Vectors and Matrices

Cambridge University Mathematical Tripos: Part IA

17th May 2024

Contents

1 Complex numbers

1.1 Definition and basic theorems

We construct the complex numbers from $\mathbb{R}$ by adding an element i such that $i^2 = -1$. By definition, any complex number $z \in \mathbb{C} = x + iy$ where $x, y \in \mathbb{R}$. We use the notation $x = \operatorname{Re} z$ and $y = \operatorname{Im} z$ to query the components of a complex number. The complex numbers contains the set of real numbers, due to the fact that $x = x + i0$. We define the operations of addition and multiplication in familiar ways, which lets us state that $\mathbb{C}$ is a field.

We also define the complex conjugate $\overline{z}$ as negating the imaginary part of z. Trivially we can see facts such as $\overline{(\overline{z})} = z$; $\overline{z + w} = \overline{z} + \overline{w}$ and $\overline{zw} = \overline{z} \cdot \overline{w}$.

The Fundamental Theorem of Algebra states that a polynomial of degree n can be written as a product of n linear factors:

$$c_n z^n + \cdots + c_1 z^1 + c_0 z^0 = c_n(z - \alpha_1)(z - \alpha_2) \cdots (z - \alpha_n) \quad (\text{where } c_i, \alpha_i \in \mathbb{C})$$

We can reformulate this statement as follows: a polynomial of degree n has n solutions α_i, counting repeats. This theorem is not proved in this course.

The modulus of complex numbers z_1, z_2 satisfies:

- (composition) $|z_1 z_2| = |z_1||z_2|$, and
- (triangle inequality) $|z_1 + z_2| \leq |z_1| + |z_2|$

Proof. The composition property is trivial. To prove the triangle inequality, we square both sides and compare.

$$\begin{aligned}
\text{LHS} &= |z_1 + z_2|^2 \\
&= (z_1 + z_2)\overline{(z_1 + z_2)} \\
&= |z_1|^2 + \overline{z_1} z_2 + z_1 \overline{z_2} + |z_2|^2 \\
\text{RHS} &= |z_1|^2 + 2|z_1||z_2| + |z_2|^2
\end{aligned}$$

Note that

$$\begin{aligned}
& \overline{z_1} z_2 + z_1 \overline{z_2} \leq 2|z_1||z_2| \\
\iff & \frac{1}{2}\left(\overline{z_1} z_2 + \overline{\overline{z_1} z_2}\right) \leq |z_1||z_2| \\
\iff & \operatorname{Re}(\overline{z_1} z_2) \leq |\overline{z_1} z_2|
\end{aligned}$$

which is true. □

We can alternatively use the map $z_2 \to z_2 - z_1$ to write the triangle inequality as

$$\begin{aligned}
|z_2 - z_1| &\geq |z_2| - |z_1| \\
\text{or } |z_2 - z_1| &\geq |z_1| - |z_2| \\
\therefore |z_2 - z_1| &\geq \big||z_2| - |z_1|\big|
\end{aligned}$$

De Moivre's Theorem states that

$$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta \quad (\forall n \in \mathbb{Z})$$

We can prove this using induction for $n \geq 0$. To show the negative case, simply use the positive result and raise it to the power of -1.

1.2 Complex valued functions

For $z \in \mathbb{C}$, we can define:

$$\begin{aligned} \exp z &= \sum_{n=0}^{\infty} \frac{1}{n!} z^n \\ \cos z &= \frac{1}{2}\left(e^{iz} + e^{-iz}\right) \\ \sin z &= \frac{1}{2i}\left(e^{iz} - e^{-iz}\right) \end{aligned}$$

By defining $\log z = w$ s.t. $e^w = z$, we have a complex logarithm function. By expanding the definition, we get that $\log z = \log r + i\theta$ where $r = |z|$ and $\theta = \arg z$. Note that because the argument of a complex number is multi-valued, so is the logarithm.

We can define exponentiation in the general case by defining $z^\alpha = e^{\alpha \log z}$. Depending on the choice of α, we have three cases:

- If $\alpha = p \in \mathbb{Z}$ then the result of z^p is unambiguous because

  $$z^p = e^{p \log z} = e^{p(\log r + i\theta + 2\pi i n)}$$

  which has a factor of $e^{2\pi i p n}$ which is 1.
- For a similar reason, a rational exponent has finitely many values.
- But in the general case, there are infinitely many values.

We can calculate results such as the square root of a complex number, which have two results as you might expect.

Note. We can't use facts like $z^\alpha z^\beta = z^{\alpha+\beta}$ in the complex case because the left and right hand sides both have infinite sets of answers, which may not be the same.

1.3 Transformations and primitives

We can represent a line passing through $x_0 \in \mathbb{C}$ parallel to $w \in \mathbb{C}$ using the formula:

$$z = z_0 + \lambda w \quad (\lambda \in \mathbb{R})$$

We can eliminate the dependency on λ by computing the conjugate of both sides:

$$\begin{aligned} \overline{z} &= \overline{z_0} + \lambda \overline{w} \\ \overline{w}z - w\overline{z} &= \overline{w}z_0 - w\overline{z_0} \end{aligned}$$

We can also write the equation for a circle with centre $c \in \mathbb{C}$ and radius $\rho \in \mathbb{R}$:

$$z = c + \rho e^{i\alpha}$$

or equivalently:

$$|z - c| = |\rho e^{i\alpha}| = \rho$$

or by squaring both sides:

$$|z|^2 - c\bar{z} - \bar{c}z = \rho^2 - |c|^2$$

2 Vectors in three dimensions

We use the normal Euclidean notions of points, lines, planes, length, angles and so on. By choosing an (arbitrary) origin point O, we may write positions as position vectors with respect to that origin point.

2.1 Vector addition and scalar multiplication

We define vector addition using the shape of a parallelogram with points $\mathbf{0}, \mathbf{a}, \mathbf{a} + \mathbf{b}, \mathbf{b}$. We define scalar multiplication of a vector using the line $\overrightarrow{OA}$ and setting the length to be multiplied by the constant. Note that this vector space is an abelian group under addition.

Definition. $\mathbf{a}$ and $\mathbf{b}$ are defined to be parallel if and only if $\mathbf{a} = \lambda\mathbf{b}$ or $\mathbf{b} = \lambda\mathbf{a}$ for some $\lambda \in \mathbb{R}$. This is denoted $\mathbf{a} \parallel \mathbf{b}$. Note that the vectors may be zero, in particular the zero vector is parallel to all vectors.

Definition. The span of a set of vectors is defined as $\operatorname{span}\{\mathbf{a}, \mathbf{b}, \cdots, \mathbf{c}\} = \{\alpha\mathbf{a} + \beta\mathbf{b} + \cdots + \gamma\mathbf{c} : \alpha, \beta, \gamma \in \mathbb{R}\}$. This is the line/plane/volume etc. containing the vectors. The span has an amount of dimensions at most equal to the amount of vectors in the input set. For example, the span of a set of two vectors may be a point, line or plane containing the vectors.

2.2 Scalar product

Definition. Given two vectors $\mathbf{a}, \mathbf{b}$, let θ be the angle between the two vectors. Then, we define

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos\theta$$

Note that if either of the vectors is zero, θ is undefined. However, the dot product is zero anyway here, so this is irrelevant.

Definition. Two vectors $\mathbf{a}$ and $\mathbf{b}$ are defined to be parallel (or orthogonal) if and only if $\mathbf{a} \cdot \mathbf{b} = 0$. This is denoted $\mathbf{a} \perp \mathbf{b}$. This is true in two cases:

(i) $\cos\theta = 0 \iff \theta = \frac{\pi}{2} \mod \pi$, or

(ii) $\mathbf{a} = 0$ or $\mathbf{b} = 0$.

Therefore, the zero vector is perpendicular to all vectors.

Definition. We can decompose a vector $\mathbf{b}$ into components relative to $\mathbf{a}$:

$$\mathbf{b} = \mathbf{b}_{\parallel} + \mathbf{b}_{\perp}$$

where $\mathbf{b}_{\parallel}$ is the component of $\mathbf{b}$ parallel to $\mathbf{a}$, and $\mathbf{b}_{\perp}$ is the component of $\mathbf{b}$ perpendicular to $\mathbf{a}$. In particular, we have that

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a} \cdot \mathbf{b}_{\parallel}$$

2.3 Vector product

Definition. Given two vectors $\mathbf{a}, \mathbf{b}$, let θ be the angle between the two vectors measured with respect to an arbitrary normal $\hat{\mathbf{n}}$. Then, we define

$$\mathbf{a} \wedge \mathbf{b} = \mathbf{a} \times \mathbf{b} = |\mathbf{a}||\mathbf{b}|\hat{\mathbf{n}} \sin\theta$$

Note that by swapping the sign of $\hat{\mathbf{n}}$, θ changes to $2\pi - \theta$, leaving the result unchanged. There are two degenerate cases:

- θ is undefined if $\mathbf{a}$ or $\mathbf{b}$ is the zero vector, but the result is zero anyway because we multiply by the magnitudes of both vectors.
- $\hat{\mathbf{n}}$ is undefined if $\mathbf{a} \parallel \mathbf{b}$, but here $\sin\theta = 0$ so the result is zero anyway.

We can provide several useful interpretations of the cross product:

- The magnitude of $\mathbf{a} \times \mathbf{b}$ is the vector area of the parallelogram defined by the points $\mathbf{0}, \mathbf{a}, \mathbf{a}+\mathbf{b}, \mathbf{b}$.
- By fixing a vector $\mathbf{a}$, we can consider the plane perpendicular to it. If $\mathbf{x}$ is another vector in the plane, $\mathbf{x} \mapsto \mathbf{a} \times \mathbf{x}$ rotates $\mathbf{x}$ by $\frac{\pi}{2}$ in the plane, scaling it by the magnitude of $\mathbf{a}$.

Note that by resolving a vector $\mathbf{b}$ perpendicular to another vector $\mathbf{a}$, we have that

$$\mathbf{a} \times \mathbf{b} = \mathbf{a} \times \mathbf{b}_{\perp}$$

A final useful property of the cross product is that since the result is perpendicular to both input vectors, we have

$$\mathbf{a} \cdot (\mathbf{a} \times \mathbf{b}) = \mathbf{b} \cdot (\mathbf{a} \times \mathbf{b}) = 0$$

2.4 Basis vectors

To represent vectors as some collection of numbers, we can choose some basis vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ which are 'orthonormal', i.e. they are unit vectors and pairwise orthogonal. Note that

$$\mathbf{e}_i \cdot \mathbf{e}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

The set $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is called a basis because any vector can be written uniquely as a linear combination of the basis vectors. Because we have orthonormal basis vectors, we can reduce this to

$$\mathbf{a} = \sum_i \mathbf{a}_i \mathbf{e}_i \implies \mathbf{a}_i = \mathbf{e}_i \cdot \mathbf{a}$$

By representing a vector as a linear combination of basis vectors, it is very easy to evaluate the scalar product algebraically. To calculate the vector product, we first need to define whether $\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3$ or $-\mathbf{e}_3$. By convention, we assume that the basis vectors are right-handed, i.e. $\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3$. Then we can calculate the formula for the cross product in terms of the vectors' components.

2.5 Scalar triple product

The scalar triple product is the scalar product of one vector with the cross product of two more.

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) = [\mathbf{a}, \mathbf{b}, \mathbf{c}]$$

The result of the scalar triple product is the signed volume of the parallelepiped starting at the origin with axes $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$. We can represent this triple product as the determinant of a matrix:

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \begin{vmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \\ \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 \\ \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 \end{vmatrix}$$

If the scalar triple product is greater than zero, then $\mathbf{a}, \mathbf{b}, \mathbf{c}$ is called a right handed set. If it is equal to zero, then the vectors are all coplanar: $\mathbf{c} \in \operatorname{span}\{\mathbf{a}, \mathbf{b}\}$.

2.6 Vector triple product

The vector triple product is the cross product of three vectors. Note that this is non-associative. The proof is covered in the subsequent lecture.

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}$$

$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{b} \cdot \mathbf{c})\mathbf{a}$$

2.7 Lines

A line through $\mathbf{a}$ parallel to $\mathbf{u}$ is defined by

$$\mathbf{r} = \mathbf{a} + \lambda \mathbf{u}$$

where λ is some real parameter. We can eliminate lambda by using the cross product with $\mathbf{u}$. This will allow us to get a $\mathbf{u} \times \mathbf{u}$ term which will cancel to zero.

$$\mathbf{u} \times \mathbf{r} = \mathbf{u} \times \mathbf{a}$$

Informally, this is saying that $\mathbf{r}$ and $\mathbf{a}$ have the same components perpendicular to $\mathbf{u}$. Note that we can also reverse this process. Consider the equation

$$\mathbf{u} \times \mathbf{r} = \mathbf{c}$$

By using the dot product with $\mathbf{u}$ we can say

$$\mathbf{u} \cdot (\mathbf{u} \times \mathbf{r}) = \mathbf{u} \cdot \mathbf{c}$$

If $\mathbf{u} \cdot \mathbf{c} \neq 0$ then the equation is inconsistent. Otherwise, we can suppose that maybe $\mathbf{r} = \mathbf{u} \times \mathbf{c}$ and use the formula for the vector product to get the left hand side to be $\mathbf{u} \times (\mathbf{u} \times \mathbf{c}) = -|\mathbf{u}|^2 \mathbf{c}$. Therefore, by inspection, $\mathbf{a} = -\frac{1}{|\mathbf{u}|^2}(\mathbf{u} \times \mathbf{c})$ is a solution. Now, note that we can add any multiple of $\mathbf{u}$ to $\mathbf{a}$ and it remains a solution. So the general solution is $\mathbf{r} = \mathbf{a} + \lambda \mathbf{u}$.

2.8 Planes

The general point on a plane that passes through $\mathbf{a}$ and has directions $\mathbf{u}$ and $\mathbf{v}$ is

$$\mathbf{r} = \mathbf{a} + \lambda\mathbf{u} + \mu\mathbf{v}$$

where $\mathbf{u}$ and $\mathbf{v}$ are not parallel, and λ and μ are real parameters. We can do a dot product with $\mathbf{n} = (\mathbf{u} \times \mathbf{v})$ to eliminate both parameters.

$$\mathbf{n} \cdot \mathbf{r} = \varkappa$$

where $\varkappa = \mathbf{n} \cdot \mathbf{a}$. Note that $|\varkappa|/|\mathbf{n}|$ is the perpendicular distance from the origin to the plane.

2.9 Other vector equations

The equation of a sphere is given by a quadratic vector equation in $\mathbf{r}$.

$$\mathbf{r}^2 + \mathbf{r} \cdot \mathbf{a} = k$$

We can complete the square to give

$$\left(\mathbf{r} + \frac{1}{2}\mathbf{a}\right)^2 = \frac{1}{4}\mathbf{a}^2 + k$$

which is clearly a sphere with centre $-\frac{1}{2}\mathbf{a}$ and radius $\left(\frac{1}{4}\mathbf{a}^2 + k\right)^{1/2}$.

Another example of a vector equation is

$$\mathbf{r} + \mathbf{a} \times (\mathbf{b} \times \mathbf{r}) = \mathbf{c} \tag{1}$$

where $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are fixed. We can dot with $\mathbf{a}$ to eliminate the second term:

$$\mathbf{a} \cdot \mathbf{r} = \mathbf{a} \cdot \mathbf{c} \tag{2}$$

Note that using the dot product loses information—this is simply a tool to make deductions; (2) does not contain the full information of (1). Combining (1) and (2), and using the formula for the vector triple product, we get

$$\begin{aligned} & \mathbf{r} + (\mathbf{a} \cdot \mathbf{r})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{r} = \mathbf{c} && (3) \\ \implies & \mathbf{r} + (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{r} = \mathbf{c} \end{aligned}$$

This eliminates the dependency on $\mathbf{r}$ inside the dot product. Now, we can factorise, leaving

$$(1 - \mathbf{a} \cdot \mathbf{b})\mathbf{r} = \mathbf{c} - (\mathbf{a} \cdot \mathbf{c})\mathbf{b} \tag{4}$$

If $1 - \mathbf{a} \cdot \mathbf{b} \neq 0$ then $\mathbf{r}$ has a single solution, a point. Otherwise, the right hand side must also be zero (otherwise the equation is inconsistent). Therefore, $\mathbf{c} - (\mathbf{a} \cdot \mathbf{c})\mathbf{b} = \mathbf{0}$. We can now combine this expression for $\mathbf{c}$ into (3), eliminating the $(1 - \mathbf{a} \cdot \mathbf{b})$ term, to get

$$(\mathbf{a} \cdot \mathbf{r} - \mathbf{a} \cdot \mathbf{c})\mathbf{b} = \mathbf{0}$$

This shows us that (given that $\mathbf{b}$ is nonzero) the solutions to the equation are given by (2), which is the equation of a plane.

3 Index notation and the summation convention

3.1 Kronecker δ and Levi-Civita ε

The Kronecker δ is defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Then $\mathbf{e}_i \mathbf{e}_j = \delta_{ij}$. We can also use δ to rewrite indices: $\sum_i \delta_{ij} \mathbf{a}_i = \mathbf{a}_j$. So

$$\begin{aligned} \mathbf{a} \cdot \mathbf{b} &= \left(\sum_i \mathbf{a}_i \mathbf{e}_i \right) \cdot \left(\sum_j \mathbf{b}_j \mathbf{e}_j \right) \\ &= \sum_{ij} \mathbf{a}_i \mathbf{b}_j (\mathbf{e}_i \cdot \mathbf{e}_j) \\ &= \sum_{ij} \mathbf{a}_i \mathbf{b}_j \delta_{ij} \\ &= \sum_i \mathbf{a}_i \mathbf{b}_i \end{aligned}$$

The Levi-Civita ε is defined by

$$\varepsilon_{ijk} = \begin{cases} +1 & \text{if } ijk \text{ is an even permutation of } [1,2,3] \\ -1 & \text{if } ijk \text{ is an odd permutation of } [1,2,3] \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\begin{aligned} \varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = +1 \\ \varepsilon_{132} = \varepsilon_{321} = \varepsilon_{213} = -1 \end{aligned}$$

and all other permutations of $[1,2,3]$ yield 0. This shows that ε is totally antisymmetric; exchanging any pair of indices changes the sign. We now have:

$$\mathbf{e}_i \times \mathbf{e}_j = \sum_k \varepsilon_{ijk} \mathbf{e}_k$$

And:

$$\begin{aligned} \mathbf{a} \times \mathbf{b} &= \left(\sum_i \mathbf{a}_i \mathbf{e}_i \right) \times \left(\sum_j \mathbf{b}_j \mathbf{e}_j \right) \\ \mathbf{a} \times \mathbf{b} &= \sum_{ij} \mathbf{a}_i \mathbf{b}_j \left(\mathbf{e}_i \times \mathbf{e}_j \right) \\ \mathbf{a} \times \mathbf{b} &= \sum_{ijk} \mathbf{a}_i \mathbf{b}_j \varepsilon_{ijk} \mathbf{e}_k \end{aligned}$$

So the individual terms of the cross product can be written

$$(\mathbf{a} \times \mathbf{b})_k = \sum_{ij} \mathbf{a}_i \mathbf{b}_j \varepsilon_{ijk}$$

We use the 'summation convention' to abbreviate the many summation symbols used throughout linear algebra.

(i) An index which occurs exactly once in some term, called a 'free' index, must appear once in every term in that equation.

(ii) An index which occurs exactly twice in a given term, called a 'repeated', 'contracted', or 'dummy' index, is implicitly summed over.

(iii) No index can occur more than twice in a given term.

3.2 Identities

The most general $\varepsilon\varepsilon$ identity is as follows:

$$\begin{aligned}\varepsilon_{ijk}\varepsilon_{pqr} &= \delta_{ip}\delta_{jq}\delta_{kr} - \delta_{jp}\delta_{iq}\delta_{kr} \\ &+ \delta_{jp}\delta_{kq}\delta_{ir} - \delta_{kp}\delta_{jq}\delta_{ir} \\ &+ \delta_{kp}\delta_{iq}\delta_{jr} - \delta_{ip}\delta_{kq}\delta_{jr}\end{aligned}$$

This is, however, very verbose and not used often throughout the course. It is provable by noting the total antisymmetry in i, j, k and p, q, r on both sides of the equation implies that both sides agree up to a constant factor. We can check that this factor is 1 by substituting in values such as $i = p = 1$, $j = q = 2$ and $k = r = 3$.

The next most generic form is a very useful identity.

$$\varepsilon_{ijk}\varepsilon_{pqk} = \delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp}$$

This is essentially the first line of the above identity, noting that $k = r$. We can prove this is true by observing the antisymmetry, and that both sides vanish under $i = j$ or $p = q$. So it suffices to check two cases: $i = p, j = q$ and $i = q, j = p$.

We can now continue making more indices equal to each other to get even more specific identities:

$$\varepsilon_{ijk}\varepsilon_{pjk} = 2\delta_{ip}$$

This is easy to prove by noting that $\delta_{jj} = \sum_j \delta_{jj} = 3$, and using the δ rewrite rule.

Finally, we have

$$\varepsilon_{ijk}\varepsilon_{ijk} = 6$$

No indices are free here, so the values of i, j, k themselves are predetermined by the fact that we are in three-dimensional space.

