf{x}_0 \in V$ satisfy the equation $A\mathbf{x}_0 = \mathbf{b}$ and let H denote the set of solutions of the associated homogeneous equation $A\mathbf{x} = \mathbf{0}$. Then the set

$$\{\mathbf{x}_0 + \mathbf{x}_h : \mathbf{x}_h \in H\}$$

is the set of all solutions of the equation $A\mathbf{x} = \mathbf{b}$.

This can be restated as

General solution of $A\mathbf{x} = \mathbf{b}$	$=$	A particular solution of $A\mathbf{x} = \mathbf{b}$	$+$	General solution of $A\mathbf{x} = \mathbf{0}$

Proof. Given a linear system $A\mathbf{x} = \mathbf{b}$, we call the system

$$A\mathbf{x} = \mathbf{0}$$

the **associated homogenous** system.

Consider a linear system $A\mathbf{x} = \mathbf{b}$ and its associated homogeneous system $A\mathbf{x} = \mathbf{0}$. Suppose $\mathbf{x}_0$ is solution of the original system, i.e. $A\mathbf{x}_0 = \mathbf{b}$, and suppose that $\mathbf{x}_h$ is a solution of the homogeneous system, i.e. $A\mathbf{x}_h = \mathbf{0}$ Then

$$A(\mathbf{x}_0 + \mathbf{x}_h) = A\mathbf{x}_0 + A\mathbf{x}_h = \mathbf{b} + \mathbf{0} = \mathbf{b}.$$

Hence $\mathbf{x}_0 + \mathbf{x}_h$ is also a solution of the original system.

Now suppose that $\mathbf{x}_1$ is another solution of the original system. Then

$$A(\mathbf{x}_1 - \mathbf{x}_0) = A\mathbf{x}_1 + A\mathbf{x}_0 = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

Hence $\mathbf{x}_1 - \mathbf{x}_0$ is also a solution of the original system. Moreover, we have

$$\mathbf{x}_1 = \mathbf{x}_0 + (\mathbf{x}_1 - \mathbf{x}_0),$$

and thus every solution of the original system is the sum of $\mathbf{x}_0$ and a solution of the homogeneous system. $\square$

Remark. *Observe that Theorem 6.1 is valid for all linear maps. In particular, it is valid for linear maps between infinite dimensional spaces where one cannot simply use elimination to determine the general solution.*

6.7 Image and kernel of a linear map

Definition 6.4. Let $A : V \to W$ be a linear map. The **image** (or **column space**) of A, denoted $\operatorname{Im} A$, is the set of all vectors $\mathbf{w} \in W$ such that $\mathbf{w} = A\mathbf{v}$ for some $\mathbf{v} \in V$, i.e.

$$\operatorname{Im} A := \{A\mathbf{v} \in W : \mathbf{v} \in V\}.$$

Definition 6.5. Let $A : V \to W$ be a linear map. The **kernel** (or **null space**) of A, denoted $\operatorname{Ker} A$ is the set of all vectors $\mathbf{v} \in V$ such that $A\mathbf{v} = \mathbf{0}$, i.e.

$$\operatorname{Ker} A := \{\mathbf{v} \in V : A\mathbf{v} = \mathbf{0}\}.$$

Equivalently, $\operatorname{Im} A$ is the set of right sides $\mathbf{b}$ such that the equation $A\mathbf{x} = \mathbf{b}$ has solutions and $\operatorname{Ker} A$ is the set of solutions of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

Remark. *The four spaces* $\operatorname{Im} A, \operatorname{Ker} A, \operatorname{Im} A^T$ *and* $\operatorname{Ker} A^T$ *are called the* ***fundamental subspaces*** *of* A.

Proposition 20. Let $A : V \to W$ be a linear map. Then $\operatorname{Im} A$ is a subspace of W and $\operatorname{Ker} A$ is a subspace of V.

Remark. *Observe that a linear map* $A : V \to W$ *is surjective if and only if* $\operatorname{Im} A = W$.