Using the summation convention (as will now be implied for the remainder of the course), we can prove the vector triple product identity

$$\begin{aligned}[\mathbf{a} \times (\mathbf{b} \times \mathbf{c})]_i &= \varepsilon_{ijk}\mathbf{a}_j(\mathbf{b} \times \mathbf{c})_k \\ &= \varepsilon_{ijk}\mathbf{a}_j\varepsilon_{pqk}\mathbf{b}_p\mathbf{c}_q \\ &= \varepsilon_{ijk}\varepsilon_{pqk}\mathbf{a}_j\mathbf{b}_p\mathbf{c}_q \\ &= (\delta_{ip}\delta_{jq})\mathbf{a}_j\mathbf{b}_p\mathbf{c}_q - (\delta_{iq}\delta_{jp})\mathbf{a}_j\mathbf{b}_p\mathbf{c}_q \\ &= (\mathbf{a} \cdot \mathbf{c})\mathbf{b}_i - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}_i\end{aligned}$$

4 Higher dimensional vectors

4.1 Multidimensional real space

We define multidimensional real space as follows:

$$\mathbb{R}^n = \{\mathbf{x} = (x_1, x_2, \cdots, x_n) : x_i \in \mathbb{R}\}$$

We can define addition and scalar multiplication by mapping these operations over each term in the tuple. Therefore, we have a notion of linear combinations of vectors and hence a concept of parallel vectors. We can say, like before in $\mathbb{R}^3$, that $\mathbf{x} \parallel \mathbf{y}$ if and only if $\mathbf{x} = \lambda\mathbf{y}$ or $\mathbf{y} = \lambda\mathbf{x}$.

We define an operator analogous to the scalar product in $\mathbb{R}^3$. The inner product is defined as $x \cdot y = x_i y_i$. Directly from this definition, we can deduce some properties:

- (symmetric) $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$
- (bilinear) $(\lambda\mathbf{x} + \lambda'\mathbf{x}') \cdot \mathbf{y} = \lambda\mathbf{x} \cdot \mathbf{y} + \lambda'\mathbf{x}' \cdot \mathbf{y}$
- (positive definite) $\mathbf{x} \cdot \mathbf{x} \geq 0$, and the equality holds if and only if $\mathbf{x} = \mathbf{0}$.

We can define the norm of a vector (similar to the concept of length in three-dimension space), denoted $|\mathbf{x}|$, by $|\mathbf{x}|^2 = \mathbf{x} \cdot \mathbf{x}$. We can now define orthogonality as follows: $\mathbf{x} \perp \mathbf{y} \iff \mathbf{x} \cdot \mathbf{y} = 0$.

We define the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ by setting each element of the tuple $\mathbf{e}_i$ to zero apart from the ith element, which is set to one. Also, we redefine the Kronecker δ to be valid in higher-dimensional space. Note that under this definition, the standard basis vectors are orthonormal because $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$.

4.2 Cauchy–Schwarz inequality

Proposition. For vectors $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^n$, $|\mathbf{x} \cdot \mathbf{y}| \leq |\mathbf{x}||\mathbf{y}|$, where the equality is true if and only if $\mathbf{x} \parallel \mathbf{y}$.

Proof. If $\mathbf{y} = \mathbf{0}$, then the result is immediate. So suppose that $\mathbf{y} \neq 0$, then for some $\lambda \in \mathbb{R}$, we have

$$\begin{aligned} |\mathbf{x} - \lambda\mathbf{y}|^2 &= (\mathbf{x} - \lambda\mathbf{y}) \cdot (\mathbf{x} - \lambda\mathbf{y}) \\ &= |\mathbf{x}|^2 - 2\lambda\mathbf{x} \cdot \mathbf{y} + \lambda^2|\mathbf{y}|^2 \geq 0 \end{aligned}$$

As this is a positive real quadratic in λ that is always greater than zero, it has at most one real root. Therefore the discriminant is less than or equal to zero.

$$(-2\mathbf{x} \cdot \mathbf{y})^2 - 4|\mathbf{x}|^2|\mathbf{y}|^2 \leq 0 \implies |\mathbf{x} \cdot \mathbf{y}| \leq |\mathbf{x}||\mathbf{y}|$$

where the equality only holds if $\mathbf{x}$ and $\mathbf{y}$ are parallel (i.e. when $\mathbf{x} - \lambda\mathbf{y}$ equals zero for some λ). □

4.3 Triangle inequality

Following from the Cauchy–Schwarz inequality,

$$\begin{aligned} |\mathbf{x} + \mathbf{y}|^2 &= |\mathbf{x}|^2 + 2(\mathbf{x} \cdot \mathbf{y}) + |\mathbf{y}|^2 \\ &\leq |\mathbf{x}|^2 + 2|\mathbf{x}||\mathbf{y}| + |\mathbf{y}|^2 \\ &= (|\mathbf{x}| + |\mathbf{y}|)^2 \end{aligned}$$

where the equality holds under the same conditions as above.

4.4 Levi-Civita ε in higher dimensions

Note that the Levi-Civita ε has three indices in $\mathbb{R}^3$. We can extend this ε to higher and lower dimensions by increasing or reducing the amount of indices. It does not make logical sense to use the same ε without changing the amount of indices to define, for example, a vector product in four-dimensional space, since we would have unused indices. The expression $(\mathbf{x} \times \mathbf{y})_k = \varepsilon_{ijk}\mathbf{a}_i\mathbf{b}_j$ works because there is one free index, k, on the right hand side, so we can use this to calculate the values of each element of the result.

We can, however, use this ε to extend the notion of a scalar triple product to other dimensions, for example two-dimensional space, with $[\mathbf{a}, \mathbf{b}] \coloneqq \varepsilon_{ij}\mathbf{a}_i\mathbf{b}_j$. This is the signed area of the parallelogram spanning $\mathbf{a}$ and $\mathbf{b}$.

4.5 General real vector spaces

Vector spaces are not studied axiomatically in this course, but the axioms are given here for completeness. A real (as in, $\mathbb{R}$) vector space V is a set of objects with two operators $+ : V \times V \to V$ and $\cdot : \mathbb{R} \times V \to V$ such that

- $(V, +)$ is an abelian group
- $\lambda(v + w) = \lambda v + \lambda w$
- $(\lambda + \mu)v = \lambda v + \mu v$
- $\lambda(\mu v) = (\lambda\mu)v$
- $1v = v$ (to exclude trivial cases for example $\lambda v = 0$ for all v)

A subspace of a real vector space V is a subset $U \subseteq V$ that is a vector space. Equivalently, if all pairs of vectors $v, w \in U$ satisfy $\lambda v + \mu w \in U$, then U is a subspace of V. Note that the span generated from a set of vectors is a subspace, as it is characterised by this equivalent definition. Also, note that the origin must be part of any subspace, because multiplying a vector by zero must yield the origin.

In some real vector space V, let $\mathbf{v}_1, \mathbf{v}_2 \cdots \mathbf{v}_r$ be vectors in V. Now consider the linear relation

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_r\mathbf{v}_r = 0$$

Then we call the set of vectors a linearly independent set if the only solution is where all λ values are zero. Otherwise, it is a linearly dependent set.

4.6 Inner product spaces

An inner product is an extra structure that we can have on a real vector space V, which is often denoted by angle brackets or parentheses. It can also be characterised by axioms (specifically the ones in Section 6.2). Features like the norm of a vector, and theorems like the Cauchy–Schwarz inequality, follow from these axioms.

For example, let us consider the vector space

$$V = \{f : [0, 1] \to \mathbb{R} : f \text{ smooth}; f(0) = f(1) = 0\}$$

We can define the inner product to be

$$f \cdot g = \langle f, g \rangle = \int_0^1 f(x)g(x)\,\mathrm{d}x$$

Then by the Cauchy–Schwarz inequality, we have

$$|\langle f, g \rangle| \leq \|f\| \cdot \|g\|$$
$$\therefore \left| \int_0^1 f(x)g(x)\,\mathrm{d}x \right| \leq \sqrt{\int_0^1 f(x)^2\,\mathrm{d}x} \sqrt{\int_0^1 g(x)^2\,\mathrm{d}x}$$

Lemma. In any real inner product space V, if $\mathbf{v}_1 \cdots v_r \neq \mathbf{0}$ are orthogonal, they are linearly independent.

Proof. If $\sum_i \alpha_i \mathbf{v}_i = 0$, then

$$\left\langle \mathbf{v}_j, \sum_i \alpha_i \mathbf{v}_i \right\rangle = 0$$

And because each vector that is not $\mathbf{v}_j$ is orthogonal to it, those terms cancel, leaving

$$\therefore \langle \mathbf{v}_j, \alpha_j \mathbf{v}_j \rangle = 0$$
$$\alpha_j \langle \mathbf{v}_j, \mathbf{v}_j \rangle = 0$$
$$\alpha_j = 0$$

So they are linearly independent. □

4.7 Bases and dimensions

In a vector space V, a basis is a set $\mathcal{B} = \{\mathbf{e}_1 \cdots \mathbf{e}_n\}$ such that

- $\mathcal{B}$ spans V; and
- $\mathcal{B}$ is linearly independent, which implies that the coefficients on these basis vectors are unique for any vector in V, since it is impossible to write one vector in terms of the others

Theorem. If $\{\mathbf{e}_1 \cdots \mathbf{e}_n\}$ and $\{\mathbf{f}_1 \cdots \mathbf{f}_m\}$ are bases for a real vector space V, then $n = m$, which we call the dimension of V.

Proof. This proof is non-examinable (without prompts). We can write each basis vector in terms of the others, since they all span the same vector space. Thus:

$$\mathbf{f}_a = \sum_i A_{ai} \mathbf{e}_i; \quad \mathbf{e}_i = \sum_a B_{ia} \mathbf{f}_a$$

Note that indices i, j span from 1 to n, while a, b span from 1 to m. We can substitute one expression into the other, forming:

$$\mathbf{f}_a = \sum_i A_{ai}\left(\sum_b B_{ib}\mathbf{f}_b\right)$$
$$\mathbf{f}_a = \sum_b\left(\sum_i A_{ai}B_{ib}\right)\mathbf{f}_b$$

Note that we have now written $\mathbf{f}_a$ as a linear combination of $\mathbf{f}_b$ for all valid b. But since they are linearly independent, the coefficient of $\mathbf{f}_b$ must be zero if $a \neq b$, and one of $a = b$. Therefore, we have

$$\delta_{ab} = \sum_i A_{ai}B_{ib}$$

We can make a similar statement about $\mathbf{e}_i$:

$$\delta_{ij} = \sum_a B_{ia}A_{aj} = \sum_a A_{aj}B_{ia}$$

Now, assigning $a = b$ and $i = j$, summing over both, and substituting into our two previous expressions for δ, we have:

$$\sum_{ia} A_{ai}B_{ia} = \sum_a \delta_{aa} = \sum_i \delta_{ii}$$
$$= m \qquad = n$$

$\square$

Note that $\{\mathbf{0}\}$ is a trivial subspace of all vector spaces, and it has dimension zero since it requires a linear combination of no vectors.

Proposition. Let V be a vector space with finite subsets $Y = \{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ that spans V, and $X = \{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ that is linearly independent. Let $n = \dim V$. Then:

(i) A basis can be found as a subset of Y by discarding vectors in Y as necessary, and that $n \leq m$.

(ii) X can be extended to a basis by adding in additional vectors from Y as necessary, and that $k \leq n$.

Proof. This proof is non-examinable (without prompts).

(i) If Y is linearly independent, then Y is a basis and $m = n$. Otherwise, Y is not linearly independent. So there exists some linear relation

$$\sum_{i=1}^{m} \lambda_i \mathbf{w}_i = \mathbf{0}$$

where there is some i such that $\lambda_i \neq 0$. Without loss of generality (because the order of elements in Y does not matter) we will reorder Y such that $\mathbf{w}_m \neq 0$. So we have

$$\mathbf{w}_m = \frac{-1}{\lambda_m}\sum_{i=1}^{m-1} \lambda_i \mathbf{w}_i$$

So span Y = span$(Y \setminus \{\mathbf{w}_m\})$. We can repeat this process of eliminating vectors from Y until linear independence is achieved. We know that this process will end because Y is a finite set. Clearly, in this case, $n < m$. So for all cases, $n \leq m$.

(ii) If X spans V, then X is a basis and $k = n$. Else, there exists some $u_{k+1} \in V$ that is not in the span of X. Then, we will construct an arbitrary linear relation

$$\sum_{i=1}^{k+1} \mu_i \mathbf{u}_i = \mathbf{0}$$

Note that this implies that $\mu_{k+1} = \mathbf{0}$ because it is not in the span of X, and that $\mu_i = 0$ for all $i \leq k$ because the original X was linearly independent. So we know that all the coefficients are zero, and therefore $X \cup \{u_{k+1}\}$ is linearly independent.

Note that we can always choose this u_{k+1} to be an element of Y because we just need to ensure that $u_{k+1} \notin \operatorname{span} X$. Suppose we cannot choose such a vector in Y. Then $Y \subseteq \operatorname{span} X \implies \operatorname{span} Y \subseteq \operatorname{span} X \implies \operatorname{span} X = V$, which is clearly false because X does not span V. This is a contradiction, so we can always choose such a vector from Y. We can repeat this process of taking vectors from Y and adding them to X until we have a basis. This process will always terminate in a finite amount of steps because we are taking new vectors from a finite set Y. Therefore $k \leq n$, as we are adding vectors (increasing k) until $k = n$.

□

It is perfectly possible to have a vector space that has infinite dimensionality. However, they will be rarely touched upon in this course apart from specific examples, like the following example. Let $V = \{f : [0,1] \to \mathbb{R} : f \text{ smooth}, f(0) = f(1) = 0\}$. Then let $S_n(x) = \sqrt{2}\sin(n\pi x)$ where n is a natural number $1, 2, \cdots$. Clearly, $S_n \in V$ for all n. The inner product of two of these S functions is given by

$$\begin{aligned}\langle S_n, S_m \rangle &= 2\int_0^1 \sin(n\pi x)\sin(m\pi x)\,\mathrm{d}x \\ &= \delta_{mn}\end{aligned}$$

So S_n are orthonormal and therefore linearly independent. So we can continue adding more vectors until it becomes a basis. However, the set of all S_n is already infinite—so V must have infinite dimensionality.

4.8 Multidimensional complex space

We define $\mathbb{C}^n$ by

$$\mathbb{C}^n \coloneqq \{\mathbf{z} = (z_1, z_2, \cdots, z_n) : \forall i, z_i \in \mathbb{C}\}$$

We define addition and scalar multiplication in obvious ways. Note that we have a choice over what the scalars are allowed to be. If we only allow scalars that are real numbers, $\mathbb{C}^n$ can be considered a real vector space with bases $(0, \cdots, 1, \cdots, 0)$ and $(0, \cdots, i, \cdots, 0)$ and dimension $2n$. Alternatively, if we let the scalars be any complex numbers, we don't need to have imaginary bases, thus giving us a complex vector space with bases $(0, \cdots, 1, \cdots, 0)$ and dimension n. We can say that $\mathbb{C}^n$ has dimension $2n$ over $\mathbb{R}$, and dimension n over $\mathbb{C}$. From here on, unless stated otherwise, we treat $\mathbb{C}^n$ to be a complex vector space.

We can define the inner product by

$$\langle \mathbf{z}, \mathbf{w} \rangle := \sum_j \overline{z_j} w_j$$

The conjugate over the z terms ensures that the inner product is positive definite. It has these properties, analogous to the properties of the inner product in the real vector space $\mathbb{R}^n$:

- (Hermitian) $\langle \mathbf{z}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{z} \rangle}$
- (linear/antilinear) $\langle \mathbf{z}, \lambda\mathbf{w} + \lambda'\mathbf{w}' \rangle = \lambda\langle \mathbf{z}, \mathbf{w} \rangle + \lambda'\langle \mathbf{z}, \mathbf{w}' \rangle$ and $\langle \lambda\mathbf{z} + \lambda'\mathbf{z}', w \rangle = \overline{\lambda}\langle \mathbf{z}, \mathbf{w} \rangle + \overline{\lambda'}\langle \mathbf{z}', \mathbf{w} \rangle$
- (positive definite) $\langle \mathbf{z}, \mathbf{z} \rangle = \sum_j \left|z_j\right|^2$ which is real and greater than or equal to zero, where the equality holds if and only if $\mathbf{z} = \mathbf{0}$.

We can also define the norm of $\mathbf{z}$ to satisfy $|\mathbf{z}| \geq 0$ and $|\mathbf{z}|^2 = \langle \mathbf{z}, \mathbf{z} \rangle$. Note that the standard basis for $\mathbb{C}^n$ is orthonormal, since the inner product of any two basis vectors $\mathbf{e}_j$ and $\mathbf{e}_k$ is given by δ_{jk}.

Here is an example of the use of the complex inner product on $\mathbb{C}^1 = \mathbb{C}$. Note first that $\langle z, w \rangle = \overline{z}w$. Let $z = a_1 + ia_2$ and $w = b_1 + ib_2$ where $a_1, a_2, b_1, b_2 \in \mathbb{R}$. Then

$$\begin{aligned}\langle z, w \rangle &= \overline{z}w \\ &= (a_1b_1 + a_2b_2) + i(a_1b_2 - a_2b_1) \\ &= (z \cdot w) + i[z, w]\end{aligned}$$

We can therefore use the inner product to compute two different scalar products at the same time.

5 Linear maps

5.1 Introduction

A linear map (or linear transformation) is some operation $T : V \to W$ between vector spaces V and W preserving the core vector space structure (specifically, the linearity). It is defined such that

$$T(\lambda\mathbf{x} + \mu\mathbf{y}) = \lambda T(\mathbf{x}) + \mu T(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in V$ where the scalars λ and μ match up with the scalar field that V and W use (so this could be $\mathbb{R}$ or $\mathbb{C}$ in our examples). Much of the language used for linear maps between vector spaces is analogous to the language used for homomorphisms between groups.

Note that a linear map is completely determined by its action on a basis $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ where $n = \dim V$, since

$$T\left(\sum_i x_i\mathbf{e}_i\right) = \sum_i x_i T(\mathbf{e}_i)$$

We denote $\mathbf{x}' = T(\mathbf{x}) \in W$, and define $\mathbf{x}'$ as the image of x under T. Further, we define

$$\mathrm{Im}(T) = \{\mathbf{x}' \in W : \mathbf{x}' = T(\mathbf{x}) \text{ for some } \mathbf{x} \in V\}$$

to be the image of T, and we define

$$\ker(T) = \{\mathbf{x} \in V : T(\mathbf{x}) = \mathbf{0}\}$$

to be the kernel of T.

Lemma. $\ker T$ is a subspace of V, and $\operatorname{Im} T$ is a subspace of W.

Proof. To verify that some subset is a subspace, it suffices to check that it is non-empty, and that it is closed under linear combinations.

$\ker T$ is non-empty because $\mathbf{0} \in \ker T$. For $\mathbf{x}, \mathbf{y} \in \ker T$, we have $T(\lambda\mathbf{x} + \mu\mathbf{y}) = \lambda T(\mathbf{x}) + \mu T(\mathbf{y}) = \mathbf{0} \in \ker T$ as required.

$\operatorname{Im} T$ is non-empty because $\mathbf{0} \in \operatorname{Im} T$. For $\mathbf{x}, \mathbf{y} \in V$, let $\mathbf{x}' = T(\mathbf{x})$ and $\mathbf{y}' = T(\mathbf{y})$, therefore $\mathbf{x}', \mathbf{y}' \in \operatorname{Im} T$. Now, $\lambda\mathbf{x}' + \mu\mathbf{y}' = T(\lambda\mathbf{x} + \mu\mathbf{y})$ so it is closed under linear combinations as required. □

Here are some examples of images and kernels.

(i) The zero linear map $\mathbf{x} \mapsto \mathbf{0}$ has:

$$\begin{aligned} \operatorname{Im} T &= \{\mathbf{0}\} \\ \ker T &= V \end{aligned}$$

(ii) The identity linear map $\mathbf{x} \mapsto \mathbf{x}$ has:

$$\begin{aligned} \operatorname{Im} T &= V \\ \ker T &= \{\mathbf{0}\} \end{aligned}$$

(iii) Let $T : \mathbb{R}^3 \to \mathbb{R}^3$, such that

$$\begin{aligned} x_1' &= 3x_1 - x_2 + 5x_3 \\ x_2' &= -x_1 - 2x_3 \\ x_3' &= 2x_1 + x_2 + 3x + 3 \end{aligned}$$

This map has

$$\begin{aligned} \operatorname{Im} T &= \left\{ \lambda \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} : \lambda, \mu \in \mathbb{R} \right\} \\ \ker T &= \left\{ \lambda \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} : \lambda \in \mathbb{R} \right\} \end{aligned}$$

5.2 Rank and nullity

We define the rank of a linear map to be the dimension of its image, and the nullity of a linear map to be the dimension of its kernel.

$$\operatorname{rank} T = \dim \operatorname{Im} T; \quad \operatorname{null} T = \dim \ker T$$

Note that therefore for $T : V \to W$, we have $\operatorname{rank} T \leq \dim W$ and $\ker T \leq \dim V$.

Theorem. For some linear map $T : V \to W$,

$$\operatorname{rank} T + \operatorname{null} T = \dim V$$

Proof. This proof is non-examinable (without prompts). Let $\mathbf{e}_1, \cdots, \mathbf{e}_k$ be a basis for $\ker T$, so $T(\mathbf{e}_i) = \mathbf{0}$ for all valid i. We may extend this basis by adding more vectors $\mathbf{e}_i$ where $k < i \leq n$ until we have a basis for V, where $n = \dim V$. We claim that the set $\mathcal{B} = \{T(\mathbf{e}_{k+1}), \cdots, T(\mathbf{e}_n)\}$ is a basis for $\operatorname{Im} T$. If this is true, then clearly the result follows because $k = \dim \ker T = \operatorname{null} T$ and $n-k = \dim \operatorname{Im} T = \operatorname{rank} T$.

To prove the claim we need to show that $\mathcal{B}$ spans $\operatorname{Im} T$ and that it is a linearly independent set.

- $\mathcal{B}$ spans $\operatorname{Im} T$ because for any $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i$, we have

$$T(\mathbf{x}) = \sum_{i=k+1}^n x_i T(\mathbf{e}_i) \in \operatorname{span} \mathcal{B}$$

- $\mathcal{B}$ is linearly independent. Consider a general linear combination of basis vectors:

$$\sum_{i=k+1}^n \lambda_i T(\mathbf{e}_i) = 0 \implies T\left(\sum_{i=k+1}^n \lambda_i \mathbf{e}_i\right) = 0$$

  so

$$\sum_{i=k+1}^n \lambda_i \mathbf{e}_i \in \ker T$$

  Because this is in the kernel, it may be written in terms of the basis vectors of the kernel. So, we have

$$\sum_{i=k+1}^n \lambda_i \mathbf{e}_i = \sum_{i=1}^k \mu_i \mathbf{e}_i$$

  This is a linear relation in terms of all basis vectors of V. So all coefficients are zero.

$\square$

5.3 Rotations

Linear maps are often used to describe geometrical transformations, such as rotations, reflections, projections, dilations and shears. A convenient way to express these maps is by describing where the basis vectors are mapped to. In $\mathbb{R}^2$, we may describe a rotation anticlockwise around the origin by angle θ with

$$\begin{aligned} \mathbf{e}_1 &\mapsto \cos\theta \mathbf{e}_1 + \sin\theta \mathbf{e}_2 \\ \mathbf{e}_2 &\mapsto -\sin\theta \mathbf{e}_1 + \cos\theta \mathbf{e}_2 \end{aligned}$$

In $\mathbb{R}^3$ we can construct a similar transformation for a rotation around the $\mathbf{e}_3$ axis with

$$\begin{aligned} \mathbf{e}_1 &\mapsto \cos\theta \mathbf{e}_1 + \sin\theta \mathbf{e}_2 \\ \mathbf{e}_2 &\mapsto -\sin\theta \mathbf{e}_1 + \cos\theta \mathbf{e}_2 \\ \mathbf{e}_3 &\mapsto \mathbf{e}_3 \end{aligned}$$

We can extend this to a general rotation in $\mathbb{R}^3$ about an axis given by a unit normal vector $\hat{\mathbf{n}}$. For any vector $\mathbf{x} \in \mathbb{R}^3$ we can resolve parallel and perpendicular to $\hat{\mathbf{n}}$ as follows.

$$\mathbf{x} = \mathbf{x}_{\parallel} + \mathbf{x}_{\perp}; \quad \mathbf{x}_{\parallel} = (\mathbf{x} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}; \quad \mathbf{x}_{\perp} = \mathbf{x} - (\mathbf{x} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}$$

Note that $\hat{\mathbf{n}}$ resembles the $\mathbf{e}_3$ axis here, and $\mathbf{x}_{\perp}$ resembles the $\mathbf{e}_1$ axis. So we can compute the equivalent of $\mathbf{e}_2$ using the cross product, $\hat{\mathbf{n}} \times \mathbf{x}_{\perp} = \hat{\mathbf{n}} \times \mathbf{x}$. Now we may define the map with

$$\begin{aligned} \mathbf{x}_{\parallel} &\mapsto \mathbf{x}_{\parallel} \\ \mathbf{x}_{\perp} &\mapsto (\cos\theta)\mathbf{x}_{\perp} + (\sin\theta)(\hat{\mathbf{n}} \times \mathbf{x}) \end{aligned}$$

So all together, we have

$$\mathbf{x} \mapsto (\cos\theta)\mathbf{x} + (1 - \cos\theta)(\hat{\mathbf{n}} \cdot \mathbf{x})\hat{\mathbf{n}} + (\sin\theta)(\hat{\mathbf{n}} \times \mathbf{x})$$

5.4 Reflections and projections

For a plane with normal $\hat{\mathbf{n}}$, we define a projection to be

$$\begin{aligned} \mathbf{x}_{\parallel} &\mapsto \mathbf{0} \\ \mathbf{x}_{\perp} &\mapsto \mathbf{x}_{\perp} \\ \mathbf{x} &\mapsto \mathbf{x}_{\perp} = \mathbf{x} - (\mathbf{x} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}} \end{aligned}$$

and a reflection to be

$$\begin{aligned} \mathbf{x}_{\parallel} &\mapsto -\mathbf{x}_{\parallel} \\ \mathbf{x}_{\perp} &\mapsto \mathbf{x}_{\perp} \\ \mathbf{x} &\mapsto \mathbf{x}_{\perp} - \mathbf{x}_{\parallel} = \mathbf{x} - 2(\mathbf{x} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}} \end{aligned}$$

The same expressions also apply in $\mathbb{R}^2$, where we replace the plane with a line.

5.5 Dilations

Given scale factors $\alpha, \beta, \gamma > 0$, we define a dilation along the axes by

$$\begin{aligned} \mathbf{e}_1 &\mapsto \alpha\mathbf{e}_1 \\ \mathbf{e}_2 &\mapsto \beta\mathbf{e}_2 \\ \mathbf{e}_3 &\mapsto \gamma\mathbf{e}_3 \end{aligned}$$

5.6 Shears

Let $\mathbf{a}, \mathbf{b}$ be orthogonal unit vectors in $\mathbb{R}^3$, i.e. $|\mathbf{a}| = |\mathbf{b}| = \mathbf{0}$ and $\mathbf{a} \cdot \mathbf{b} = 0$, and we define a real parameter λ. A shear is defined as

$$\begin{aligned} \mathbf{x} &\mapsto \mathbf{x}' = \mathbf{x} + \lambda\mathbf{a}(\mathbf{x} \cdot \mathbf{b}) \\ \mathbf{a} &\mapsto \mathbf{a} \\ \mathbf{b} &\mapsto \mathbf{b} + \lambda\mathbf{a} \end{aligned}$$

This definition holds equivalently in $\mathbb{R}^2$.

5.7 Matrices

Consider a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$, with standard bases $\{\mathbf{e}_i\} \in \mathbb{R}^n$, $\{\mathbf{f}_a\}, \in \mathbb{R}^m$, and with $T(\mathbf{x}) = \mathbf{x}'$. Let further

$$\mathbf{x} = \sum_i x_i \mathbf{e}_i = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}; \quad x' = \sum_a x'_a \mathbf{f}_a = \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_m \end{pmatrix}$$

Linearity implies that T is fixed by specifying

$$T(\mathbf{e}_i) = \mathbf{e}'_i = \mathbf{C}_i \in \mathbb{R}^m$$

We take these $\mathbf{C}$ as columns of an $m \times n$ array or matrix M, with rows denoted as $\mathbf{R}_a \in \mathbb{R}^n$.

$$\begin{pmatrix} \uparrow & & \uparrow \\ \mathbf{C}_1 & \cdots & \mathbf{C}_n \\ \downarrow & & \downarrow \end{pmatrix} = M = \begin{pmatrix} \leftarrow & \mathbf{R}_1 & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{R}_m & \rightarrow \end{pmatrix}$$

M has entries $M_{ai} \in \mathbb{R}$, where a labels rows and i labels columns, so

$$(\mathbf{C}_i)_a = M_{ai} = (\mathbf{R}_a)_i$$

The action of T is then given by the matrix M multiplying the vector $\mathbf{x}$ in the following way:

$$\mathbf{x}' = M\mathbf{x}$$

defined by

$$x'_a = M_{ai} x_i$$

or explicitly:

$$\begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_m \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} & \cdots & M_{1n} \\ M_{21} & M_{22} & \cdots & M_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ M_{m1} & M_{m2} & \cdots & M_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} M_{11}x_1 + M_{12}x_2 + \cdots + M_{1n}x_n \\ M_{21}x_1 + M_{22}x_2 + \cdots + M_{2n}x_n \\ \vdots \\ M_{m1}x_1 + M_{m2}x_2 + \cdots + M_{mn}x_n \end{pmatrix}$$

To check that the matrix multiplication above gives the action of T, we can plug in a generic value $\mathbf{x}$, and we get

$$\mathbf{x}' = T\left(\sum_i x_i \mathbf{e}_i\right) = \sum_i x_i T(\mathbf{e}_i) = \sum_i x_i \mathbf{C}_i$$

and by taking component a of the vector, we have

$$x'_a = \sum_i x_i (\mathbf{C}_i)_a = \sum_i x_i M_{ai}$$

as required. Note also that

$$x'_a = M_{ai} x_i = (\mathbf{R}_a)_i x_i = \mathbf{R}_a \cdot \mathbf{x}$$

We can now regard the properties of T as properties of M (suitably interpreted). For example:

- $\mathrm{Im}(T) = \mathrm{Im}(M) = \mathrm{span}\{\mathbf{C}_1, \cdots, \mathbf{C}_n\}$. In words, the image of a matrix is the span of its columns.
- $\ker(T) = \ker(M) = \{\mathbf{x} : \forall a, \mathbf{R}_a \cdot \mathbf{x} = 0\}$. In some sense, the kernel of M is the subspace perpendicular to all of its rows.

Example. (i) The zero map $\mathbb{R}^n \to \mathbb{R}^m$ corresponds to the zero matrix

$$M = 0 \text{ with } M_{ai} = 0$$

(ii) The identity map $\mathbb{R}^n \to \mathbb{R}^n$ corresponds to the identity (or unit) matrix

$$M = I \text{ with } I_{ij} = \delta_{ij}$$

(iii) The map $\mathbb{R}^3 \to \mathbb{R}^3$ given by $\mathbf{x}' = T(\mathbf{x}) = M\mathbf{x}$ with

$$M = \begin{pmatrix} 3 & 1 & 5 \\ -1 & 0 & -2 \\ 2 & 1 & 3 \end{pmatrix}$$

gives

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = \begin{pmatrix} 3x_1 + x_2 + 5x_3 \\ -x_1 - 2x_3 \\ 2x_1 + x_2 + 3x_3 \end{pmatrix}$$

In this case, we may read off the column vectors $\mathbf{C}_a$ from the matrix. Note that since they form a linearly dependent set, we have

$$\mathrm{Im}(T) = \mathrm{Im}(M) = \mathrm{span}\{\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3\} = \mathrm{span}\{\mathbf{C}_1, \mathbf{C}_2\}$$

Here, $\mathbf{R}_2 \times \mathbf{R}_3 = \begin{pmatrix} 2 & -1 & -1 \end{pmatrix}^\intercal = \mathbf{u}$ is actually perpendicular to all rows as they form a linearly dependent set. So

$$\ker(T) = \ker(M) = \{\lambda \mathbf{u}\}$$

(iv) A rotation through θ in $\mathbb{R}^2$ is given by (building from the images of the basis vectors):

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

(v) A dilation $\mathbf{x}' = M\mathbf{x}$ with scale factors α, β, γ along axes in $\mathbb{R}^3$ is given by

$$\begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix}$$

(vi) A reflection in a plane perpendicular to a unit vector $\hat{\mathbf{n}}$ is given by a matrix H that must have the property that

$$\mathbf{x}' = H\mathbf{x} = \mathbf{x} - 2(\mathbf{x} - \hat{\mathbf{n}})\hat{\mathbf{n}}$$
$$x_i' = x_i - 2x_j n_j n_i = H_{ij} x_j$$

And by comparing coefficients of x_j, and using δ to rewrite x_i using the j index, we have

$$H_{ij} = \delta_{ij} - 2n_i n_j$$

For example, with $\hat{\mathbf{n}} = \frac{1}{\sqrt{3}}\begin{pmatrix}1 & 1 & 1\end{pmatrix}$, then $n_i n_j = \frac{1}{3}$ for all i, j, so

$$H = \frac{1}{3}\begin{pmatrix}1 & -2 & -2\\ -2 & 1 & -2\\ -2 & -2 & 1\end{pmatrix}$$

(vii) A shear is defined by a matrix S such that

$$\mathbf{x}' = S\mathbf{x} = \mathbf{x} + \lambda(\mathbf{b}\cdot\mathbf{x})\mathbf{a}$$

where $\mathbf{a}$, $\mathbf{b}$ are unit vectors with $\mathbf{a} \perp \mathbf{b}$, and where λ is a real scale factor. Therefore:

$$\begin{aligned} x_i' &= x_i + \lambda b_j x_j a_i = S_{ij}x_j \\ \therefore S_{ij} &= \delta_{ij} + \lambda a_i b_j \end{aligned}$$

For example in $\mathbb{R}^2$ with $\mathbf{a} = \begin{pmatrix}1\\0\end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix}0\\1\end{pmatrix}$, we have

$$S = \begin{pmatrix}1 & \lambda\\ 0 & 1\end{pmatrix}$$

(viii) A rotation matrix R in $\mathbb{R}^3$ with axis $\hat{\mathbf{n}}$ and angle θ must satisfy

$$\begin{aligned} \mathbf{x}' &= R\mathbf{x} = (\cos\theta)\mathbf{x} + (1-\cos\theta)(\hat{\mathbf{n}}\cdot\mathbf{x})\hat{\mathbf{n}} + (\sin\theta)(\hat{\mathbf{n}}\times\mathbf{x}) \\ x_i' &= (\cos\theta)x_i + (1-\cos\theta)n_j x_j n_i - (\sin\theta)\varepsilon_{ijk}x_j n_k = R_{ij}x_j \\ \therefore R_{ij} &= \delta_{ij}(\cos\theta) - (1-\cos\theta)n_i n_j - (\sin\theta)\varepsilon_{ijk}n_k \end{aligned}$$

5.8 Matrix of a general linear map

Consider a linear map $T : V \to W$ between general real or complex vector spaces of dimension n, m respectively. We will choose bases $\{\mathbf{e}_i\}$ for V and $\{\mathbf{f}_a\}$ for W. The matrix representing the linear map T with respect to these bases is an $m \times n$ array with entries $M_{ai} \in \mathbb{R}$ or $\mathbb{C}$ as appropriate, defined by

$$T(\mathbf{e}_i) = \sum_a \mathbf{f}_a M_{ai}$$

Then

$$\mathbf{x}' = T(\mathbf{x}) \iff x_a' = \sum_i M_{ai}x_i = M_{ai}x_i$$

where

$$\mathbf{x} = \sum_i x_i\mathbf{e}_i; \quad \mathbf{x}' = \sum_a x_a\mathbf{f}_a$$

Note therefore that (in real vector spaces) given choices of bases $\{\mathbf{e}_i\}$ and $\{\mathbf{f}_a\}$, V is identified with $\mathbb{R}_n$ in the sense that any vector has n real components, and that W is identified with R_m analogously, and that therefore T is identified with an $m \times n$ real matrix M. Note further that entries in column i of M are components of $T(\mathbf{e}_i)$ with respect to basis $\{\mathbf{f}_a\}$.

5.9 Linear combinations

If $T : V \to W$ and $S : V \to W$, between real or complex vector spaces V, W of dimension n, m respectively, are linear, then

$$\alpha T + \beta S : V \to W$$

is also a linear map, where

$$(\alpha T + \beta S)(\mathbf{x}) = \alpha T(\mathbf{x}) + \beta S(\mathbf{x})$$

for any $\mathbf{x} \in V$. So the set of linear maps is a vector space. If M and N are the $m \times N$ matrices for T, S then $\alpha M + \beta N$ is the $m \times n$ matrix for the linear combination above, where

$$(\alpha M + \beta N)_{ai} + \alpha M_{ai} + \beta N_{ai}; \quad a = 1, \cdots, m; \quad i = 1, \cdots, n$$

with respect to the same bases.

5.10 Matrix multiplication

If A is an $m \times n$ matrix with entries A_{ai}, and B is an $n \times p$ matrix with entries B_{ir}, then we define AB to be an $m \times p$ matrix with entries

$$(AB)_{ar} = A_{ai}B_{ir}; \quad a = 1, \cdots, m; \quad i = 1, \cdots, n; \quad r = 1, \cdots, p$$

The product is not defined unless the amount of columns of A matches the number of rows of B.

Matrix multiplication corresponds to composition of linear maps. Consider linear maps:

$$\begin{aligned} S : \mathbb{R}^p \to \mathbb{R}^n;\ S(\mathbf{x}) = B\mathbf{x},\ \mathbf{x} \in \mathbb{R}^p \\ T : \mathbb{R}^n \to \mathbb{R}^m;\ T(\mathbf{x}) = A\mathbf{x},\ \mathbf{x} \in \mathbb{R}^n \\ \implies T \circ S : \mathbb{R}^p \to \mathbb{R}^m;\ (T \circ S)(\mathbf{x}) = (AB)x \end{aligned}$$

since

$$[(AB)\mathbf{x}]_a = (AB)_{ar}x_r$$

and

$$A(B(\mathbf{x})) = A_{ai}(B\mathbf{x})_i = A_{ai}B_{ir}x_r = (AB)_{ar}x_r$$

as required. The definition of matrix multiplication ensures that these answers agree. Of course, this proof works for complex or general vector spaces.

Whenever the products are defined, then for any scalars λ and μ:

- $(\lambda M + \mu N)P = \lambda MP + \mu NP$
- $P(\lambda M + \mu N) = \lambda PM + \mu PN$
- $(MN)P = M(NP)$
- $IM = MI = M$ where $I_{ij} = \delta_{ij}$

We may view matrix multiplication in the following ways.

(i) Regarding a vector $\mathbf{x} \in \mathbb{R}^n$ as a column vector (an $n \times 1$ matrix), then the matrix-vector and matrix-matrix multiplication rules agree.

(ii) Consider the product AB where A is an $m \times n$ matrix and B is an $n \times p$, with columns $\mathbf{C}_r(B) \in \mathbb{R}^n$ and columns $\mathbf{C}_r(AB) \in \mathbb{R}^m$, where $1 \le r \le p$. The columns are related by $\mathbf{C}_r(AB) = A\mathbf{C}_r(B)$. Less formally, each column in the right matrix is acted on by the left matrix as if it were a vector, then the resultant vectors are combined into the output matrix.

(iii) In terms of rows and columns,

$$AB = \begin{pmatrix} & \vdots & \\ \leftarrow & \mathbf{R}_n(A) & \rightarrow \\ & \vdots & \end{pmatrix} \begin{pmatrix} & \uparrow & \\ \cdots & \mathbf{C}_r(B) & \cdots \\ & \downarrow & \end{pmatrix}$$

gives

$$\begin{aligned} (AB)_{ar} &= [\mathbf{R}_a(A)]_i \, [\mathbf{C}_r(B)]_i \\ &= \mathbf{R}_a(A) \cdot \mathbf{C}_r(B) \text{ for real matrices, where the } \cdot \text{ is the dot product in } R^n \end{aligned}$$

5.11 Matrix inverses

If A is an $m \times n$ then B, an $n \times m$ matrix, is a left inverse of A if $BA = I$ (the $n \times n$ identity matrix). C is a right inverse of A if $AC = I$ (the $m \times m$ identity matrix). If $m = n$ (A is square), then one of these implies the other; there is no distinction between left and right inverses. We say that $B = C = A^{-1}$, *the* inverse of the matrix A, such that $AA^{-1} = A^{-1}A = I$. Not every matrix has an inverse. If such an inverse exists, A is called invertible, or non-singular.

Consider $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^n$ or $\mathbb{C}^n$, and M is an $n \times n$ matrix. If M^{-1} exists, we can solve the equation $\mathbf{x}' = M\mathbf{x}$ for $\mathbf{x}$, given $\mathbf{x}'$, because we can apply the matrix inverse on the left. For example, where $n = 2$, we have

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$$

and

$$\begin{aligned} x_1' &= M_{11}x_1 + M_{12}x_2 \\ x_2' &= M_{21}x_1 + M_{22}x_2 \end{aligned}$$

We can solve these simultaneous equations to construct the general matrix inverse.

$$\begin{aligned} M_{22}x_1' - M_{12}x_2' &= (\det M)x_1 \\ -M_{21}x_1' + M_{11}x_2' &= (\det M)x_2 \end{aligned}$$

where $\det M = M_{11}M_{22} - M_{12}M_{21}$, called the determinant of the matrix. Where the determinant is nonzero, the matrix inverse

$$M^{-1} = \frac{1}{\det M} \begin{pmatrix} M_{22} & -M_{12} \\ -M_{21} & M_{11} \end{pmatrix}$$

exists. Note that

$$\begin{aligned} \mathbf{C}_1 &= M\mathbf{e}_1 = \begin{pmatrix} M_{11} \\ M_{21} \end{pmatrix} \\ \mathbf{C}_2 &= M\mathbf{e}_2 = \begin{pmatrix} M_{12} \\ M_{22} \end{pmatrix} \\ \iff \det M &= [\mathbf{C}_1, \mathbf{C}_2] = [M\mathbf{e}_1, M\mathbf{e}_2] \text{ in } \mathbb{R}^2 \end{aligned}$$

So the determinant gives the signed factor by which areas are scaled under the action of M. $\det M$ is nonzero if and only if $M\mathbf{e}_1$ and $M\mathbf{e}_2$ are linearly independent, which is true if and only if the image of M has dimension 2, i.e. M has maximal rank. For example, a shear

$$S(\lambda) = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}$$

has determinant 1, so areas are preserved. In particular, in this case,

$$S^{-1}(\lambda) = \begin{pmatrix} 1 & -\lambda \\ 0 & 1 \end{pmatrix} = S(-\lambda)$$

As another example, we know that a matrix $R(\theta)$ for a rotation about a fixed axis $\hat{\mathbf{n}}$ through angle θ has formula

$$R(\theta)_{ij}R(-\theta)_{jk} = (\delta_{ij}\cos\theta + (1-\cos\theta)n_in_j - \varepsilon_{ijp}n_p\sin\theta) \times (\delta_{jk}\cos\theta + (1-\cos\theta)n_jn_k + \varepsilon_{jkq}n_q\sin\theta)$$

Expanding out, noting that $n_in_i = 1$ as $\hat{\mathbf{n}}$ is a unit vector, and cancelling:

$$= \delta_{ik}\cos^2\theta + 2\cos\theta(1-\cos\theta)n_in_k + (1-\cos\theta)^2 n_in_k - \varepsilon_{ijp}\varepsilon_{jkq}n_pn_q\sin^2\theta$$

By using an $\varepsilon\varepsilon$ identity:

$$\begin{aligned}
&= \delta_{ik}\cos^2\theta + (1-\cos^2\theta)n_in_k + \delta_{ik}n_pn_p\sin^2\theta - (\sin^2\theta)n_in_k \\
&= \delta_{ik}\cos^2\theta + \delta_{ik}n_pn_p\sin^2\theta \\
&= \delta_{ik}\cos^2\theta + \delta_{ik}\sin^2\theta \\
&= \delta_{ik}
\end{aligned}$$

as required.