Proposition 21. Let $A : V \to W$ be a linear map. A is injective if and only $\operatorname{Ker} A = \{\mathbf{0}\}$.

Proof. First, let us assume that $\operatorname{Ker} A = \{\mathbf{0}\}$. Take $\mathbf{v}_1, \mathbf{v}_2 \in V$ with $A\mathbf{v}_1 = A\mathbf{v}_2$. We need to show that $\mathbf{v}_1 = \mathbf{v}_2$. We have that $A(\mathbf{v}_1 - \mathbf{v}_2) = A\mathbf{v}_1 - A\mathbf{v}_2 = \mathbf{0}$, and so $\mathbf{v}_1 - \mathbf{v}_2 \in \operatorname{Ker} A$. Since $\operatorname{Ker} A = \{\mathbf{0}\}$ we have that $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{0}$, i.e $\mathbf{v}_1 = \mathbf{v}_2$.

Now let us assume that there exists a non-zero vector $\mathbf{v} \in \operatorname{Ker} A$. Then $A\mathbf{0} = \mathbf{0}$ and $A\mathbf{v} = \mathbf{0}$ with $\mathbf{v} \neq \mathbf{0}$. Hence A is not injective. □

Definition 6.6. Let $A : V \to W$ be a linear map.
The **rank** of A, denoted $r(A)$, is the dimension of the image of A, i.e. $r(A) := \dim \operatorname{Im} A$.
The **nullity** of A, denoted $n(A)$, is the dimension of the kernel of A, i.e. $n(A) = \dim \operatorname{Ker} A$.

Note:

1. The pivot columns of A, i.e. columns of A (not its echelon form) corresponding to pivot variables, form a basis for $\operatorname{Im} A$.

2. To compute a basis for the kernel of A, we have to solve the homogenous equation $A\mathbf{x} = \mathbf{0}$. The vectors that form the basis will be the coefficient vectors of the parameters.

3. The rank of A is the number of pivots in the reduced echelon form of A.

4. The nullity of A is the number free columns (i.e columns without pivots) in the echelon form of A.

Theorem 6.2. (Rank-nullity theorem). Let $A : V \to W$ be a linear map, where V is finite dimensional. Then

$$r(A) + n(A) = \dim V.$$

Corollary 6.2.1. *Let V and W be vector spaces of dimension n and let $A : V \to W$ be a linear map. Then the following properties of V are equivalent:*

(i) *A is surjective.*

(ii) $r(A) = n$.

(iii) $n(A) = 0$.

(iv) *A is injective.*

Proposition 22. For any matrix A we have that $r(A) = r(A^T)$.

Remark. *As a consequence of Proposition 6.8 we have that for $A : V \to W$*

$$r(A) + n(A^T) = \dim W.$$

6.8 Change of bases. Similarity

Recall for that a linear map $T : V \to W$ and bases $\mathcal{E}$ and $\mathcal{F}$ for V and W respectively, we have

$$[T\mathbf{v}]_{\mathcal{F}} = [T]^{\mathcal{E}}_{\mathcal{F}}[\mathbf{v}]_{\mathcal{E}}, \quad \mathbf{v} \in V. \tag{1}$$

Let $I : V \to V$ be the identity map given by $I\mathbf{v} = \mathbf{v}, \mathbf{v} \in V$. Then (1) we have that for each $\mathbf{v} \in V$,

$$[\mathbf{v}]_{\mathcal{F}} = [I]^{\mathcal{E}}_{\mathcal{F}}[\mathbf{v}]_{\mathcal{E}}.$$

> Multiplying by $[I]^{\mathcal{E}}_{\mathcal{F}}$ transforms the coordinates in the $\mathcal{E}$ basis to coordinates in the $\mathcal{F}$ basis.

The matrix $[I]^{\mathcal{E}}_{\mathcal{F}}$ is known as the **transition** matrix (or **change of coordinates** matrix) from $\mathcal{E}$ to $\mathcal{F}$. Observe that we necessarily have

$$[I]^{\mathcal{F}}_{\mathcal{E}} = ([I]^{\mathcal{E}}_{\mathcal{F}})^{-1}.$$

Let $\mathcal{G}$ be another basis of V. Then

$$[I]^{\mathcal{E}}_{\mathcal{G}} = [I]^{\mathcal{F}}_{\mathcal{G}}[I]^{\mathcal{E}}_{\mathcal{F}}.$$

Recall to obtain $[I]^{\mathcal{E}}_{\mathcal{F}}$ we must express $\mathbf{e}_k$ in terms of $\mathbf{f}_k$.

Definition 6.7. Let A and B be square matrices. Then A is similar to B if there exists an invertible matrix P such that

$$A = P^{-1}BP.$$

Note. *If A is similar to B then B is also similar to A.*

Remark. *Similarity is an equivalence relation on the set of $n \times n$ matrices.*

> Two $n \times n$ matrices are similar if and only if they represent the same linear map with respect to (possibly) different bases.

7 Determinants

Determinants of a 1×1 matrix

Let $A = (a)$ be a 1×1 matrix. Then A is invertible if and only if $a \neq 0$. We define

$$\det\{(A)\} := a.$$

Determinants of a 2×2 matrix

Let

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

If A is invertible then at least one of a or c must be non-zero (otherwise A would have a zero column). We suppose that $a \neq 0$. Then we can row reduce A as follows:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \longrightarrow \begin{pmatrix} a & b \\ 0 & d - \frac{cb}{a} \end{pmatrix}.$$

This is invertible if and only if it has a pivot in each row. Hence A is invertible if and only if

$$a\left(d - \frac{cb}{a}\right) = ad - bc \neq 0.$$

We define

$$\det\{(A)\} := ad - bc.$$

7.1 Geometric interpretation

The determinant must satisfy these properties:

1. **Multilinearity:** The determinant is linear as a function of each column separately.
2. **Antisymmetry:** The determinant is multiplied by -1 whenever we interchange two columns.
3. **Normalization:** The determinant of the identity matrix is 1.

- For $\mathbb{R}^2$ the determinant represents the signed area of the parallelogram determined by $A\mathbf{e}_1$ and $A\mathbf{e}_2$.
- Applying the transformation A scales areas by a factor of $|\det\{(A)\}|$.
- For $\mathbb{R}^3$ the determinant represents the volume of the parallelepiped determined by $A\mathbf{e}_1, A\mathbf{e}_2, A\mathbf{e}_3$.

Theorem 7.1. For each $n \in \mathbb{N}$, there exists a unique function $\det : M_n(\mathbb{C}) \to \mathbb{C}$ satisfying the basic properties above. In other words, the determinant is well-defined and uniquely determined by the basic properties.

Proposition 23. For a square matrix A, the following statements hold:

(i) If A has a zero column, the determinant of A is zero.

(ii) If the columns of A are linearly dependent, the determinant of A is zero.

(iii) Adding a multiple of one column to another leaves the determinant unchanged.

Corollary 7.1.1. *If A is not invertible, the determinant of A is zero.*

7.2 Determinants of diagonal and triangular matrices

Definition 7.1. A square matrix $A = (a_{jk})$ is diagonal if $a_{jk} = 0$ whenever $j \neq k$, that is, if A is of the form

$$A = \begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n \end{pmatrix}$$

Remark. *The entries a_{jk} of A with $j = k$ are called the main diagonal. So we say that a diagonal matrix is zero away from the main diagonal. The main diagonal is from the top-left to the bottom-right.*

$$\det\{(A)\} = \det \begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n \end{pmatrix} = a_1 a_2 \dots a_n \det\{(I)\} = a_1 a_2 \dots a_n.$$

Definition 7.2. A square matrix $A = (a_{jk})$ is called upper triangular if $a_{jk} = 0$ whenever $k < j$ and lower triangular if $a_{jk} = 0$ whenever $k > j$. A matrix is triangular if it is either upper or lower triangular.