6 Transpose and Hermitian conjugate

6.1 Transpose

If M is an $m \times n$ (real or complex) matrix, the transpose $M^\intercal$ is an $n \times m$ matrix defined by

$$(M^\intercal)_{ia} = M_{ai}$$

which essentially exchanges rows and columns. Here are some key properties.

- $(\alpha A + \beta B)^\intercal = \alpha A^\intercal + \beta B^\intercal$ for α, β scalars, and A, B both $m \times n$ matrices.
- $(AB)^\intercal = B^\intercal A^\intercal$, where A is $m \times n$ and B is $n \times p$. This is because

$$\begin{aligned}
[(AB)^\intercal]_{ra} &= (AB)_{ar} \\
&= A_{ai}B_{ir} \\
&= (A^\intercal)_{ia}(B^\intercal)_{ri} \\
&= (B^\intercal)_{ri}(A^\intercal)_{ia} \\
&= (B^\intercal A^\intercal)_{ra}
\end{aligned}$$

- If $\mathbf{x}$ is a column vector (or an $n \times 1$ matrix), $\mathbf{x}^\intercal$ is the equivalent row vector (a $1 \times n$ matrix).

- The inner product in $\mathbb{R}^n$ can therefore be written $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^\intercal \mathbf{y}$. Note that this is not equivalent to $\mathbf{x}\mathbf{y}^\intercal$, which is known as the outer product, which results in a matrix not a scalar.
- If M is $n \times n$ (square) then M is:
  - symmetric iff $M^\intercal = M$, or $M_{ij} = M_{ji}$
  - antisymmetric iff $M^\intercal = -M$, or $M_{ij} = -M_{ji}$
- Any M which is square can be written as a sum of a symmetric and and an antisymmetric part

  $$M = S + A \quad \text{where } S = \frac{1}{2}(M + M^\intercal); \quad A = \frac{1}{2}(M - M^\intercal)$$

  as S is symmetric and A is antisymmetric by construction.
- If A is 3×3 and antisymmetric, then we can write

  $$A_{ij} = \varepsilon_{ijk} a_k \text{ where } A = \begin{pmatrix} 0 & a_3 & -a_2 \\ -a_3 & 0 & a_1 \\ a_2 & -a_1 & 0 \end{pmatrix}$$

  Then, we have

  $$(A\mathbf{x})_i = \varepsilon_{ijk} a_k x_j = (\mathbf{x} \times \mathbf{a})_i$$

6.2 Hermitian conjugate

Let M be an $m \times n$ matrix. Then the Hermitian conjugate (also known as the conjugate transpose) $M^\dagger$ is an $n \times m$ matrix defined by

$$(M^\dagger)_{ia} = \overline{M_{ai}}$$

If M is square, then M is Hermitian if and only if $M^\dagger = M$, or alternatively $M_{ia} = \overline{M_{ai}}$; M is anti-Hermitian if $M^\dagger = -M$, or alternatively $M_{ia} = -\overline{M_{ai}}$. Similarly to above, if $\mathbf{z}$ is a column vector in $\mathbb{C}^n$ (an $n \times 1$ matrix), then the complex inner product is given by $\mathbf{z} \cdot \mathbf{w} = \mathbf{z}^\dagger \mathbf{w}$.

6.3 Trace

For a complex $n \times n$ (square) matrix M, the trace of the matrix, denoted $\operatorname{tr}(M)$, is defined by

$$\operatorname{tr}(M) = M_{ii} = M_{11} + M_{22} + \cdots + M_{nn}$$

It has a number of key properties.

- $\operatorname{tr}(\alpha M + \beta N) = \alpha \operatorname{tr} M + \beta \operatorname{tr} N$ where α and β are scalars, and M and N are $n \times n$ matrices.
- $\operatorname{tr}(MN) = \operatorname{tr}(NM)$ where M is $m \times n$ and N is $n \times m$. MN and NM need not have the same dimension, but their traces are identical. We can check this as follows: $\operatorname{tr}(MN) = (MN)_{aa} = M_{ai} N_{ia} = N_{ia} M_{ai} = (NM)_{ii} = \operatorname{tr}(NM)$.
- $\operatorname{tr}(M^\intercal) = \operatorname{tr}(M)$
- $\operatorname{tr}(I) = \delta_{ii} = n$ where n is the dimensionality of the vector space.

- If S is $n \times n$ and symmetric, let

$$
\begin{aligned}
T &= S - \frac{1}{n}\operatorname{tr}(S)I \\
\text{or } T_{ij} &= S_{ij} - \frac{1}{n}\operatorname{tr}(S)\delta_{ij} \\
\text{then } \operatorname{tr}(T) = T_{ii} &= S_{ii} = \frac{1}{n}\operatorname{tr}(S)\delta_{ii} \\
&= \operatorname{tr}(S) - \frac{1}{n}\operatorname{tr}(S) = 0
\end{aligned}
$$

  Then $S = T + \frac{1}{n}\operatorname{tr}(S)I$ where T is traceless and the right hand term $\frac{1}{n}\operatorname{tr}(S)I$ is 'pure trace'.

- If A is $n \times n$ antisymmetric, $\operatorname{tr}(A) = A_{ii} = 0$.

6.4 Orthogonal matrices

A real $n \times n$ matrix U is orthogonal if and only if its transpose is its inverse.

$$U^\intercal U = UU^\intercal = I$$

These conditions can be written

$$U_{ki}U_{kj} = U_{ik}U_{jk} = \delta_{ij}$$

In words, the left hand side says that the columns of U are orthonormal, and the middle part of the equation says that the rows of U are orthonormal.

$$U^\intercal U = \begin{pmatrix} & \vdots & \\ \leftarrow & \mathbf{C}_i & \rightarrow \\ & \vdots & \end{pmatrix}\begin{pmatrix} & \uparrow & \\ \cdots & \mathbf{C}_j & \cdots \\ & \downarrow & \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}$$

For example, if $U = R(\theta)$ is a rotation through θ around an axis $\hat{\mathbf{n}}$, then $U^\intercal = R(\theta)^\intercal = R(-\theta) = R(\theta)^{-1} = U^{-1}$. An equivalent definition for orthogonality is: U is orthogonal if and only if it preserves the inner product on $\mathbb{R}^n$.

$$(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y} \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

To check equivalence:

$$
\begin{aligned}
(U\mathbf{x}) \cdot (U\mathbf{y}) &= (U\mathbf{x})^\intercal (U\mathbf{y}) \\
&= (\mathbf{x}^\intercal U^\intercal)(U\mathbf{y}) \\
&= \mathbf{x}^\intercal (U^\intercal U)\mathbf{y} \\
&= \mathbf{x}^\intercal \mathbf{y} \\
&= \mathbf{x} \cdot \mathbf{y}
\end{aligned}
$$

which is true if and only if $U^\intercal U = I$. Note that in $\mathbb{R}^n$, the columns of U are $U\mathbf{e}_i, \cdots, U\mathbf{e}_n$ so the inner product is preserved when U acts on the standard basis vectors if and only if

$$(U\mathbf{e}_i) \cdot (U\mathbf{e}_j) = \mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$$

i.e. the columns of U are orthonormal.

Let us now try to find a general 2×2 orthogonal matrix. We begin by transforming the basis vectors. $\mathbf{e}_i = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ must be transformed to a unit vector. Therefore, in the most general sense:

$$U \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$$

for some parameter θ. Now, the other basis vector $\mathbf{e}_2$ must be orthogonal to it, and so it must be

$$U \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \pm \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$$

So we have two cases:

$$U = R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}; \quad U = H = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$$

where R is a rotation by θ and H is a reflection in $\mathbb{R}^2$, where

$$\hat{\mathbf{n}} = \begin{pmatrix} -\sin\frac{\theta}{2} \\ \cos\frac{\theta}{2} \end{pmatrix}$$

because

$$H_{ij} = \delta_{ij} - 2n_i n_j \therefore H = \begin{pmatrix} 1 - 2\sin^2\frac{\theta}{2} & 2\sin\frac{\theta}{2}\cos\frac{\theta}{2} \\ 2\sin\frac{\theta}{2}\cos\frac{\theta}{2} & 1 - 2\cos^2\frac{\theta}{2} \end{pmatrix}$$

which simplifies as required. Note that $\det R = +1$, but $\det H = -1$.

6.5 Unitary matrices

A complex $n \times n$ matrix U is called unitary if and only if

$$U^\dagger U = UU^\dagger = I$$

Equivalently, U is unitary if and only if it preserves the complex inner product on $\mathbb{C}_n$:

$$\langle U\mathbf{z}, U\mathbf{w} \rangle = \langle \mathbf{z}, \mathbf{w} \rangle \quad \forall \mathbf{z}, \mathbf{w} \in \mathbb{C}^n$$

To check equivalence:

$$\begin{aligned} \langle U\mathbf{z}, U\mathbf{w} \rangle &= (U\mathbf{z})^\dagger (U\mathbf{w}) \\ &= (\mathbf{z}^\dagger U^\dagger)(U\mathbf{w}) \\ &= \mathbf{z}^\dagger (U^\dagger U)\mathbf{w} \\ &= \mathbf{z}^\dagger \mathbf{w} \end{aligned}$$

which of course matches if and only if $U^\dagger U = I$.

7 Adjugates and alternating forms

7.1 Inverses in two dimensions

Consider a linear map $T : \mathbb{R}^n \to \mathbb{R}^n$. If T is invertible (i.e. bijective), then $\ker T = \{\mathbf{0}\}$ as T is injective, and $\operatorname{Im} T = \mathbb{R}^n$ as T is surjective. These conditions are actually equivalent due to the rank-nullity theorem. Conversely, if the conditions hold, then $T(\mathbf{e}_1), T(\mathbf{e}_2), \cdots, T(\mathbf{e}_n)$ must be a basis of the image, so we can just define T^{-1} by defining its actions on the basis vectors $T(\mathbf{e}_1), T(\mathbf{e}_2) \cdots T(\mathbf{e}_n)$, specifically mapping them to the standard basis.

How can we test whether the conditions above hold for a matrix M representing T, and how can we find M^{-1} from M explicitly? For any $n \times n$ matrix M (not necessarily invertible), we will define the adjugate matrix $\widetilde{M}$ and the determinant $\det M$ such that

$$\widetilde{M}M = (\det M)I \tag{$*$}$$

Then if $\det M \neq 0$, M is invertible, where

$$M^{-1} = \frac{1}{\det M}\widetilde{M}$$

From $n = 2$, recall that $(*)$ holds with

$$M = \begin{pmatrix} M_{11} & M_{21} \\ M_{12} & M_{22} \end{pmatrix}; \quad \widetilde{M} = \begin{pmatrix} M_{22} & -M_{21} \\ -M_{12} & M_{11} \end{pmatrix}; \quad \det M = [M\mathbf{e}_1, M\mathbf{e}_2] = \varepsilon_{ij} M_{i1} M_{j2}$$

The determinant in this case is the factor by which areas scale under M. $\det M \neq 0$ if and only if $M\mathbf{e}_1, M\mathbf{e}_2$ are linearly independent.

7.2 Three dimensions

For $n = 3$, we will define similarly

$$\det M = [M\mathbf{e}_1, M\mathbf{e}_2, M\mathbf{e}_3] = \varepsilon_{ijk} M_{i1} M_{j2} M_{k3}$$

We define it like this because this is the factor by which volumes scale under M in three dimensions. So

$$\det M \neq 0 \iff \{M\mathbf{e}_1, M\mathbf{e}_2, M\mathbf{e}_3\} \text{ linearly independent, or } \operatorname{Im} M = \mathbb{R}^3$$

Now we define $\widetilde{M}$ from M using row/column notation.

$$\begin{aligned}
\mathbf{R}_1(\widetilde{M}) &= \mathbf{C}_2(M) \times \mathbf{C}_3(M) \\
\mathbf{R}_2(\widetilde{M}) &= \mathbf{C}_3(M) \times \mathbf{C}_1(M) \\
\mathbf{R}_3(\widetilde{M}) &= \mathbf{C}_1(M) \times \mathbf{C}_2(M)
\end{aligned}$$

Note that therefore,

$$(\widetilde{M}M)_{ij} = \mathbf{R}_i(\widetilde{M}) \cdot \mathbf{C}_j(M) = \underbrace{(\mathbf{C}_1(M) \times \mathbf{C}_2(M) \cdot \mathbf{C}_3(M))}_{\det M} \delta_{ij}$$

as claimed. For example, let us invert the following matrix.

$$
\begin{aligned}
M &= \begin{pmatrix} 1 & 3 & 0 \\ 0 & -1 & -2 \\ 4 & 1 & -1 \end{pmatrix} \\
\mathbf{C}_2 \times \mathbf{C}_3 &= \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix} \times \begin{pmatrix} 0 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ 6 \end{pmatrix} \\
\mathbf{C}_3 \times \mathbf{C}_1 &= \begin{pmatrix} 0 \\ 2 \\ -1 \end{pmatrix} \times \begin{pmatrix} 1 \\ 0 \\ 4 \end{pmatrix} = \begin{pmatrix} 8 \\ -1 \\ -2 \end{pmatrix} \\
\mathbf{C}_1 \times \mathbf{C}_2 &= \begin{pmatrix} 1 \\ 0 \\ 4 \end{pmatrix} \times \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 11 \\ -1 \end{pmatrix} \\
\widetilde{M} &= \begin{pmatrix} -1 & 3 & 6 \\ 8 & -1 & -2 \\ 4 & 11 & -1 \end{pmatrix} \\
\det M &= \mathbf{C}_1 \cdot \mathbf{C}_2 \times \mathbf{C}_3 = 23 \\
\widetilde{M}M &= 23I
\end{aligned}
$$

7.3 Levi-Civita ε in higher dimensions

Recall (from IA Groups):

- A permutation σ on the set $\{1, 2, \cdots, n\}$ is a bijection from the set to itself, specified by an ordered list $\sigma(1), \sigma(2), \cdots, \sigma(n)$.
- Permutations form a group S_n, called the symmetric group of order $n!$
- A transposition $\tau = (p, q)$ where $p \neq q$ is a permutation that swaps p and q.
- Any permutation is a product of of k transpositions, where k is unique modulo 2 for a given σ. In this course, we will write $\varepsilon(\sigma)$ to mean the sign (or signature) of the permutation, $(-1)^k$. σ is even if the sign is 1, and odd if the sign is -1.

The alternating symbol ε in $\mathbb{R}^n$ or $\mathbb{C}^n$ is an n-index object (tensor) defined by

$$
\varepsilon_{\underbrace{ij\dots l}_{n \text{ indices}}} = \begin{cases} +1 & \text{if } i, j \cdots, l \text{ is an even permutation of } 1, 2, \cdots, n \\ -1 & \text{if } i, j \cdots, l \text{ is an odd permutation of } 1, 2, \cdots, n \\ 0 & \text{otherwise, i.e. if any indices take the same value} \end{cases}
$$

Thus if σ is any permutation, then

$$
\varepsilon_{\sigma(1)\cdots\sigma(n)} = \varepsilon(\sigma)
$$

So $\varepsilon_{ij\dots l}$ is totally antisymmetric and changes sign whenever a pair of indices are exchanged.

Definition. Given vectors $\mathbf{v}_1, \cdots \mathbf{v}_n \in \mathbb{R}^n$ or $\mathbb{C}^n$, the alternating form combines them to give the scalar

$$
\begin{aligned}
[\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n] &= \varepsilon_{ij\dots l}(\mathbf{v}_1)_i(\mathbf{v}_2)_j \cdots (\mathbf{v}_n)_l \\
&= \sum_{\sigma \in S_n} \varepsilon(\sigma) \cdot (\mathbf{v}_1)_{\sigma(1)} \cdot (\mathbf{v}_2)_{\sigma(2)} \cdots (\mathbf{v}_n)_{\sigma(n)}
\end{aligned}
$$

7.4 Properties

(i) The alternating form is multilinear.

$$
\begin{aligned}
[\mathbf{v}_1, \cdots, \mathbf{v}_{p-1}, \alpha\mathbf{u} + \beta\mathbf{w}, \mathbf{v}_{p+1} \cdots, \mathbf{v}_n] &= \alpha[\mathbf{v}_1, \cdots, \mathbf{v}_{p-1}, \mathbf{u}, \mathbf{v}_{p+1} \cdots, \mathbf{v}_n] \\
&+ \beta[\mathbf{v}_1, \cdots, \mathbf{v}_{p-1}, \mathbf{w}, \mathbf{v}_{p+1} \cdots, \mathbf{v}_n]
\end{aligned}
$$

(ii) It is totally antisymmetric. $[\mathbf{v}_{\sigma(1)}, \mathbf{v}_{\sigma(2)}, \cdots, \mathbf{v}_{\sigma(n)}] = \varepsilon(\sigma)[\mathbf{v}_1, \cdots, \mathbf{v}_n]$

(iii) Standard basis vectors give a positive result: $[\mathbf{e}_i, \cdots, \mathbf{e}_n] = 1$.

These three properties fix the alternating form completely, and they also imply

(iv) If $\mathbf{v}_p = \mathbf{v}_q$ where $p \neq q$, then

$$
[\mathbf{v}_1, \cdots, \mathbf{v}_p, \cdots, \mathbf{v}_q, \cdots, \mathbf{v}_n] = 0
$$

(v) If $\mathbf{v}_p$ can be written as a non-trivial linear combination of the other vectors, then

$$
[\mathbf{v}_1, \cdots, \mathbf{v}_p, \cdots, \mathbf{v}_n] = 0
$$

Property (iv) follows from property (ii), where we swap $\mathbf{v}_p$ and $\mathbf{v}_q$. Property (v) follows from substituting the linear combination representation of $\mathbf{v}_p$ into the alternating form expression, the using properties (i) and (iv). To justify (ii) above, it suffices to check a transposition $\tau = (p\ q)$ where (without loss of generality) $p < q$, then since transpositions generate all permutations the result follows.

$$
\begin{aligned}
&[\mathbf{v}_1, \cdots, \mathbf{v}_{p-1}, \mathbf{v}_q, \mathbf{v}_{p+1}, \cdots, \mathbf{v}_{q-1}, \mathbf{v}_p, \mathbf{v}_{q+1}, \cdots, \mathbf{v}_n] \\
&= \sum_\sigma \varepsilon(\sigma)(\mathbf{v}_1)_{\sigma(1)} \cdots (\mathbf{v}_{p-1})_{\sigma(p-1)}(\mathbf{v}_q)_{\sigma(p)}(\mathbf{v}_{p+1})_{\sigma(p+1)} \\
&\qquad \cdots (\mathbf{v}_{q-1})_{\sigma(q-1)}(\mathbf{v}_p)_{\sigma(q)}(\mathbf{v}_{q+1})_{\sigma(q+1)} \\
&= \sum_\sigma \varepsilon(\sigma)(\mathbf{v}_1)_{\sigma'(1)} \cdots (\mathbf{v}_{p-1})_{\sigma'(p-1)}(\mathbf{v}_q)_{\sigma'(q)}(\mathbf{v}_{p+1})_{\sigma'(p+1)} \\
&\qquad \cdots (\mathbf{v}_{q-1})_{\sigma'(q-1)}(\mathbf{v}_p)_{\sigma'(p)}(\mathbf{v}_{q+1})_{\sigma'(q+1)}
\end{aligned}
$$

where $\sigma' = \sigma\tau$

$$
\begin{aligned}
&= -\sum_{\sigma'} \varepsilon(\sigma')(\mathbf{v}_1)_{\sigma'(1)} \cdots (\mathbf{v}_{p-1})_{\sigma'(p-1)}(\mathbf{v}_p)_{\sigma'(p)}(\mathbf{v}_{p+1})_{\sigma'(p+1)} \\
&\qquad \cdots (\mathbf{v}_{q-1})_{\sigma'(q-1)}(\mathbf{v}_q)_{\sigma'(q)}(\mathbf{v}_{q+1})_{\sigma'(q+1)} \\
&= -[\mathbf{v}_1, \cdots, \mathbf{v}_{p-1}, \mathbf{v}_p, \mathbf{v}_{p+1}, \cdots, \mathbf{v}_{q-1}, \mathbf{v}_q, \mathbf{v}_{q+1}, \cdots, \mathbf{v}_n]
\end{aligned}
$$

as required.

Proposition. $[\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n] \neq 0$ if and only if $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ are linearly independent.