For example,

$$\begin{pmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{pmatrix} \text{ is upper triangular,} \qquad \begin{pmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{pmatrix} \text{ is lower triangular.}$$

Note. *The determinant of a diagonal matrix is the product of its diagonal entries.*

Let E be an elementary matrix that corresponds to a particular row operation. Applying the same operation to columns of a matrix A (i.e. to the rows of A^T) produces the matrix

$$(EA^T)^T = AE^T.$$

Observe that E^T is also an elementary matrix of the same type as E. Hence every elementary column operation can be realised as multiplication on the *right* by an elementary matrix.

Lemma 7.2. *Let E be an elementary matrix. Then the following statements hold:*

- $\det\{(E)\} = \det\big\{(E^T)\big\}$.
- *For every square matrix* $A, \det\{(AE)\} = \det\{(A)\}\det\{(E)\}$.

Theorem 7.3. Let A be a square matrix. Then

$$\det\{(A)\} = \det\big\{(A^T)\big\}.$$

Corollary 7.3.1. *The following properties hold for the determinant:*

- *The determinant is linear in each row separately (i.e. it is a multilinear function of the rows).*
- *Elementary row operations have the same effect on the determinant as the corresponding column operations.*

Theorem 7.4. Let A and B be square matrices. Then

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

Corollary 7.4.1. *Let A be a square matrix. Then A is invertible if and only if* $\det(A) \neq 0$

7.3 Cofactor Formula

Definition 7.3. Let A be an $n \times n$ matrix. For $1 \leq i, j \leq n$, the **minor** $M_{ij}(A)$ is the determinant of the $(n-1) \times (n-1)$ matrix obtained by deleting the row i and column j from A.
The **cofactor** $C_{ij(A)}$ is $(-1)^{i+j}M_{ij}(A)$.

Definition 7.4. Let $A = (a_{ij})$ be an $n \times n$ matrix. Then the determinant of A, $\det(A)$, is defined inductively as follows:

- if $n = 1, \det\{(A)\} = a$;
- if $n \geq 2, \det\{(A)\} = \sum_{j=1}^{n} a_{1j}C_{1j}(A) = a_{11}C_{11}(A) + a_{12}C_{12}(A) + \ldots + a_{1n}C_{1n}(A)$.

Proposition 24. For each $n \in \mathbb{N}$, the determinant det : $M_n(\mathbb{C}) \to \mathbb{C}$ satisfies the basic properties given (linearity in each column, antisymmetry and normalization).

Proposition 25. Let A be an $n \times n$ matrix with $\det(A) \neq 0$. Let C be the $n \times n$ matrix whose i, j-entry is the cofactor $C_{ij}(A)$, i.e.

$$C = \begin{pmatrix} C_{11}(A) & C_{12}(A) & \dots & C_{1n}(A) \\ C_{21}(A) & C_{22}(A) & \dots & C_{2n}(A) \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1}(A) & C_{n2}(A) & \dots & C_{nn}(A) \end{pmatrix}.$$

Then

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

8 Introduction to eigenvalues and canonical forms

8.1 Eigenvalues and eigenvectors

Definition 8.1. Let $A : \mathbb{F}^n \to \mathbb{F}^n$ be a linear map. Then $\lambda \in \mathbb{C}$ is an **eigenvalue** of A if there exists a **non-zero** vector $\mathbf{x} \in \mathbb{F}^n$ such that

$$A\mathbf{x} = \lambda \mathbf{x}. \tag{2}$$

A non-zero vector $\mathbf{x}$ is an **eigenvector** of A corresponding to the eigenvalue λ if it satisfies (2).