Proof. To show the forward implication, let us suppose that they are not linearly independent and use property (v). Then we can express some $\mathbf{v}_p$ as a linear combination of the others. Then $[\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n] = 0$.

To show the other direction, note that $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_3$ means that they span, and if they span then each of the standard basis vectors $\mathbf{e}_i$ can be written as a linear combination of the $\mathbf{v}$ vectors, i.e. $\mathbf{e}_i = U_{ai}\mathbf{v}_a$. Then

$$\begin{aligned}
[\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n] &= [U_{a1}\mathbf{v}_a, U_{b2}\mathbf{v}_b, \cdots, U_{cn}\mathbf{v}_c] \\
&= U_{a1}U_{b2}\cdots U_{cn}[\mathbf{v}_a, \mathbf{v}_b, \cdots, \mathbf{v}_c] \\
&= U_{a1}U_{b2}\cdots U_{cn}\varepsilon_{ab\cdots c}[\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n]
\end{aligned}$$

By definition, the left hand side is $+1$, so $[\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n]$ is nonzero. □

As an example of these ideas, let

$$\mathbf{v}_1 = \begin{pmatrix} i \\ 0 \\ 0 \\ 2 \end{pmatrix}; \quad \mathbf{v}_2 = \begin{pmatrix} 0 \\ 0 \\ 5i \\ 0 \end{pmatrix}; \quad \mathbf{v}_3 = \begin{pmatrix} 3 \\ 2i \\ 0 \\ 0 \end{pmatrix}; \quad \mathbf{v}_4 = \begin{pmatrix} 0 \\ 0 \\ i \\ 1 \end{pmatrix}; \quad \text{where } \mathbf{v}_j \in \mathbb{C}_4$$

Then

$$\begin{aligned}
[\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4] &= 5i[\mathbf{v}_1, \mathbf{e}_3, \mathbf{v}_3, \mathbf{v}_4] \\
&= 5i[i\mathbf{e}_1 + 2\mathbf{e}_4, \mathbf{e}_3, 3\mathbf{e}_1 + 2i\mathbf{e}_2, -i\mathbf{e}_3 + \mathbf{e}_4]
\end{aligned}$$

By multilinearity, we can eliminate all $\mathbf{e}_3$ terms not in the second position because they will cancel with it, giving

$$= 5i[i\mathbf{e}_1 + 2\mathbf{e}_4, \mathbf{e}_3, 3\mathbf{e}_1 + 2i\mathbf{e}_2, \mathbf{e}_4]$$

And likewise with $\mathbf{e}_4$:

$$= 5i[i\mathbf{e}_1, \mathbf{e}_3, 3\mathbf{e}_1 + 2i\mathbf{e}_2, \mathbf{e}_4]$$

And again with $\mathbf{e}_1$:

$$\begin{aligned}
&= 5i[i\mathbf{e}_1, \mathbf{e}_3, 2i\mathbf{e}_2, \mathbf{e}_4] \\
&= 5i \cdot 2i \cdot i[\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_2, \mathbf{e}_4] \\
&= 10i[\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4] \\
&= 10i
\end{aligned}$$

8 Determinant

8.1 Definition

For an $n \times n$ matrix M with columns $\mathbf{C}_a = M\mathbf{e}_a$, then the determinant $\det(M) = |M| \in \mathbb{R}$ or $\mathbb{C}$ is given by any of the following equivalent definitions.

$$\begin{aligned}
\det M &= [\mathbf{C}_1, \mathbf{C}_2, \cdots, \mathbf{C}_n] \\
&= [M\mathbf{e}_1, M\mathbf{e}_2, \cdots, M\mathbf{e}_n] \\
&= \varepsilon_{ij\dots l} M_{i1} M_{j2} \cdots M_{ln} \\
&= \sum_\sigma \varepsilon(\sigma) M_{\sigma(1)1} M_{\sigma(2)2} \cdots M_{\sigma(n)n}
\end{aligned}$$

Here are some examples.

(i) $n = 2$

$$\det M = \sum_\sigma M_{\sigma(1)1} M_{\sigma(2)2} = \begin{vmatrix} M_{11} & M_{21} \\ M_{12} & M_{22} \end{vmatrix} = M_{11}M_{22} - M_{12}M_{21}$$

(ii) M diagonal, i.e. $M_{ij} = 0$ for $i \neq j$

$$M = \begin{pmatrix} M_{11} & 0 & \cdots & 0 \\ 0 & M_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_{nn} \end{pmatrix} \implies \det M = M_{11}M_{22}\cdots M_{nn}$$

(iii) Let M be $n \times n$, A be $(n-1) \times (n-1)$, where

$$M = \left(\begin{array}{c|c} A & 0 \\ \hline 0 & 1 \end{array}\right)$$

We call M a matrix 'in block form'. So $M_{ni} = M_{in} = 0$ if $i \neq n$. So we can restrict the permutation σ to only transmuting the first $(n-1)$ terms, i.e. $\sigma(n) = n$. So $\det M = \det A$.

Proposition. If $\mathbf{R}_a$ are the rows of M, $\det M$ is given by

$$\begin{aligned}
\det M &= [\mathbf{R}_1, \mathbf{R}_2, \cdots, \mathbf{R}_n] \\
&= \varepsilon_{ij\dots l} M_{1i} M_{2j} \cdots M_{nl} \\
&= \sum_\sigma \varepsilon(\sigma) M_{1\sigma(1)} M_{2\sigma(2)} \cdots M_{n\sigma(n)}
\end{aligned}$$

i.e. $\det M = \det M^\intercal$.

Proof. Recall that $(\mathbf{C}_a)_i = M_{ia} = (\mathbf{R}_i)_a$. We need to show that one of these definitions is equivalent to one of the previous definitions, then all other equivalent definitions follow. We use the Σ definition by considering the product $M_{1\sigma(1)}M_{2\sigma(2)}\cdots M_{n\sigma(n)}$. We may rewrite this product in a different order: $M_{\rho(1)1}M_{\rho(2)2}\cdots M_{\rho(n)n}$. Then $\rho = \sigma^{-1}$. But then $\varepsilon(\sigma) = \varepsilon(\rho)$, and a sum over σ is equivalent to a sum over ρ. $\square$

8.2 Expanding by rows or columns

For an $n \times n$ matrix M with entries M_{ia}, we define the minor M^{ia} to be the $(n-1)\times(n-1)$ determinant of the matrix obtained by deleting row i and column a from M.

Proposition. The determinant of a generic $n \times n$ matrix M is given by

$$\begin{aligned} \det M &= \sum_i (-1)^{i+a} M_{ia} M^{ia} \text{ for a fixed } a \\ &= \sum_a (-1)^{i+a} M_{ia} M^{ia} \text{ for a fixed } i \end{aligned}$$

This process is known as expanding by row i or by column a. As an example, let us take the following 4×4 complex matrix

$$M = \begin{pmatrix} i & 0 & 3 & 0 \\ 0 & 0 & 2i & 0 \\ 0 & 5i & 0 & -i \\ 2 & 0 & 0 & 1 \end{pmatrix}$$

Then, the determinant is given by (expanding by row 3)

$$\begin{aligned} \det M &= -5i \begin{vmatrix} i & 3 & 0 \\ 0 & 2i & 0 \\ 2 & 0 & 1 \end{vmatrix} + i \begin{vmatrix} i & 0 & 3 \\ 0 & 0 & 2i \\ 2 & 0 & 0 \end{vmatrix} \\ &= -5i \left[i \begin{vmatrix} 2i & 0 \\ 0 & 1 \end{vmatrix} - 3 \begin{vmatrix} 0 & 0 \\ 2 & 1 \end{vmatrix} \right] + i \left[-2i \begin{vmatrix} i & 0 \\ 2 & 0 \end{vmatrix} \right] \\ &= -5i[i \cdot 2i - 3 \cdot 0] + i[-2i \cdot 0] \\ &= -5i[-2] + i[0] \\ &= 10i \end{aligned}$$

8.3 Row and column operations

Consider the following consequences of the properties of the determinant:

- (row and column scaling) If $\mathbf{R}_i \mapsto \lambda \mathbf{R}_i$ for a fixed i, or $\mathbf{C}_a \mapsto \lambda \mathbf{C}_a$, then $\det M \mapsto \lambda \det M$ by multilinearity. If we scale all rows or columns, then $M \mapsto \lambda M$, so $\det M \mapsto \lambda^n \det M$ where M is an $n \times n$ matrix.
- (row and column operations) If $\mathbf{R}_i \mapsto \mathbf{R}_i + \lambda \mathbf{R}_j$ where $i \neq j$ (or the corresponding conversion with columns), then $\det M \mapsto \det M$.
- (row and column exchanges) If we swap $\mathbf{R}_i$ and $\mathbf{R}_j$ (or two columns), then $\det M \mapsto -\det M$.

For example, let us find the determinant of matrix A, where

$$A = \begin{pmatrix} 1 & 1 & a \\ a & 1 & 1 \\ 1 & a & 1 \end{pmatrix}; \quad a \in \mathbb{C}$$

Then:

$$
\begin{aligned}
& & \det A &= \begin{vmatrix} 1 & 1 & a \\ a & 1 & 1 \\ 1 & a & 1 \end{vmatrix} \\
\mathbf{C}_1 \mapsto \mathbf{C}_1 - \mathbf{C}_3 &: & \det A &= \begin{vmatrix} 1-a & 1 & a \\ a-1 & 1 & 1 \\ 0 & a & 1 \end{vmatrix} \\
& & \det A &= (1-a) \begin{vmatrix} 1 & 1 & a \\ -1 & 1 & 1 \\ 0 & a & 1 \end{vmatrix} \\
\mathbf{C}_2 \mapsto \mathbf{C}_2 - \mathbf{C}_3 &: & \det A &= (1-a) \begin{vmatrix} 1 & 1-a & a \\ -1 & 0 & 1 \\ 0 & a-1 & 1 \end{vmatrix} \\
& & \det A &= (1-a)^2 \begin{vmatrix} 1 & 1 & a \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{vmatrix} \\
\mathbf{R}_1 \mapsto \mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3 &: & \det A &= (1-a)^2 \begin{vmatrix} 0 & 0 & a+2 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{vmatrix} \\
& & \det A &= (1-a)^2(a+2) \begin{vmatrix} -1 & 0 \\ 0 & -1 \end{vmatrix} = (1-a)^2(a+2)
\end{aligned}
$$

8.4 Multiplicative property of determinants

Theorem. For $n \times n$ matrices M, N, $\det(MN) = \det M \cdot \det N$.

We can prove this using the following elaboration on the definition of the determinant:

Lemma.

$$\varepsilon_{i_1 i_2 \cdots i_n} M_{i_1 a_1} M_{i_2 a_2} \cdots M_{i_n a_n} = (\det M) \varepsilon_{a_1 a_2 \cdots a_n}$$

Proof. The left hand side and right hand side are each totally antisymmetric (alternating) in $a_1, a_2, \cdots, a_n$, so they must be related by a constant of proportionality. To fix the constant, we can simply consider taking $a_i = i$ and the result follows. $\square$

Now, we prove the above theorem.

Proof. Using the lemma above:

$$
\begin{aligned}
\det MN &= \varepsilon_{i_1 i_2 \cdots i_n} (MN)_{i_1 1} (MN)_{i_2 2} \cdots (MN)_{i_n n} \\
&= \varepsilon_{i_1 i_2 \cdots i_n} \begin{matrix} M_{i_1 k_1} & M_{i_2 k_2} & \cdots & M_{i_n k_n} \\ N_{k_1 1} & N_{k_2 2} & & N_{k_n n} \end{matrix} \\
&= (\det M) \varepsilon_{a_1 a_2 \cdots a_n} N_{k_1 1} N_{k_2 2} \cdots N_{k_n n} \\
&= (\det M)(\det N)
\end{aligned}
$$

as required. □

Note the following consequences.

(i) $M^{-1}M = I \implies \det\!\left(M^{-1}\right)\det(M) = \det I = 1$. Therefore, $\det\!\left(M^{-1}\right) = (\det M)^{-1}$, so $\det M$ must be nonzero for M to be invertible.

(ii) For R real and orthogonal, $R^\intercal R = I \implies \det(R^\intercal)\det(R) = 1$. But $\det(R^\intercal) = \det R$, so $(\det R)^2 = 1$, so $\det R = \pm 1$.

(iii) For U complex and unitary, $U^\dagger U = I \implies \det\!\left(U^\dagger\right)\det(U) = 1$. But since $U^\dagger = \overline{U^\intercal}$, we have $\overline{\det U}\det U = 1$, so $\left|(\det U)^2\right| = 1$, so $|\det U| = 1$.

8.5 Cofactors and determinants

Consider a column of some $n \times n$ matrix M, written in the form

$$\mathbf{C}_a = \sum_i M_{ia}\mathbf{e}_i$$

$$\begin{aligned}
\implies \det M &= [\mathbf{C}_1, \cdots, \mathbf{C}_a, \cdots, \mathbf{C}_n] \\
&= [\mathbf{C}_1, \cdots, \mathbf{C}_{a-1}, \sum_i M_{ia}\mathbf{e}_i, \mathbf{C}_{a+1}, \cdots, \mathbf{C}_n] \\
&= \sum_i M_{ia}\Delta_{ia}
\end{aligned}$$

where

$$\begin{aligned}
\Delta_{ia} &= [\mathbf{C}_1, \cdots, \mathbf{C}_{a-1}, \mathbf{e}_i, \mathbf{C}_{a+1}, \cdots, \mathbf{C}_n] \\
&= \begin{vmatrix}
 & & & 0 & & & \\
 & A & & \vdots & & B & \\
 & & & 0 & & & \\
0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\
 & & & 0 & & & \\
 & C & & \vdots & & D & \\
 & & & 0 & & &
\end{vmatrix}
\end{aligned}$$

where the zero entries in the rows arise from antisymmetry, giving

$$\begin{aligned}
&= \underbrace{(-1)^{n-a}}_{\text{amount of column transpositions}} \cdot \underbrace{(-1)^{n-i}}_{\text{amount of row transpositions}} \begin{vmatrix} A & B \\ C & D \end{vmatrix} \\
&= (-1)^{i+a}M^{ia}
\end{aligned}$$

where M^{ia} is the minor in this position; the determinant of the matrix with this particular row and column removed. We call Δ_{ia} the cofactor.

$$\det M = \sum_i M_{ia}\Delta_{ia} = \sum_i (-1)^{i+a} M_{ia} M^{ia}$$

Similarly, by considering rows,

$$\det M = \sum_a M_{ia}\Delta_{ia} = \sum_a (-1)^{i+a} M_{ia} M^{ia}$$

8.6 Adjugates and inverses

Reasoning as above, consider $\mathbf{C}_b = \sum_i M_{ib}\mathbf{e}_i$. Then,

$$[\mathbf{C}_1, \cdots, \mathbf{C}_{a-1}, \mathbf{C}_b, \mathbf{C}_{a+1}, \cdots, \mathbf{C}_n] = \sum_i M_{ib}\Delta_{ia}$$

If $a = b$ then clearly this is $\det M$. Otherwise, $\mathbf{C}_b$ is equal to one of the other columns, so $\sum_i M_{ib}\Delta_{ia} = 0$.

$$\sum_i M_{ib}\Delta_{ia} = (\det M)\delta_{ab}$$

Similarly,

$$\sum_a M_{ja}\Delta_{ia} = (\det M)\delta_{ij}$$

Now, let Δ be the matrix of cofactors (i.e. entries Δ_{ia}), and we define the adjugate $\widetilde{M} = \Delta^\intercal$. Then

$$\Delta_{ia}M_{ib} = \widetilde{M}_{ai}M_{ib} = (\widetilde{M}M)_{ab} = (\det M)\delta_{ab}$$

Therefore,

$$\widetilde{M}M = (\det M)I$$

We can reach this result similarly considering the other index. Hence, if $\det M \neq 0$ then $M^{-1} = \frac{1}{\det M}\widetilde{M}$.

8.7 Systems of linear equations

Consider a system of n linear equations in n unknowns x_i written in matrix-vector form:

$$A\mathbf{x} = \mathbf{b}, \quad \mathbf{x}, \mathbf{b} \in \mathbb{R}^n,$$

where A is an $n \times n$ matrix. There are three possibilities:

(i) $\det A \neq 0 \implies A^{-1}$ exists so there is a unique solution $\mathbf{x} = A^{-1}\mathbf{b}$

(ii) $\det A = 0$ and $b \notin \operatorname{Im} A$ means that there is no solution

(iii) $\det A = 0$ and $b \in \operatorname{Im} A$ means that there are infinitely many solutions of the form

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{u}$$

where $\mathbf{u} \in \ker A$ and $\mathbf{x}_0$ is a particular solution

A solution therefore exists if and only if $A\mathbf{x}_0 = \mathbf{b}$ for some $\mathbf{x}_0$, which is true if and only if $\mathbf{b} \in \operatorname{Im} A$. Then $\mathbf{x}$ is also a solution if and only if $\mathbf{u} = \mathbf{x} - \mathbf{x}_0$ satisfies

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

This equation is known as the equivalent homogeneous problem. Now, $\det A \neq 0 \iff \operatorname{Im} A = \mathbb{R}^n \iff \ker A = \{\mathbf{0}\}$. So in case (i), there is always a unique solution for any $\mathbf{b}$. But $\det A = 0 \iff \operatorname{rank}(A) < n \iff \operatorname{null} A > 0$. Then either $b \notin \operatorname{Im} A$ as in case (ii), or $b \in \operatorname{Im} A$ as in case (iii).

If $\mathbf{u}_1, \dots, \mathbf{u}_k$ is a basis for $\ker A$, then the general solution to the homogeneous problem is some linear combination of these basis vectors, i.e.

$$\mathbf{u} = \sum_{i=1}^k \lambda_i \mathbf{u}_i, \quad k = \operatorname{null} A$$

This is similar to the complementary function and particular integral technique used to solve linear differential equations.

For example, in $A\mathbf{x} = \mathbf{b}$, let

$$A = \begin{pmatrix} 1 & 1 & a \\ a & 1 & 1 \\ 1 & a & 1 \end{pmatrix}; \quad \mathbf{b} = \begin{pmatrix} 1 \\ c \\ 1 \end{pmatrix}; \quad a, c \in \mathbb{R}$$

We have previously found that $\det A = (a-1)^2(a+2)$. So the cases are:

- $(a \neq 1, a \neq -2)$ $\det A \neq 0$ and A^{-1} exists; we previously found this to be

  $$A^{-1} = \frac{1}{(1-a)(2+a)} \begin{pmatrix} 1 & 1+a & 1 \\ 1 & 1 & -1-a \\ -1-a & 1 & 1 \end{pmatrix}$$

  For these values of a, there is a unique solution for any c, demonstrating case (i) above:

  $$\mathbf{x} = A^{-1}\mathbf{b} = \frac{1}{(1-a)(2+a)} \begin{pmatrix} 2-c-ca \\ c-a \\ c-a \end{pmatrix}$$

  Geometrically, this solution is simply a point.

- $(a = 1)$ In this case, the matrix is simply

  $$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \implies \operatorname{Im} A = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\} = \left\{ \lambda \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}; \quad \ker A = \operatorname{span}\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

  Note that $\mathbf{b} \in \operatorname{Im} A$ if and only if $c = 1$, where a particular solution is

  $$\mathbf{x}_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

  So the general solution is given by

  $$\mathbf{x} = \mathbf{x}_0 + \mathbf{u} = \begin{pmatrix} 1-\lambda-\mu \\ \lambda \\ \mu \end{pmatrix}$$

  In summary, for $a = 1, c = 1$ we have case (iii). Geometrically this is a plane. For $a = 1, c \neq 1$, we have case (ii) where there are no solutions.

- $(a = -2)$ The matrix becomes

  $$A = \begin{pmatrix} 1 & 1 & -2 \\ -2 & 1 & 1 \\ 1 & -2 & 1 \end{pmatrix} \implies \operatorname{Im} A = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \right\}; \quad \ker A = \left\{ \lambda \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$$

  Now, $\mathbf{b} \in \operatorname{Im} A$ if and only if $c = -2$, the particular solution is

  $$\mathbf{x}_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

The general solution is therefore

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{u} = \begin{pmatrix} 1+\lambda \\ \lambda \\ \lambda \end{pmatrix}$$

In summary, for $a = -2$ and $c = -2$ we have case (iii). Geometrically this is a line. For $a = -2$, $c \neq -2$, we have case (ii) where there are no solutions.

8.8 Geometrical interpretation of solutions of linear equations

Let $\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3$ be the rows of the 3×3 matrix A. Then the rows represent the normals of planes. This is clear by expanding the matrix multiplication of the homogeneous form:

$$\begin{aligned} A\mathbf{u} = \mathbf{0} \iff \mathbf{R}_1 \cdot \mathbf{u} &= 0 \\ \mathbf{R}_2 \cdot \mathbf{u} &= 0 \\ \mathbf{R}_3 \cdot \mathbf{u} &= 0 \end{aligned}$$

So the solution of the homogeneous problem (i.e. finding the general solution) amounts to determining where the planes intersect.

- ($\operatorname{rank} A = 3$) The rows are linearly independent, so the three planes' normals are linearly independent and the planes intersect at $\mathbf{0}$ only.
- ($\operatorname{rank} A = 2$) The normals span a plane, so the planes intersect in a line.
- ($\operatorname{rank} A = 1$) The normals are parallel and therefore the planes coincide.
- ($\operatorname{rank} A = 0$) The normals are all zero, so any vector in $\mathbb{R}^3$ solves the equation.

Now, let us consider instead the original problem $A\mathbf{x} = \mathbf{b}$:

$$\begin{aligned} A\mathbf{b} = \mathbf{0} \iff \mathbf{R}_1 \cdot \mathbf{u} &= b_1 \\ \mathbf{R}_2 \cdot \mathbf{u} &= b_2 \\ \mathbf{R}_3 \cdot \mathbf{u} &= b_3 \end{aligned}$$

The planes still have normals $\mathbf{R}_i$ as before, but they do not necessarily pass through the origin.

- ($\operatorname{rank} A = 3$) The planes' normals are linearly independent and the planes intersect at a point; this is the unique solution.
- ($\operatorname{rank} A < 3$) The existence of a solution depends on the value of $\mathbf{b}$.
  - ($\operatorname{rank} A = 2$) The planes may intersect in a line as before, but they may instead form a sheaf (the planes pairwise intersect in lines but they do not as a triple), or two planes could be parallel and not intersect each other at all.
  - ($\operatorname{rank} A = 1$) The normals are parallel, so the planes may coincide or they might be parallel. There is no solution unless all three planes coincide.

9 Properties of matrices

9.1 Eigenvalues and eigenvectors

For a linear map $T: V \to V$, a vector $\mathbf{v} \in V$ with $\mathbf{v} \neq 0$ is called an eigenvector of T with eigenvalue λ if $T(\mathbf{v}) = \lambda \mathbf{v}$. If $V = \mathbb{R}^n$ or $\mathbb{C}^n$, and T is given by an $n \times n$ matrix A, then

$$A\mathbf{v} = \lambda v \iff (A - \lambda I)\mathbf{v} = \mathbf{0}$$

and for a given λ, this holds for some $\mathbf{v} \neq 0$ if and only if

$$\det(A - \lambda I) = 0$$

This is called the characteristic equation for A. So λ is an eigenvalue if and only if it is a root of the characteristic polynomial

$$\chi_A(t) = \det(A - tI) = \begin{vmatrix} A_{11} - t & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} - t & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} - t \end{vmatrix}$$

We can look for eigenvalues as roots of the characteristic polynomial or characteristic equation, and then determine the corresponding eigenvectors once we've deduced what the possibilities are. Here are a few examples.

(i) $V = \mathbb{C}^2$:

$$A = \begin{pmatrix} 2 & i \\ -i & 2 \end{pmatrix} \implies \det(A - \lambda I) = (2 - \lambda)^2 - 1 = 0$$

So we have $(2 - \lambda)^2 = 1$ so $\lambda = 1$ or 3.

- $(\lambda = 1)$

$$(A - I)\mathbf{v} = \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \mathbf{0} \implies \mathbf{v} = \alpha \begin{pmatrix} 1 \\ i \end{pmatrix}$$

for any $\alpha \neq 0$.

- $(\lambda = 3)$

$$(A - 3I)\mathbf{v} = \begin{pmatrix} -1 & i \\ -i & -1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \mathbf{0} \implies \mathbf{v} = \beta \begin{pmatrix} 1 \\ -i \end{pmatrix}$$

for any $\beta \neq 0$.

(ii) $V = \mathbb{R}^2$:

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \implies \det(A - \lambda I) = (1 - \lambda)^2 = 0$$

So $\lambda = 1$ only, a repeated root.

$$(A - I)\mathbf{v} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \mathbf{0} \implies \mathbf{v} = \alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

for any $\alpha \neq 0$. There is only one (linearly independent) eigenvector here.

(iii) $V = \mathbb{R}^2$ or $\mathbb{C}^2$:

$$U = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \implies \chi_U(t) = \det(U - tI) = t^2 - 2t\cos\theta + 1$$

The eigenvalues λ are $e^{\pm i\theta}$. The eigenvectors are

$$\mathbf{v} = \alpha \begin{pmatrix} 1 \\ \mp i \end{pmatrix}; \quad \alpha \neq 0$$

So there are no real eigenvalues or eigenvectors except when $\theta = n\pi$.

(iv) $V = \mathbb{C}^n$:

$$A = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \implies \chi_A(t) = \det(A - tI) = (\lambda_1 - t)(\lambda_2 - t)(\lambda_3 - t) \dots (\lambda_n - t)$$

So the eigenvalues are all the λ_i, and the eigenvectors are $\mathbf{v} = \alpha\mathbf{e}_i$ ($\alpha \neq 0$) for each i.

9.2 The characteristic polynomial

For an $n \times n$ matrix A, the characteristic polynomial $\chi_A(t)$ has degree n:

$$\chi_A(t) = \sum_{j=0}^{n} c_j t^j = (-1)^n (t - \lambda_1) \dots (t - \lambda_n)$$

(i) There exists at least one eigenvalue (solution to χ_A), due to the fundamental theorem of algebra, or n roots counted with multiplicity.

(ii) $\mathrm{tr}(A) = A_{ii} = \sum_{i=1}^{n} \lambda_i$, the sum of the eigenvalues. Compare terms of degree $n - 1$ in t, and from the determinant we get

$$(-t)^{n-1} A_{11} + (-t)^{n-1} A_{22} + \dots + (-t)^{n-1} A_{nn}$$

The overall sign matches with the expansion of $(-1)^n (t - \lambda_1)(t - \lambda_2) \dots (t - \lambda_n)$.

(iii) $\det(A) = \chi_A(0) = \prod_{i=1}^{n} \lambda_i$, the product of the eigenvalues.

(iv) If A is real, then the coefficients c_i in the characteristic polynomial are real, so $\chi_A(\lambda) = 0 \iff \chi_A(\overline{\lambda}) = 0$. So the non-real roots occur in conjugate pairs if A is real.

9.3 Eigenspaces and multiplicities

For an eigenvalue λ of a matrix A, we define the eigenspace

$$E_\lambda = \{\mathbf{v} : A\mathbf{v} = \lambda\mathbf{v}\} = \ker(A - \lambda I)$$

All nonzero vectors in this space are eigenvectors. The geometric multiplicity is

$$m_\lambda = \dim E_\lambda = \mathrm{null}(A - \lambda I)$$

equivalent to the number of linearly independent eigenvectors with the given eigenvalue λ. The algebraic multiplicity is

$$M_\lambda = \text{the multiplicity of } \lambda \text{ as a root of } \chi_A(t)$$

i.e. $\chi_A(t) = (t - \lambda)^{M_t} f(t)$, where $f(\lambda) \neq 0$.

Proposition. $M_\lambda \geq m_\lambda$ (and $m_\lambda \geq 1$ since λ is an eigenvalue). The proof of this proposition is delayed until the next section where we will then have the tools to prove it.

Here are some examples.

(i)

$$A = \begin{pmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{pmatrix} \implies \chi_A(t) = \det(A - tI) = (5-t)(t+3)^2$$

So $\lambda = 5, -3$. $M_5 = 1, M_{-3} = 2$. We will now find the eigenspaces.

- ($\lambda = 5$)

$$E_5 = \left\{ \alpha \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \right\}$$

- ($\lambda = -3$)

$$E_{-3} = \left\{ \alpha \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix} \right\}$$

Note that to compute the eigenvectors, we just need to solve the equation $(A - \lambda I)\mathbf{x} = \mathbf{0}$. In the case of $\lambda = -3$, for example, we then have

$$\begin{pmatrix} 1 & 2 & -3 \\ 2 & 4 & -6 \\ -1 & -2 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{0}$$

We can use the first line of the matrix to get a linear combination for x_1, x_2, x_3, specifically $x_1 + 2x_2 = 3x_3 = 0$, so we can eliminate one of the variables (here, x_1) to get

$$\mathbf{x} = \begin{pmatrix} -2x_2 + 3x_3 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{0}$$

Now, $\dim E_5 = m_5 = 1 = M_5$. Similarly, $\dim E_{-3} = m_{-3} = 2 = M_{-3}$.

(ii)

$$A = \begin{pmatrix} -3 & -1 & 1 \\ -1 & -3 & 1 \\ -2 & -2 & 0 \end{pmatrix} \implies \chi_A(t) = \det(A - tI) = -(t+2)^3$$

We have a root $\lambda = -2$ with $M_{-2} = 3$. To find the eigenspace, we will look for solutions of:

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

So

$$E_{-2} = \left\{ \alpha \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

Further, $m_{-2} = 2 < 3 = M_{-2}$.

(iii) A reflection in a plane through the origin with unit normal $\hat{\mathbf{n}}$ satisfies

$$H\hat{\mathbf{n}} = -\hat{\mathbf{n}}; \quad \forall \mathbf{u} \perp \hat{\mathbf{n}}, H\mathbf{u} = \mathbf{u}$$

The eigenvalues are therefore ± 1 and $E_{-1} = \{\alpha\hat{\mathbf{n}}\}$, and $E_1 = \{\mathbf{x} : \mathbf{x} \cdot \hat{\mathbf{n}} = 0\}$. The multiplicities are given by $M_{-1} = m_{-1} = 1, M_1 = m_1 = 2$.

(iv) A rotation about an axis $\hat{\mathbf{n}}$ through angle θ in $\mathbb{R}^3$ satisfies

$$R\hat{\mathbf{n}} = \hat{\mathbf{n}}$$

So the axis of rotation is the eigenvector with eigenvalue 1. There are no other real eigenvalues unless $\theta = n\pi$. The rotation restricted to the plane perpendicular to $\hat{\mathbf{n}}$ has eigenvalues $e^{\pm i\theta}$ as shown above.

9.4 Linear independence of eigenvectors

Proposition. Let $\mathbf{v}_1, \dots, \mathbf{v}_r$ be eigenvectors of an $n \times n$ matrix A with eigenvalues $\lambda_1, \dots, \lambda_r$. If the eigenvalues are distinct, then the eigenvectors are linearly independent.

Proof. Note that if we take some linear combination $\mathbf{w} = \sum_{j=1}^r \alpha_j \mathbf{v}_j$, then $(A - \lambda I)\mathbf{w} = \sum_{j=1}^r \alpha_j(\lambda_j - \lambda)\mathbf{v}_j$. Here are two methods for getting this proof.

(i) Suppose the eigenvectors are linearly dependent, so there exist linear combinations $\mathbf{w} = \mathbf{0}$ where some α are nonzero. Let p be the amount of nonzero α values. So, $2 \leq p \leq r$. Now, pick such a $\mathbf{w}$ for which p is least. Without loss of generality, let α_1 be one of the nonzero coefficients. Then

$$(A - \lambda_1 I)\mathbf{w} = \sum_{j=2}^r \alpha_j(\lambda_j - \lambda_1)\mathbf{v}_j = \mathbf{0}$$

This is a linear relation with $p - 1$ nonzero coefficients #.

(ii) Alternatively, given a linear relation $\mathbf{w} = \mathbf{0}$,

$$\prod_{j \neq k}(A - \lambda_j I)\mathbf{w} = \alpha_k \prod_{j \neq k}(\lambda_k - \lambda_j)\mathbf{v}_k = \mathbf{0}$$

for some fixed k. So $\alpha_k = 0$. So the eigenvectors are linearly independent as claimed.

□

Corollary. With conditions as in the proposition above, let $\mathcal{B}_{\lambda_i}$ be a basis for the eigenspace E_{λ_i}. Then $\mathcal{B} = \mathcal{B}_{\lambda_1} \cup \mathcal{B}_{\lambda_2} \cup \dots \cup \mathcal{B}_{\lambda_r}$ is linearly independent.

Proof. Consider a general linear combination of all these vectors, it has the form

$$\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2 + \dots + \mathbf{w}_r$$

where each $\mathbf{w}_i \in E_i$. Applying the same arguments as in the proposition, we find that

$$\mathbf{w} = 0 \implies \forall i\, \mathbf{w}_i = 0$$

So each $\mathbf{w}_i$ is the trivial linear combination of elements of $\mathcal{B}_{\lambda_i}$ and the result follows. □

9.5 Diagonalisability

Proposition. For an $n \times n$ matrix A acting on $V = \mathbb{R}^n$ or $\mathbb{C}^n$, the following conditions are equivalent:

(i) there exists a basis of eigenvectors of A for V, named $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ which $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ for each i; and

(ii) there exists an $n \times n$ invertible matrix P with the property that

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

If either of these conditions hold, then A is diagonalisable.

Proof. Note that for any matrix P, AP has columns $A\mathbf{C}_i(P)$, and PD has columns $\lambda_i \mathbf{C}_i(P)$. Then (i) and (ii) are related by choosing $\mathbf{v}_i = \mathbf{C}_i(P)$. Then $P^{-1}AP = D \iff AP = PD \iff A\mathbf{v}_i = \lambda_i \mathbf{v}_i$.

In essence, given a basis of eigenvectors as in (i), the relation above defines P, and if the eigenvectors are linearly independent then P is invertible. Conversely, given a matrix P as in (ii), its columns are a basis of eigenvectors. □

Let's try some examples.

(i) Let
$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \implies E_1 = \left\{ \alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$$
This is a single eigenvalue $\lambda = 1$ with one linearly independent eigenvector. So there is no basis of eigenvectors for $\mathbb{R}^2$ or $\mathbb{C}^2$, so A is not diagonalisable.

(ii) Let
$$U = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \implies E_{e^{i\theta}} = \left\{ \alpha \begin{pmatrix} 1 \\ -i \end{pmatrix} \right\}; \quad E_{e^{-i\theta}} = \left\{ \beta \begin{pmatrix} 1 \\ i \end{pmatrix} \right\}$$
which are two linearly independent complex eigenvectors. So,
$$P = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}; \quad P^{-1} = \frac{1}{2} \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix}; \quad P^{-1}UP = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{i\theta} \end{pmatrix}$$
So U is diagonalisable over $\mathbb{C}^2$ but not over $\mathbb{R}^2$.

9.6 Criteria for diagonalisability

Proposition. Consider an $n \times n$ matrix A.

(i) A is diagonalisable if it has n distinct eigenvalues (sufficient condition).

(ii) A is diagonalisable if and only if for every eigenvalue λ, $M_\lambda = m_\lambda$ (necessary and sufficient condition).

Proof. Use the proposition and corollary above.

(i) If we have n distinct eigenvalues, then we have n linearly independent eigenvectors. Hence they form a basis.

(ii) If λ_i are all the distinct eigenvalues, then $\mathcal{B}_{\lambda_1} \cup \dots \cup \mathcal{B}_{\lambda_r}$ are linearly independent. The number of elements in this new basis is $\sum_i m_{\lambda_i} = \sum_i M_{\lambda_i} = n$ which is the degree of the characteristic polynomial. So we have a basis.

Note that case (i) is just a specialisation of case (ii) where both multiplicities are 1. □

Let us consider some examples.

(i) Let

$$A = \begin{pmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{pmatrix} \implies \lambda = 5, -3; \quad M_5 = m_5 = 1; \quad M_{-3} = m_{-3} = 2$$

So A is diagonalisable by case (ii) above, and moreover

$$P = \begin{pmatrix} 1 & -2 & 3 \\ 2 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}; \quad P^{-1} = \frac{1}{8}\begin{pmatrix} 1 & 2 & -3 \\ -2 & 4 & 6 \\ 1 & 2 & 5 \end{pmatrix} \implies P^{-1}AP = \begin{pmatrix} 5 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -3 \end{pmatrix}$$

(ii) Let

$$A = \begin{pmatrix} -3 & -1 & 1 \\ -1 & -3 & 1 \\ -2 & 2 & 0 \end{pmatrix} \implies \lambda = -2; \quad M_{-2} = 3 > m_{-2} = 2$$

So A is not diagonalisable. As a check, if it were diagonalisable, then there would be some matrix P such that $P^{-1}AP = -2I \implies A = P(-2I)P^{-1} = -2I$ #.

9.7 Similarity

Matrices A and B (both $n \times n$) are similar if $B = P^{-1}AP$ for some invertible $n \times n$ matrix P. This is an equivalence relation.

Proposition. If A and B are similar, then
(i) $\operatorname{tr} B = \operatorname{tr} A$
(ii) $\det B = \det A$
(iii) $\chi_B = \chi_A$

Proof. (i)

$$\begin{aligned} \operatorname{tr} B &= \operatorname{tr}(P^{-1}AP) \\ &= \operatorname{tr}(APP^{-1}) \\ &= \operatorname{tr} A \end{aligned}$$

(ii)

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

(iii)

$$\begin{aligned}
\det(B - tI) &= \det\bigl(P^{-1}AP - tI\bigr) \\
&= \det\bigl(P^{-1}AP - tP^{-1}P\bigr) \\
&= \det\bigl(P^{-1}(A - tI)P\bigr) \\
&= \det P^{-1} \det(A - tI) \det P \\
&= \det(A - tI)
\end{aligned}$$

□

9.8 Real eigenvalues and orthogonal eigenvectors

Recall that an $n \times n$ matrix A is hermitian if and only if $A^\dagger = \overline{A}^\intercal = A$, or $\overline{A_{ij}} = A_{ji}$. If A is real, then it is hermitian if and only if it is symmetric. The complex inner product for $\mathbf{v}, \mathbf{w} \in \mathbb{C}^n$ is $\mathbf{v}^\dagger\mathbf{w} = \sum_i \overline{v_i} w_i$, and for $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, this reduces to the dot product in $\mathbb{R}^n$, $\mathbf{v}^\intercal\mathbf{w}$.

Here is a key observation. If A is hermitian, then

$$(A\mathbf{v})^\dagger\mathbf{w} = \mathbf{v}^\dagger(A\mathbf{w})$$

Theorem. For an $n \times n$ matrix A that is hermitian:

(i) Every eigenvalue λ is real;

(ii) Eigenvectors $\mathbf{v}, \mathbf{w}$ with different eigenvalues λ, μ respectively, are orthogonal, i.e. $\mathbf{v}^\dagger\mathbf{w} = 0$; and

(iii) If A is real and symmetric, then for each eigenvalue λ we can choose a real eigenvector, and part (ii) becomes $\mathbf{v} \cdot \mathbf{w} = 0$.

Proof. (i) Using the observation above with $\mathbf{v} = \mathbf{w}$ where $\mathbf{v}$ is any eigenvector with eigenvalue λ, we get

$$\begin{aligned}
\mathbf{v}^\dagger(A\mathbf{v}) &= (A\mathbf{v})^\dagger\mathbf{v} \\
\mathbf{v}^\dagger(\lambda\mathbf{v}) &= (\lambda\mathbf{v})^\dagger\mathbf{v} \\
\lambda\mathbf{v}^\dagger(\mathbf{v}) &= \overline{\lambda}(\mathbf{v})^\dagger\mathbf{v}
\end{aligned}$$

As $\mathbf{v}$ is an eigenvector, it is nonzero, so $\mathbf{v}^\dagger\mathbf{v} \neq 0$, so

$$\lambda = \overline{\lambda}$$

(ii) Using the same observation,

$$\begin{aligned}
\mathbf{v}^\dagger(A\mathbf{w}) &= (A\mathbf{v})^\dagger\mathbf{w} \\
\mathbf{v}^\dagger(\mu\mathbf{w}) &= (\lambda\mathbf{v})^\dagger\mathbf{w} \\
\mu\mathbf{v}^\dagger\mathbf{w} &= \lambda\mathbf{v}^\dagger\mathbf{w}
\end{aligned}$$

Since $\lambda \neq \mu$, $\mathbf{v}^\dagger\mathbf{w} = 0$, so the eigenvectors are orthogonal.

(iii) Given $A\mathbf{v} = \lambda\mathbf{v}$ with $\mathbf{v} \in \mathbb{C}^n$ but A is real, let

$$\mathbf{v} = \mathbf{u} + i\mathbf{u}'; \quad \mathbf{u}, \mathbf{u}' \in \mathbb{R}^n$$

Since $\mathbf{v}$ is an eigenvector, and this is a linear equation, we have

$$A\mathbf{u} = \lambda\mathbf{u}; \quad A\mathbf{u}' = \lambda\mathbf{u}'$$

So $\mathbf{u}$ and $\mathbf{u}'$ are eigenvectors. $\mathbf{v} \neq 0$ implies that at least one of $\mathbf{u}$ and $\mathbf{u}'$ are nonzero, so there is at least one real eigenvector with this eigenvalue.

□

Case (ii) is a stronger claim for hermitian matrices than just showing that eigenvectors are linearly independent. Furthermore, previously we considered bases $\mathcal{B}_\lambda$ for each eigenspace E_λ, and it is now natural to choose bases $\mathcal{B}_\lambda$ to be orthonormal when we are considering hermitian matrices. Here are some examples.

(i) Let

$$A = \begin{pmatrix} 2 & i \\ -i & 2 \end{pmatrix}; \quad A^\dagger = A; \quad \lambda = 1, 3; \quad \mathbf{u}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}; \quad \mathbf{u}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}$$

We have chosen coefficients for the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ such that they are unit vectors. As shown above, they are then orthonormal. We know that having distinct eigenvalues means that a matrix is diagonalisable. So let us set

$$P = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \implies P^{-1}AP = D = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$$

Since the eigenvectors are orthonormal, so are the columns of P, so $P^{-1} = P^\dagger$ (i.e. P is unitary).