Observe that (2) is satisfied if and only if $(A - \lambda I)\mathbf{x} = \mathbf{0}$, so that $\mathbf{x} \in \operatorname{Ker}(A - \lambda I)$. Thus λ is an eigenvalue of A if and only $\operatorname{Ker}(A - \lambda I) \neq \{\mathbf{0}\}$ and the eigenvalues corresponding to λ are the non-zero elements of $\operatorname{Ker}(A - \lambda I)$. Recall that if the kernel of an $n \times n$ matrix contains a non-zero vector, it is not invertible and so its determinant is zero. This gives us the following characterisation of eigenvalues:

$\lambda \in \mathbb{C}$ is an eigenvalue of A if and only if $\det\{(A - \lambda I)\} = 0$.

Remark. *The polynomial* $p(\lambda) = \det\{(A - \lambda I)\}$ *is called the* ***characteristic polynomial*** *of* A.

Remark. *An eigenvalue is by definition a non-zero number consequently, an eigenvector is non-zero therefore, the* $\operatorname{Ker}(A - \lambda I) \neq \{\mathbf{0}\}$. *So, the matrix* $A - \lambda I$ *is not injective hence, not invertible so it must have* $\det\{(A - \lambda I) = 0\}$.

8.2 Canonical forms of 2×2 matrices

We say that a matrix is **diagonalisable** if it is similar to a diagonal matrix. More precisely, we say that it is diagonalisable over $\mathbb{F}$ if it is similar to a diagonal matrix with entries in $\mathbb{F}$.

Note. *It is possible that a real matrix is diagonalisable over $\mathbb{C}$ but not over $\mathbb{R}$.*

Theorem 8.1. Let A be a 2×2 matrix. Then exactly one of the following hold:

(i) A is diagonalisable over $\mathbb{C}$ so that there exists an invertible matrix P and $\lambda_1, \lambda_2 \in \mathbb{C}$ such that
$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

(ii) A is not diagonalisable over $\mathbb{C}$ but there exists an invertible matrix P and $\lambda_0 \in \mathbb{C}$ such that
$$P^{-1}AP = \begin{pmatrix} \lambda_0 & 0 \\ 0 & \lambda_0 \end{pmatrix}.$$

To determine which class a particular matrix A falls into, we need to determine its eigenvalues.

Theorem 8.2. Let A be a real 2×2 matrix. Then exactly one of the following hold:

(i) A is diagonalisable over $\mathbb{R}$ so that there exists an invertible matrix P and $\lambda_1, \lambda_2 \in \mathbb{R}$ such that
$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

(ii) A is not diagonalisable over $\mathbb{R}$ but there exists an invertible matrix P and $\lambda \in \mathbb{R}$ such that
$$P^{-1}AP = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}.$$

(iii) A is not diagonalisable over $\mathbb{R}$ but there exists an invertible matrix P and $\alpha, \beta \in \mathbb{R}, \beta \neq 0$, such that
$$P^{-1}AP = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}.$$

As before, we determine which of these cases applies to a particular matrix A by considering its eigenvalues.

8.3 Application to ODEs

For simplicity, we will only consider first order homogeneous constant coefficient equations.

If A is diagonalisable and $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are linearly independent eigenvectors corresponding to $\lambda_1, \ldots, \lambda_n$ respectively, then the general solution of

$$\frac{d\mathbf{u}}{dt} = A\mathbf{u},$$

is

$$\mathbf{u}(t) = C_1 e^{\lambda_1 t}\mathbf{x}_1 + C_2 e^{\lambda_2 t}\mathbf{x}_2 + \ldots + C_n e^{\lambda_n t}\mathbf{x}_n, \quad C_1, \ldots, C_n \in \mathbb{F}.$$

Remark. *If A is not diagonalisable, there are other solutions. In particular, if λ is an eigenvalue which is repeated k times but has only one linearly independent eigenvector $\mathbf{x}$, then $e^{\lambda t}\mathbf{x}, te^{\lambda t}\mathbf{x}, \ldots, t^{k-1}e^{\lambda t}\mathbf{x}$ will be linearly independent solutions.*