(ii) Let

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

A is real and symmetric, with eigenvalues $\lambda = -1, 2$ with $M_{-1} = 2, M_2 = 1$. Further,

$$E_{-1} = \operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2\}; \quad \mathbf{w}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}; \quad \mathbf{w}_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

So $m_{-1} = 2$, and the matrix is diagonalisable. Let us choose an orthonormal basis for E_{-1} by taking

$$\mathbf{u}_1 = \frac{1}{|\mathbf{w}_1|}\mathbf{w}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

and we can consider

$$\mathbf{w}_2' = \mathbf{w}_2 - (\mathbf{u}_1 \cdot \mathbf{w}_2)\mathbf{u}_1 = \begin{pmatrix} 1/2 \\ 1/2 \\ -1 \end{pmatrix}$$

so that $\mathbf{w}_2'$ is orthogonal to $\mathbf{u}_1$ by construction. We can then normalise this vector to get

$$\mathbf{u}_2 = \frac{1}{|\mathbf{w}_2'|}\mathbf{w}_2' = \frac{1}{\sqrt{6}}\begin{pmatrix}1\\1\\-2\end{pmatrix}$$

and therefore

$$\mathcal{B}_{-1} = \{\mathbf{u}_1, \mathbf{u}_2\}$$

is an orthonormal basis. For E_2, let us choose $\mathcal{B}_2 = \{\mathbf{u}_3\}$ where

$$\mathbf{u}_3 = \frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}$$

Together,

$$\mathcal{B} = \left\{\frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\\0\end{pmatrix}, \frac{1}{\sqrt{6}}\begin{pmatrix}1\\1\\-2\end{pmatrix}, \frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}\right\}$$

is an orthonormal basis for $\mathbb{R}^3$. Let P be the matrix with columns $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$, then $P^{-1}AP = D$ as required. Since we have chosen an orthonormal basis, P is orthogonal, so $P^\intercal AP = D$.

9.9 Unitary and orthogonal diagonalisation

Theorem. Any $n \times n$ hermitian matrix A is diagonalisable.

(i) There exists a basis of eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_n \in \mathbb{C}^n$ with $A\mathbf{u}_i = \lambda\mathbf{u}_i$; equivalently

(ii) There exists an $n \times n$ invertible matrix P with $P^{-1}AP = D$ where D is the matrix with eigenvalues on the diagonal, where the columns of P are the eigenvectors $\mathbf{u}_i$.

In addition, the eigenvectors $\mathbf{u}_i$ can be chosen to be orthonormal, so

$$\mathbf{u}_i^\dagger \mathbf{u}_j = \delta_{ij}$$

or equivalently, the matrix P can be chosen to be unitary,

$$P^\dagger = P^{-1} \implies P^\dagger AP = D$$

In the special case that the matrix A is real, the eigenvectors can be chosen to be real, and so

$$\mathbf{u}^\intercal \mathbf{u}_j = \mathbf{u}_i \cdot \mathbf{u}_j = \delta_{ij}$$

so P is orthogonal, so

$$P^\intercal = P^{-1} \implies P^\intercal AP = D$$

10 Quadratic forms

10.1 Simple example

Consider a function $\mathcal{F} : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\mathcal{F}(\mathbf{x}) = 2x_1^2 - 4x_1x_2 + 5x_2^2$$

This can be simplified by writing

$$\mathcal{F}(\mathbf{x}) = x_1'^2 + 6x_2'^2$$

where

$$x_1' = \frac{1}{\sqrt{5}}(2x_1 + x_2); \quad x_2' = \frac{1}{\sqrt{5}}(-x_1 + 2x_2)$$

This can be found by writing $\mathcal{F}(\mathbf{x}) = \mathbf{x}^\intercal A\mathbf{x}$ where

$$A = \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix}$$

by inspection from the original equation, and then diagonalising A. We find the eigenvalues to be $\lambda = 1, 6$, with eigenvectors

$$\frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ 1 \end{pmatrix}; \quad \frac{1}{\sqrt{5}}\begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

10.2 Diagonalising quadratic forms

In general, a quadratic form is a function $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}$ given by

$$\mathcal{F}(\mathbf{x}) = \mathbf{x}^\intercal A\mathbf{x} \implies \mathcal{F}(\mathbf{x})_{ij} = x_i A_{ij} x_j$$

where A is a real symmetric $n \times n$ matrix. Any antisymmetric part of A would not contribute to the result, so there is no loss of generality under this restriction. From the section above, we know we can write $P^\intercal AP = D$ where D is a diagonal matrix containing the eigenvalues, and P is constructed from the eigenvectors, with orthonormal columns $\mathbf{u}_i$. Setting $\mathbf{x}' = P^\intercal\mathbf{x}$, or equivalently $\mathbf{x} = P\mathbf{x}'$, we have

$$\begin{aligned} \mathcal{F}(\mathbf{x}) &= \mathbf{x}^\intercal A\mathbf{x} \\ &= (P\mathbf{x}')^\intercal A(P\mathbf{x}') \\ &= (\mathbf{x}')^\intercal P^\intercal AP\mathbf{x}' \\ &= (\mathbf{x}')^\intercal D\mathbf{x}' \\ &= \sum_i \lambda_i x_i'^2 = \lambda_1 x_1'^2 + \lambda_2 x_2'^2 + \dots \end{aligned}$$

We say that $\mathcal{F}$ has been diagonalised. Now, note that

$$\begin{aligned} \mathbf{x}' &= x_1'\mathbf{e}_1 + \dots + x_n'\mathbf{e}_n \\ \mathbf{x} &= x_1\mathbf{e}_1 + \dots + x_n\mathbf{e}_n \\ &= x_1'\mathbf{u}_1 + \dots + x_n'\mathbf{u}_n \end{aligned}$$

where the $\mathbf{e}_i$ are the standard basis vectors, since

$$\mathbf{x}_i' = \mathbf{u}_i \cdot \mathbf{x} \iff \mathbf{x}' = P^\intercal\mathbf{x}$$

Hence the $\mathbf{x}_i'$ can be regarded as coordinates with respect to a new set of axes defined by the orthonormal eigenvector basis, known as the principal axes of the quadratic form. They are related to the standard axes (given by basis vectors $\mathbf{e}_i$) by the orthogonal transformation P.

Example (two dimensions)**.** Consider $\mathcal{F}(\mathbf{x}) = \mathbf{x}^\intercal A\mathbf{x}$ with

$$A = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}$$

The eigenvalues are $\lambda = \alpha + \beta, \alpha - \beta$ and

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}; \quad \mathbf{u}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

So in terms of the standard basis vectors,

$$\mathcal{F}(\mathbf{x}) = \alpha x_1^2 + 2\beta x_1 x_2 + \alpha x_2^2$$

And in terms of our new basis vectors,

$$\mathcal{F}(\mathbf{x}) = (\alpha + \beta)x_1'^2 + (\alpha - \beta)x_2'^2$$

where

$$\begin{aligned} \mathbf{x}_1' &= \mathbf{u}_1 \cdot \mathbf{x} = \frac{1}{\sqrt{2}}(x_1 + x_2) \\ \mathbf{x}_2' &= \mathbf{u}_2 \cdot \mathbf{x} = \frac{1}{\sqrt{2}}(-x_1 + x_2) \end{aligned}$$

Taking for example $\alpha = \frac{3}{2}, \beta = \frac{-1}{2}$, we have $\lambda_1 = 1, \lambda_2 = 2$. If we choose $\mathcal{F} = 1$, this represents an ellipse in our new coordinate system:

$$x_1'^2 + 2x_2'^2 = 1$$

If instead we chose $\alpha = \frac{-1}{2}, \beta = \frac{3}{2}$. We now have $\lambda_1 = 1, \lambda_2 = -2$. The locus at $\mathcal{F} = 1$ gives a hyperbola:

$$x_1'^2 - 2x_2'^2 = 1$$

Example (three dimensions)**.** In $\mathbb{R}^3$, note that if $\lambda_1, \lambda_2, \lambda_3$ are all strictly positive, then $\mathcal{F} = 1$ gives an ellipsoid. This is analogous to the $\mathbb{R}^2$ case above.

Let us consider an example. Earlier, we found that the eigenvalues of the matrix A where

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

are $\lambda_1 = \lambda_2 = -1, \lambda_3 = 2$, where

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}; \quad \mathbf{u}_2 = \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}; \quad \mathbf{u}_3 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Then

$$\begin{aligned} \mathcal{F}(\mathbf{x}) &= 2x_1x_2 + 2x_2x_3 + 2x_3x_1 \\ &= -x_1'^2 - x_2'^2 + 2x_3'^2 \end{aligned}$$

Now, $\mathcal{F} = 1$ corresponds to

$$2x_3'^2 = 1 + x_1'^2 + x_2'^2$$

So we can more clearly see that this is a hyperboloid of two sheets in $\mathbb{R}^3$ with rotational symmetry between the x_1 and x_2 axes. Further, $\mathcal{F} = -1$ corresponds to

$$1 + 2x_3'^2 = x_1'^2 + x_2'^2$$

Here, this is a hyperboloid of one sheet since for any fixed x_3 coordinate, it defines a circle in the x_1 and x_2 axes.

10.3 Hessian matrix as a quadratic form

Consider a smooth function $f : \mathbb{R}^n \to \mathbb{R}$ with a stationary point at $\mathbf{x} = \mathbf{a}$, i.e. $\frac{\partial f}{\partial x_i} = 0$ at $\mathbf{x} = \mathbf{a}$. By Taylor's theorem,

$$f(\mathbf{a} + \mathbf{h}) + f(\mathbf{a}) + \mathcal{F}(\mathbf{h}) + O(|\mathbf{h}|^3)$$

where $\mathcal{F}$ is a quadratic form with

$$A_{ij} = \frac{1}{2}\frac{\partial^2 f}{\partial x_i \partial x_j}$$

all evaluated at $\mathbf{x} = \mathbf{a}$. Note that this A is half of the Hessian matrix, and that the linear term vanishes since we are at a stationary point. Rewriting this $\mathbf{h}$ in terms of the eigenvectors of A (the principal axes), we have

$$\mathcal{F} = \lambda_1 h_1'^2 + \lambda_2 h_2'^2 + \cdots + \lambda_n h_n'^2$$

So clearly if $\lambda_i > 0$ for all i, then f has a minimum at $\mathbf{x} = \mathbf{a}$. If $\lambda_i < 0$ for all I, then f has a maximum at $\mathbf{x} = \mathbf{a}$. Otherwise, it has a saddle point. Note that it is often sufficient to consider the trace and determinant of A, since $\operatorname{tr} A = \lambda_1 + \lambda_2$ and $\det A = \lambda_1\lambda_2$.

11 Cayley–Hamilton theorem

11.1 Matrix polynomials

If A is an $n \times n$ complex matrix and

$$p(t) = c_0 + c_1 t + c_2^2 + \cdots + c_k t^k$$

is a polynomial, then

$$p(A) = c_0 I + c_1 A + c_2 A^2 + \cdots + c_k A^k$$

We can also define power series on matrices (subject to convergence). For example, the exponential series which always converges:

$$\exp(A) = I + A + \frac{1}{2}A^2 + \cdots + \frac{1}{r!}A^r + \dots$$

For a diagonal matrix, polynomials and power series can be computed easily since the power of a diagonal matrix just involves raising its diagonal elements to said power. Therefore,

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \implies p(D) = \begin{pmatrix} p(\lambda_1) & 0 & \cdots & 0 \\ 0 & p(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p(\lambda_n) \end{pmatrix}$$

Therefore,

$$\exp(D) = \begin{pmatrix} e^{\lambda_1} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\lambda_n} \end{pmatrix}$$

If $B = P^{-1}AP$ (similar to A) where P is an $n \times n$ invertible matrix, then

$$B^r = P^{-1}A^rP$$

Therefore,

$$p(B) = p(P^{-1}AP) = P^{-1}p(A)P$$

Of special interest is the characteristic polynomial,

$$\chi_A(t) = \det(A - tI) = c_0 + c_1t + c_2t^2 + \cdots + c_nt^n$$

where $c_0 = \det A$, and $c_n = (-1)^n$.

Theorem (Cayley–Hamilton Theorem)**.**

$$\chi_A(A) = c_0I + c_1A + c_2A^2 + \cdots + c_nA^n = 0$$

Less formally, a matrix satisfies its own characteristic equation.

Remark. We can find an expression for the matrix inverse.

$$-c_0I = A(c_1 + c_2A + \cdots + c_nA^{n-1})$$

If $c_0 = \det A \neq 0$, then

$$A^{-1} = \frac{-1}{c_0}(c_1 + c_2A + \cdots + c_nA^{n-1})$$

11.2 Proofs of special cases of Cayley–Hamilton theorem

Proof for a 2×2 *matrix.* Let A be a general 2×2 matrix.

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \implies \chi_A(t) = t^2 - (a+d)t + (ad - bc)$$

We can check the theorem by substitution.

$$\chi_A(A) = A^2 - (a+d)A - (ad - bc)I$$

This is shown on the last example sheet. □

Proof for diagonalisable $n \times n$ *matrices.* Consider A with eigenvalues λ_i, and an invertible matrix P such that $P^{-1}AP = D$, where D is diagonal.

$$\chi_A(D) = \begin{pmatrix} \chi_A(\lambda_1) & 0 & \cdots & 0 \\ 0 & \chi_A(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \chi_A(\lambda_n) \end{pmatrix} = 0$$

since the λ_i are eigenvalues. Then

$$\chi_A(A) = \chi_A(PDP^{-1}) = P\chi_A(D)P^{-1} = 0$$

□

11.3 Proof in general case (non-examinable)

Proof. Let $M = A - tI$. Then $\det M = \det(A - tI) = \chi_A(t) = \sum_{r=0} c_r t^r$. We can construct the adjugate matrix.

$$\widetilde{M} = \sum_{r=0}^{n-1} B_r t^r$$

Therefore,

$$\begin{aligned}
\widetilde{M}M = (\det M)I &= \left(\sum_{r=0}^{n-1} B_r t^r\right)(A - tI) \\
&= B_0 A + (B_1 A - B_0)t + (B_2 A - B_1)t^2 + \dots + (B_{n-1}A - B_{n-2})t^{n-1} - B_{n-1}t
\end{aligned}$$

Now by comparing coefficients,

$$\begin{aligned}
C_0 I &= B_0 A \\
C_1 I &= B_1 A - B_0 \\
&\vdots \\
C_{n-1} I &= B_{n-1}A - B_{n-2} \\
C_n I &= -B_{n-1}
\end{aligned}$$

Summing all of these coefficients, multiplying by the relevant powers,

$$\begin{aligned}
&C_0 I + C_1 A + C_2 A^2 + \dots + C_n A^n \\
&= B_0 A + (B_1 A^2 - B_0 A) + (B_2 A^3 - B_1 A^2) + \dots + (B_{n-1}A^n - B_{n-2}A^{n-1}) - B_{n-1}A^n \\
&= 0
\end{aligned}$$

$\square$

12 Changing bases

12.1 Change of basis formula

Recall that given a linear map $T\colon V \to W$ where V and W are real or complex vector spaces, and choices of bases $\{\mathbf{e}_i\}$ for $i = 1, \dots, n$ and $\{\mathbf{f}_a\}$ for $a = 1, \dots, m$, then the $m \times n$ matrix A with respect to these bases is defined by

$$T(\mathbf{e}_i) = \sum_a \mathbf{f}_a A_{ai}$$

So the entries in column i of A are the components of $T(\mathbf{e}_i)$ with respect to the basis $\{\mathbf{f}_a\}$. This is chosen to ensure that the statement $\mathbf{y} = T(\mathbf{x})$ is equivalent to the statement that $y_a = A_{ai}x_i$, where $\mathbf{y} = \sum_a y_a \mathbf{f}_a$ and $\mathbf{x} = \sum_i x_i \mathbf{e}_i$. This equivalence holds since

$$T\left(\sum_i x_i \mathbf{e}_i\right) = \sum_i x_i T(\mathbf{e}_i) = \sum_i x_i \left(\sum_a \mathbf{f}_a A_{ai}\right) = \sum_a \underbrace{\left(\sum_i A_{ai} x_i\right)}_{y_a} \mathbf{f}_a$$

as required. For the same linear map T, there is a different matrix representation A' with respect to different bases $\{\mathbf{e}_i'\}$ and $\{\mathbf{f}_a'\}$. To relate A with A', we need to understand how the new bases relate to the original bases. The change of base matrices P $(n \times n)$ and Q $(m \times m)$ are defined by

$$\mathbf{e}'_i = \sum_j \mathbf{e}_j P_{ji}; \quad \mathbf{f}'_a = \sum_b \mathbf{f}_b Q_{ba}$$

The entries in column i of P are the components of the new basis $\mathbf{e}'_i$ in terms of the old basis vectors $\{\mathbf{e}_j\}$, and similarly for Q. Note, P and Q are invertible, and in the relation above we could exchange the roles of $\{\mathbf{e}_i\}$ and $\{\mathbf{e}'_i\}$ by replacing P with P^{-1}, and similarly for Q.

Proposition (Change of base formula for a linear map)**.** With the definitions above,

$$A' = Q^{-1} A P$$

First we will consider an example, then we will construct a proof. Let $n = 2, m = 3$, and

$$T(\mathbf{e}_1) = \mathbf{f}_1 + 2\mathbf{f}_2 - \mathbf{f}_3 = \sum_a \mathbf{f}_a A_{a1}$$
$$T(\mathbf{e}_2) = -\mathbf{f}_1 + 2\mathbf{f}_2 + \mathbf{f}_3 = \sum_a \mathbf{f}_a A_{a2}$$

Therefore,

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

Consider a new basis for V, given by

$$\mathbf{e}'_1 = \mathbf{e}_1 - \mathbf{e}_2 = \sum_i \mathbf{e}_i P_{i1}$$
$$\mathbf{e}'_2 = \mathbf{e}_1 + \mathbf{e}_2 = \sum_i \mathbf{e}_i P_{i2}$$

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

Consider further a new basis for W, given by

$$\mathbf{f}'_1 = \mathbf{f}_1 - \mathbf{f}_3 = \sum_a \mathbf{f}_a Q_{a1}$$
$$\mathbf{f}'_2 = \mathbf{f}_2 = \sum_a \mathbf{f}_a Q_{a2}$$
$$\mathbf{f}'_3 = \mathbf{f}_1 + \mathbf{f}_3 = \sum_a \mathbf{f}_a Q_{a3}$$

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

From the change of base formula,

$$\begin{aligned}
A' &= Q^{-1}AP \\
&= \begin{pmatrix} 1/2 & 0 & -1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 2 & 2 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 2 & 0 \\ 0 & 4 \\ 0 & 0 \end{pmatrix}
\end{aligned}$$

Now checking this result directly,

$$\begin{aligned}
T(\mathbf{e}_1') &= 2\mathbf{f}_1 - 2\mathbf{f}_3 = 2\mathbf{f}_1' \\
T(\mathbf{e}_2') &= 4\mathbf{f}_2 = 4\mathbf{f}_2'
\end{aligned}$$

which matches the content of the matrix as required. Now, let us prove the proposition in general.

Proof.

$$\begin{aligned}
T(\mathbf{e}_i') &= T\left(\sum_j \mathbf{e}_j P_{ji}\right) \\
&= \sum_j T(\mathbf{e}_j) P_{ji} \\
&= \sum_j \left(\sum_a \mathbf{f}_a A_{aj}\right) P_{ji} \\
&= \sum_{ja} \mathbf{f}_a A_{aj} P_{ji}
\end{aligned}$$

But on the other hand,

$$\begin{aligned}
T(\mathbf{e}_i') &= \sum_b \mathbf{f}_b' A_{bi}' \\
&= \sum_b \left(\sum_a \mathbf{f}_a Q_{ab}\right) A_{bi}' \\
&= \sum_{ab} \mathbf{f}_a Q_{ab} A_{bi}'
\end{aligned}$$

which is a sum over the same set of basis vectors, so we may equate coefficients of $\mathbf{f}_a$.

$$\begin{aligned}
\sum_j A_{aj} P_{ji} &= \sum_b Q_{ab} A_{bi}' \\
(AP)_{ai} &= (QA')_{ai}
\end{aligned}$$

Therefore

$$AP = QA' \implies A' = Q^{-1}AP$$

as required. □

12.2 Changing bases of vector components

Here is another way to arrive at the formula $A' = Q^{-1}AP$. Consider changes in vector components

$$
\begin{aligned}
\mathbf{x} = \sum_i x_i \mathbf{e}_i &= \sum_j x'_j \mathbf{e}'_j \\
&= \sum_i \left(\sum_j P_{ij} x'_j \right) \mathbf{e}_i \\
\implies x_i &= P_{ij} x'_j
\end{aligned}
$$

We will write

$$
X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}; \quad X' = \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}
$$

Then $X = PX'$ or $X' = P^{-1}X$. Similarly,

$$
\begin{aligned}
\mathbf{y} = \sum_a y_a \mathbf{f}_a &= \sum_b y'_b \mathbf{f}'_b \\
\implies y_a &= Q_{ab} y'_b
\end{aligned}
$$

Then $Y = QY'$ or $Y' = Q^{-1}Y$. So the matrices are defined to ensure that

$$
Y = AX; \quad Y' = A'X'
$$

Therefore,

$$
QY' = APX' \implies Y' = (Q^{-1}AP)X' \implies A' = Q^{-1}AP
$$

12.3 Specialisations of changes of basis

Now, let us consider some special cases (in increasing order of specialisation).

(i) Let $V = W$ with $\mathbf{e}_i = \mathbf{f}_i$ and $\mathbf{e}'_i = \mathbf{f}'_i$. So $P = Q$ and the change of basis is

$$
A' = P^{-1}AP
$$

Matrices representing the same linear map but with respect to different bases are similar. Conversely, if A, A' are similar, then we can construct an invertible change of basis matrix P which relates them, so they can be regarded as representing the same linear map. In an earlier section we noted that $\operatorname{tr}(A') = \operatorname{tr}(A)$, $\det(A') = \det(A)$ and $\chi_A(t) = \chi_{A'}(t)$. so these are intrinsic properties of the linear map, not just the particular matrix we choose to represent it.

(ii) Let $V = W = \mathbb{R}^n$ or $\mathbb{C}^n$ where $\mathbf{e}_i$ is the standard basis, with respect to which, T has matrix A. If there exists a basis of eigenvectors, $\mathbf{e}'_i = \mathbf{v}_i$ with $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$. Then

$$
A' = P^{-1}AP = D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}
$$

and

$$
\mathbf{v}_i = \sum_k \mathbf{e}_j P_{ji}
$$

so the eigenvectors are the columns of P.

(iii) Let A be hermitian, i.e. $A^\dagger = A$, then we always have a basis of orthonormal eigenvectors $\mathbf{e}'_i = \mathbf{u}_i$. Then the relations in (ii) apply, and P is unitary, $P^\dagger = P^{-1}$.

12.4 Jordan normal form

Also known as the (Jordan) Canonical Form, this result classifies $n \times n$ complex matrices up to similarity.

Proposition. Any 2×2 complex matrix A is similar to one of the following:

(i) $A' = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ with $\lambda_1 \neq \lambda_2$, so $\chi_A(t) = (t-\lambda_1)(t-\lambda_2)$.

(ii) $A' = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$, so $\chi_A(t) = (t-\lambda)^2$.

(iii) $A' = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$, so $\chi_A(t) = (t-\lambda)^2$ as in case (ii).

Proof. $\chi_A(t)$ has two roots over $\mathbb{C}$.

(i) For distinct roots λ_1, λ_2, we have $M_{\lambda_1} = m_{\lambda_1} = M_{\lambda_2} = m_{\lambda_2} = 1$. So the eigenvectors $\mathbf{v}_1, \mathbf{v}_2$ provide a basis. Hence $A' = P^{-1}AP$ with the eigenvectors as the columns of P.

(ii) For a repeated root λ with $M_\lambda = m_\lambda = 2$, the same argument applies.

(iii) For a repeated root λ with $M_\lambda = 2$, $m_\lambda = 1$, we do not have a basis of eigenvectors so we cannot diagonalise the matrix. We only have one linearly independent eigenvector, which we will call $\mathbf{v}$. Let $\mathbf{w}$ be any other vector such that $\{\mathbf{v}, \mathbf{w}\}$ are linearly independent. Then

$$A\mathbf{v} = \lambda\mathbf{v}$$
$$A\mathbf{w} = \alpha\mathbf{v} + \beta\mathbf{w}$$

The matrix representing this linear map with respect to the basis vectors $\{\mathbf{v}, \mathbf{w}\}$ is therefore

$$\begin{pmatrix} \lambda & \alpha \\ 0 & \beta \end{pmatrix}$$

Let us solve for some of these unknowns. We know that the characteristic polynomial of this matrix must be $(t-\lambda)^2$, so $\beta = \lambda$. Also, $\alpha \neq 0$, otherwise we have case (ii) above. So now we can set $\mathbf{u} = \alpha\mathbf{v}$, so

$$A(\alpha\mathbf{v}) = \lambda(\alpha\mathbf{v})$$
$$A\mathbf{w} = \alpha\mathbf{v} + \beta\mathbf{w}$$

So with respect to the basis $\{\mathbf{u}, \mathbf{w}\}$ we get the matrix A to be

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

□

Alternative Proof. Here is an alternative approach for case (iii). If A has characteristic polynomial

$$\chi_A(t) = (t - \lambda)^2$$

but $A \neq \lambda I$, then there exists some vector $\mathbf{w}$ for which $\mathbf{u} = (A-\lambda I)\mathbf{w} \neq \mathbf{0}$. So $(A-\lambda I)\mathbf{u} = (A-\lambda I)^2\mathbf{w} = \mathbf{0}$ by the Cayley–Hamilton theorem. So

$$\begin{aligned} A\mathbf{u} &= \lambda\mathbf{u} \\ A\mathbf{w} &= \mathbf{u} + \lambda\mathbf{w} \end{aligned}$$

So with basis $\{\mathbf{u}, \mathbf{w}\}$ we have the matrix

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

□

Here is a concrete example using this alternative proof method.

$$A = \begin{pmatrix} 1 & 4 \\ -1 & 5 \end{pmatrix} \implies \chi_A(t) = (t-3)^2$$

So

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

We will choose $\mathbf{w} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and we find $\mathbf{u} = (A-3I)\mathbf{w} = \begin{pmatrix} -2 \\ -1 \end{pmatrix}$. $\mathbf{w}$ is not an eigenvector, as required for the construction. By the reasoning in the alternative argument above, $\mathbf{u}$ is an eigenvector by construction.

$$\begin{aligned} A\mathbf{u} &= 3\mathbf{u} \\ A\mathbf{w} &= \mathbf{u} + 3\mathbf{w} \end{aligned}$$

So

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

and we can check that

$$P^{-1}AP = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix} = A'$$

12.5 Jordan normal forms in n dimensions

To extend the arguments above to larger matrices, consider the $n \times n$ matrix

$$N = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

When applied to the standard basis vectors in $\mathbb{C}^n$, the action of this matrix sends $\mathbf{e}_n \mapsto \mathbf{e}_{n-1} \mapsto \cdots \mapsto \mathbf{e}_1 \mapsto \mathbf{0}$. This is consistent with the property that $N^n = 0$. The kernel of this matrix has dimension 1. Now consider the matrix $J = \lambda I + N$, as follows:

$$N = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda \end{pmatrix}$$

This matrix has

$$\chi_J(t) = (\lambda - t)^n$$

with $M_\lambda = n$ and $m_\lambda = 1$, since the kernel of $J - \lambda I = N$ has dimension 1 as before. The general result is as follows.

Theorem. Any $n \times n$ complex matrix A is similar to a matrix of the form

$$A' = \left(\begin{array}{c|c|c|c} J_{n_1}(\lambda_1) & 0 & \cdots & 0 \\ \hline 0 & J_{n_2}(\lambda_2) & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & J_{n_r}(\lambda_r) \end{array}\right)$$

where each diagonal block is a Jordan block $J_{n_r}(\lambda_r)$ which is an $n_r \times n_r$ matrix J with eigenvalue λ_r. $\lambda_1, \dots, \lambda_r$ are eigenvalues of A and A', and the same eigenvalue may appear in different blocks. Further, $n_1 + n_2 + \cdots + n_r = n$ so we end up with an $n \times n$ matrix. A is diagonalisable if and only if A' consists entirely of 1×1 blocks. The expression above is the Jordan Normal Form.

The proof is non-examinable and depends on the Part IB courses Linear Algebra, and Groups, Rings and Modules, so is not included here.

13 Conics and quadrics

13.1 Quadrics in general

A quadric in $\mathbb{R}^n$ is a hypersurface defined by an equation of the form

$$Q(\mathbf{x}) = \mathbf{x}^\intercal A\mathbf{x} + \mathbf{b}^\intercal \mathbf{x} + c = 0$$

for some nonzero, symmetric, real $n \times n$ matrix A, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$. In components,

$$Q(\mathbf{x}) = A_{ij}x_ix_j + b_ix_i + c = 0$$

We will classify solutions for $\mathbf{x}$ up to geometrical equivalence, so we will not distinguish between solutions here which are related by isometries in $\mathbb{R}^n$ (distance-preserving maps, i.e. translations and orthogonal transformations about the origin).

Note that A is invertible if and only if it has no zero eigenvalues. In this case, we can complete the square in the equation $Q(\mathbf{x}) = 0$ by setting $\mathbf{y} = \mathbf{x} + \frac{1}{2}A^{-1}\mathbf{b}$. This is essentially a translation isometry, moving the origin to $\frac{1}{2}A^{-1}\mathbf{b}$.

$$\begin{aligned}
\mathbf{y}^\intercal A\mathbf{y} &= (\mathbf{x} + \frac{1}{2}A^{-1}\mathbf{b})^\intercal A(\mathbf{x} + \frac{1}{2}A^{-1}\mathbf{b}) \\
&= (\mathbf{x}^\intercal + \frac{1}{2}\mathbf{b}^\intercal(A^{-1})^\intercal)A(\mathbf{x} + \frac{1}{2}A^{-1}\mathbf{b}) \\
&= \mathbf{x}^\intercal A\mathbf{x} + \mathbf{b}^\intercal\mathbf{x} + \frac{1}{4}\mathbf{b}^\intercal A^{-1}\mathbf{b}
\end{aligned}$$

since $(A^\intercal)^{-1} = (A^{-1})^\intercal$. Then,

$$Q(\mathbf{x}) = 0 \iff \mathcal{F}(\mathbf{y}) = k$$

with

$$\mathcal{F}(\mathbf{y}) = \mathbf{y}^\intercal A\mathbf{y}$$

which is a quadratic form with respect to a new origin $\mathbf{y} = \mathbf{0}$, and where $k = \frac{1}{4}\mathbf{b}^\intercal A^{-1}\mathbf{b} - c$. Now we can diagonalise $\mathcal{F}$ as in the above section, in particular, orthonormal eigenvectors give the principal axes, and the eigenvalues of A and the value of k determine the geometrical nature of the solution of the quadric. In $\mathbb{R}^3$, the geometrical possibilities are (as we saw before):

(i) eigenvalues positive, k positive gives an ellipsoid;

(ii) eigenvalues different signs, k nonzero gives a hyperboloid

If A has one or more zero eigenvalues, then the analysis we have just provided changes, since we can no longer construct such a $\mathbf{y}$ vector, since A^{-1} does not exist. The simplest standard form of Q may have both linear and quadratic terms.

13.2 Conics as quadrics

Quadrics in $\mathbb{R}^2$ are curves called conics. Let us first consider the case where $\det A \neq 0$. By completing the square and diagonalising A, we get a standard form

$$\lambda_1 {x_1'}^2 + \lambda_2 {x_2'}^2 = k$$

The variables x_i' correspond to the principal axes and the new origin. We have the following cases.

- ($\lambda_1, \lambda_2 > 0$) This is an ellipse for $k > 0$, and a point for $k = 0$. There are no solutions for $k < 0$.
- ($\lambda_1 > 0, \lambda_2 < 0$) This gives a hyperbola for $k > 0$, and a hyperbola in the other axis if $k < 0$. If $k = 0$, this is a pair of lines. For instance, ${x_1'}^2 - {x_2'}^2 = 0 \implies (x_1' - x_2')(x_1' + x_2') = 0$.

If $\det A = 0$, then there is exactly one zero eigenvalue since $A \neq 0$. Then:

- ($\lambda_1 > 0, \lambda_2 = 0$) We will diagonalise A in the original expression for the quadric. This gives

  $$\lambda_1 {x_1'}^2 + b_1' x_1' + b_2' x_2' + c = 0$$

  This is a new equation in the coordinate system defined by A's principal axes. Completing the square here in the x_1' term, we have

  $$\lambda_1 {x_1''}^2 + b_2' x_2' + c' = 0$$

where $x_1'' = x_1' + \frac{1}{2\lambda_1}b_1'$, and $c' = c - \frac{b_1'^2}{4\lambda_1^2}$. If $b_2' = 0$, then x_2 can take any value; and we get a pair of lines if $c' < 0$, a single line if $c' = 0$, and no solutions if $c' > 0$. Otherwise, $b_2' \neq 0$, and the equation becomes

$$\lambda_1 x_1''^2 + b_2' x_2'' = 0$$

where $x_2'' = x_2' + \frac{1}{b_2'}c'$, and clearly this equation is a parabola.

All changes of coordinates correspond to translations (shifts of the origin) or orthogonal transformations, both of which preserve distance and angles.

13.3 Standard forms for conics

The general forms of conics can be written in terms of lengths a, b (the semi-major and semi-minor axes), or equivalently a length scale ℓ and a dimensionless eccentricity constant e.

- First, let us consider Cartesian coordinates. The formulas are:

conic	formula	eccentricity	foci
ellipse	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$	$b^2 = a^2(1-e^2)$, and $e < 1$	$x = \pm ae$
parabola	$y^2 = 4ax$	one quadratic term vanishes, $e = 1$	$x = +a$
hyperbola	$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$	$b^2 = a^2(e^2-1)$, and $e > 1$	$x = \pm ae$

- Polar coordinates are a convenient alternative to Cartesian coordinates. In this coordinate system, we set the origin to be at a focus. Then, the formulas are

$$r = \frac{\ell}{1 + e\cos\theta}$$

  - For the ellipse, $e < 1$ and $\ell = a(1-e^2)$;
  - For the parabola, $e = 1$ and $\ell = 2a$; and
  - For the hyperbola, $e > 1$ and $\ell = a(e^2 - 1)$. There is only one branch for the hyperbola given by this polar form.

13.4 Conics as sections of a cone

The equation for a cone in $\mathbb{R}^3$ given by an apex $\mathbf{c}$, an axis $\hat{\mathbf{n}}$, and an angle $\alpha < \frac{\pi}{2}$, is

$$(\mathbf{x} - \mathbf{c}) \cdot \hat{\mathbf{n}} = |\mathbf{x} - \mathbf{c}| \cos\alpha$$

Less formally, the angle of $\mathbf{x}$ away from $\hat{\mathbf{n}}$ must be α. By squaring this equation, we can essentially define two cones which stretch out infinitely far and meet at the centre point $\mathbf{c}$.

$$((\mathbf{x} - \mathbf{c}) \cdot \hat{\mathbf{n}})^2 = |\mathbf{x} - \mathbf{c}|^2 \cos^2\alpha$$

Let us choose a set of coordinate axes so that our equations end up slightly easier. Let $\mathbf{c} = c\mathbf{e}_3, \hat{\mathbf{n}} = \cos\beta\mathbf{e}_1 - \sin\beta\mathbf{e}_3$. Then essentially the cone starts at $(0, 0, c)$ and points 'downwards' in the $\mathbf{e}_1$–$\mathbf{e}_3$ plane. Then the conic section is the intersection of this cone with the $\mathbf{e}_1$–$\mathbf{e}_2$ plane, i.e. $x_3 = 0$.

$$(x_1 \cos\beta - c\sin\beta)^2 = (x_1^2 + x_2^2 + c^2)\cos^2\alpha$$

$$\iff (\cos^2\alpha - \cos^2\beta)x_1^2 + (\cos^2\alpha)x_2^2 + 2x_1 c\sin\beta\cos\beta = \text{const.}$$

Now we can compare the signs of the x_1^2 and x_2^2 terms. Clearly the x_2^2 term is always positive, so we consider the sign of the x_1^2 term.

- If $\cos^2\alpha > \cos^2\beta$ (i.e. $\alpha < \beta$), then we have an ellipse.
- If $\cos^2\alpha = \cos^2\beta$ (i.e. $\alpha = \beta$), then we have a parabola.
- If $\cos^2\alpha < \cos^2\beta$ (i.e. $\alpha > \beta$), then we have a hyperbola.

14 Symmetries and transformation groups

14.1 Orthogonal transformations and rotations

We know that if a matrix R is orthogonal, we have $R^\intercal R = I \iff (R\mathbf{x})\cdot(R\mathbf{y}) = \mathbf{x}\cdot\mathbf{y} \iff$ the rows or columns are orthonormal. The set of $n\times n$ matrices R forms the orthogonal group $O_n = O(n)$. If $R \in O(n)$ then $\det R = \pm 1$. $SO_n = SO(n)$ is the special orthogonal group, which is the subgroup of $O(n)$ defined by $\det R = 1$. If some matrix R is an element of $O(n)$, then R preserves the modulus of n-dimensional volume. If $R \in SO(n)$, then R preserves not only the modulus but also the sign of such a volume.

$SO(n)$ consists precisely of all rotations in $\mathbb{R}^n$. $O(n)\setminus SO(n)$ consists of all reflections. For some specific $H \in O(n)\setminus SO(n)$, any element of $O(n)$ can be written as a product of H with some element in $SO(n)$, i.e. R or RH with $R \in SO(n)$. For example, if n is odd, we can choose $H = -I$.

Now, we can consider the transformation $x_i' = R_{ij}x_j$ under two distinct points of view.

- (active) The rotation R acts on the vector $\mathbf{x}$ and yields a new vector $\mathbf{x}'$. The x_i' are components of the transformed vector in terms of the standard basis vectors.
- (passive) The x_i' are components of the same vector $\mathbf{x}$ but with respect to new orthonormal basis vectors $\mathbf{u}_i$. In general, $\mathbf{x} = \sum_i x_i\mathbf{e}_i = \sum_i x_i'\mathbf{u}_i$ which is true where $\mathbf{u}_i = \sum_j R_{ij}\mathbf{e}_j = \sum_j \mathbf{e}_j P_{ji}$. So $P = R^{-1} = R^\intercal$ where P is the change of basis matrix.

14.2 2D Minkowski space

Consider a new 'inner product' on $\mathbb{R}^2$ given by

$$(\mathbf{x},\mathbf{y}) = \mathbf{x}^\intercal J\mathbf{y};\quad J = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$$

$$\therefore \left(\begin{pmatrix}x_0\\x_1\end{pmatrix}, \begin{pmatrix}y_0\\y_1\end{pmatrix}\right) = x_0y_0 - x_1y_1$$

We start indexing these vectors from zero, not one. Here are some important properties.

- This 'inner product' is not positive definite. In fact, $(\mathbf{x},\mathbf{x}) = x_0^2 - x_1^2$. (This is a quadratic form for $\mathbf{x}$ with eigenvalues ± 1.)
- It is bilinear and symmetric.
- Defining $\mathbf{e}_0 = \begin{pmatrix}1\\0\end{pmatrix}$ and $\mathbf{e}_1 = \begin{pmatrix}0\\1\end{pmatrix}$, they obey

$$(\mathbf{e}_0,\mathbf{e}_0) = -(\mathbf{e}_1,\mathbf{e}_1) = 1;\quad (\mathbf{e}_0,\mathbf{e}_1) = 0$$

This is similar to orthonormality, in this generalised sense.

This inner product is known as the Minkowski metric on $\mathbb{R}^2$. $\mathbb{R}^2$ with this metric is called Minkowski space.

14.3 Lorentz transformations

Let us consider a matrix

$$M = \begin{pmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{pmatrix}$$

giving a map $\mathbb{R}^2 \to \mathbb{R}^2$; this preserves the Minkowski metric if and only if $(M\mathbf{x}, M\mathbf{y}) = (\mathbf{x}, \mathbf{y})$ for any vectors $\mathbf{x}, \mathbf{y}$. Expanded, this condition is

$$(M\mathbf{x})^\intercal J(M\mathbf{y}) = \mathbf{x}^\intercal M^\intercal J M\mathbf{y} = \mathbf{x}^\intercal J\mathbf{y}$$

$$\implies M^\intercal J M = J$$

The set of such matrices form a group. Also, $\det M = \pm 1$ for the same reason as before. Furthermore, $|M_{00}|^2 \geq 1$, so either $M_{00} \geq 1$ or $M_{00} \leq -1$. The subgroup with $\det M = +1$ and $M_{00} \geq 1$ is known as the Lorentz group.

Let us find the general form of M, by using the fact that the columns $M\mathbf{e}_0$ and $M\mathbf{e}_i$ are orthonormal with respect to the Minkowski metric.

$$(M\mathbf{e}_0, M\mathbf{e}_0) = M_{00}^2 - M_{10}^2 = (\mathbf{e}_0, \mathbf{e}_0) = 1 \quad (\text{hence } |M_{00}|^2 \geq 1)$$

Taking $M_{00} \geq 1$, we can write

$$M\mathbf{e}_0 = \begin{pmatrix} \cosh\theta \\ \sinh\theta \end{pmatrix}$$

for some real value θ. For the other column,

$$(M\mathbf{e}_0, M\mathbf{e}_1) = 0;\ (M\mathbf{e}_1, M\mathbf{e}_1) = -1 \implies M\mathbf{e}_1 = \pm\begin{pmatrix} \sinh\theta \\ \cosh\theta \end{pmatrix}$$

The sign is fixed to be positive by the condition that $\det M = +1$.

$$M = \begin{pmatrix} \cosh\theta & \sinh\theta \\ \sinh\theta & \cosh\theta \end{pmatrix}$$

The curves defined by $(\mathbf{x}, \mathbf{x}) = k$ where k is a constant are hyperbolas. This is analogous to how the curves defined by $\mathbf{x} \cdot \mathbf{x} = k$ are circles. So applying M to any vector on a given branch of a hyperbola, the resultant vector remains on the hyperbola. Note that these matrices obey the rule $M(\theta_1)M(\theta_2) = M(\theta_1 + \theta_2)$. This confirms that they form a group.

14.4 Application to special relativity

Let

$$M(\theta) = \gamma(v)\begin{pmatrix} 1 & v \\ v & 1 \end{pmatrix}; \quad v = \tanh\theta; \quad \gamma = (1 - v^2)^{-\frac{1}{2}}$$

Here, v lies in the range $-1 < v < 1$. We will rename x_0 to be t, which is now our time coordinate. x_1 will just be written x, our one-dimensional space coordinate. Then,

$$\mathbf{x}' = M\mathbf{x} \iff \begin{cases} t' &= \gamma \cdot (t + vx) \\ x' &= \gamma \cdot (x + vt) \end{cases}$$

This is a Lorentz transformation, or 'boost', relating the time and space coordinates for observers moving with relative velocity v in Special Relativity, in units where the speed of light c is taken to be 1. The γ factor in the Lorentz transformation gives rise to time dilation and length contraction effects. The group property $M(\theta_3) = M(\theta_1)M(\theta_2)$ with $\theta_3 = \theta_1 + \theta_2$ corresponds to the velocities

$$v_i = \tanh\theta_i \implies v_3 = \frac{v_1 + v_2}{1 + v_1 v_2}$$

This is consistent with the fact that all velocities are less than the speed of light, 1.