Natural Sciences Tripos Part IA
Mathematics III (B course)
Easter 2006

Professor R.R. Horgan

April 21, 2006

Contents

1 Linear Algebra

1.1 Linear vector spaces

The idea of a **linear vector space** is central to the analysis of many problems in physics and mathematics and it is the basic object of study in **linear algebra**. In particular, it applies to the study of

- linear simultaneous equations. This involves the study of matrices and their properties;
- the solutions to linear partial (and ordinary) differential equations abbreviated to PDE (ODE).

Physical problems that can be tackled include

- the harmonic vibrations of a system about an equilibrium and the natural frequencies of oscillation. E.g., molecules and vibrational frequencies of absorption of radiation;
- waves in various media;
- problems in diffusion;
- the electrostatic potential of charge distributions;
- Fourier series;
- quantum mechanics.

In the first part of this course we will concentrate on **linear algebra** applied to matrices but it is important to understand that we are discussing a particular kind of realization, or representation, of a linear vector space and that there are many others. For this reason, it is important to give a formal definition.

1.1.1 Definition of a linear vector space

Notation:

V: a set of elements denoted by bold letters: $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}, \boldsymbol{v}$ etc..

K: a field consisting of elements called **scalars**, denoted by unbold letters: a, b, c, k etc.. For us these will be real or complex numbers.

Rules:

- **addition:** This is a binary operation denoted "+". To any $\boldsymbol{x}, \boldsymbol{y} \in V$ this rule assigns an element $\boldsymbol{z} \in V$: $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}$.
- **scalar multiplication:** To any $a \in K$ and $\boldsymbol{x} \in V$ this rule assigns an element $\boldsymbol{z} \in V$: $\boldsymbol{z} = a\boldsymbol{x}$.

Definition. V is called a **vector space over** K, and the elements of V are called **vectors**, if the following **axioms** hold:

A1 For any vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$, $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$. (**Associativity.**)

A2 For any vectors $\boldsymbol{u}, \boldsymbol{v} \in V$, $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$. (**Commutativity.**)

A3 There is a vector in V denoted $\mathbf{0}$, called the **zero vector** for which $\boldsymbol{u} + \mathbf{0} = \boldsymbol{u}$ $\forall\, \boldsymbol{u} \in V$.

A4 For each vector $\boldsymbol{u} \in V$ there is a vector in V denoted $-\boldsymbol{u}$ for which $\boldsymbol{u} + (-\boldsymbol{u}) = \mathbf{0}$. (**Inverse.**)

A5 For any $a \in K$ and any $\boldsymbol{u}, \boldsymbol{v} \in V$, $a(\boldsymbol{u} + \boldsymbol{v}) = a\boldsymbol{u} + a\boldsymbol{v}$.

A6 For any $a, b \in K$ and any $\boldsymbol{u} \in V$, $(a + b)\boldsymbol{u} = a\boldsymbol{u} + b\boldsymbol{u}$.

A7 For any $a, b \in K$ and any $\boldsymbol{u} \in V$, $(ab)\boldsymbol{u} = a(b\boldsymbol{u})$.

A8 For the unit scalar $1 \in K$ and any $\boldsymbol{u} \in V$, $1\boldsymbol{u} = \boldsymbol{u}$.

Other results follow from these axioms. E.g.,

$$0\boldsymbol{u} = \mathbf{0}, \qquad a\mathbf{0} = \mathbf{0}, \qquad (-a)\boldsymbol{u} = -a\boldsymbol{u}, \qquad a\boldsymbol{u} = \mathbf{0} \implies a = 0 \text{ or } \boldsymbol{u} = \mathbf{0}.$$

1.1.2 Examples of vector spaces

i) Let K be an arbitrary field. A vector space is the set of all n-tuples of elements of K with vector addition and scalar multiplication defined by

$$\begin{aligned} (a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n) &= (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n)\ , \\ k(a_1, a_2, \ldots, a_n) &= (ka_1, ka_2, \ldots, ka_n)\ , \end{aligned}$$

where $a_i, b_i, k \in K$. This space is denoted K^n.

ii) The set of all n-tuples of real numbers $(u_1, \ldots, u_n)$, denoted $\mathcal{R}^n$, is a vector space over the field $\mathcal{R}$. This follows as an example of i). Likewise, the set of all n-tuples of complex numbers $(z_1, \ldots, z_n)$, denoted $\mathcal{C}^n$, is a vector space over the field $\mathcal{C}$. Examples of vectors in $\mathcal{R}^3$ are

$$(1,\ 2,\ 5)\ , \qquad (-0.5,\ 6.3,\ 234.8)\ , \qquad (0,\ 0,\ 0)\ .$$

The last of these is the zero, or null, vector $\mathbf{0}$.

iii) V is the set of all polynomials in t of degree $\leq n$

$$a_0 + a_1 t + a_2 t^2 + \ldots + a_n t^n\ ,$$

with coefficients a_i from a field K. V is a vector space over K with respect to the usual operations of addition of polynomials and multiplication by a constant.

1.1.3 Linear combinations and linear spans

Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m \in V$ and $a_1, \ldots, a_m \in K$ and let

$$\boldsymbol{x} = a_1 \boldsymbol{v}_1 + \ldots + a_m \boldsymbol{v}_m\ .$$

Then $\boldsymbol{x}$ is called a **linear combination** of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$.

The set of all such linear combinations of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ is a subspace, S, of V. In other words, S contains all vectors of the form of $\boldsymbol{x}$ above that are **generated** by all possible choices of $a_1, \ldots, a_m \in K$. This is written

$$S = \{a_1 \boldsymbol{v}_1 + \ldots + a_m \boldsymbol{v}_m :\ a_i \in K,\ i = 1, 2, \ldots, m\}\ .$$

Then we say that the subspace S is **spanned** or **generated** by the $\boldsymbol{v}$'s, and that the $\boldsymbol{v}$'s **span** or **generate** S.

1.1.4 Linear independence

Suppose that for some $a_1, \ldots, a_m \in K$ we have

$$a_1 \boldsymbol{v}_1 + \ldots + a_m \boldsymbol{v}_m = \mathbf{0}\ ,$$

Then the vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ are said to be **linearly independent** if the only solution is $a_i = 0, \forall i$.

Conversely, if there **is** a solution with at least one of the a's non-zero then the vectors are **linearly dependent**. Note, that if any of the $\boldsymbol{v}$'s is the zero vector, $\mathbf{0}$, then the vectors are linearly dependent.

1.1.5 Dimension and basis

A vector space V is said to be of **finite dimension n** or to be **n-dimensional**, written $\dim V = n$, if there exist linearly independent vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ which span V. That is, every $\boldsymbol{v} \in V$ can be written as a linear combination of the $\boldsymbol{e}$'s. The sequence $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ is then called a **basis** of V.

Note that a set of vectors might **span** V but they do not necessarily form a basis since they might not be linearly independent. However, given such a set we can systematically reduce the number of elements until we do have an independent set which then will form a basis.

The definition of dimension is well defined because it can be shown that every basis of V has the **same number** of elements.

1.1.6 Examples of bases

1. A basis for K^3 over the field K is

$$\boldsymbol{e}_1 = (1,0,0)\ , \qquad \boldsymbol{e}_2 = (0,1,0)\ , \qquad \boldsymbol{e}_3 = (0,0,1)\ .$$

An alternative basis is

$$\boldsymbol{w}_1 = (1,1,0)\ , \qquad \boldsymbol{w}_2 = (1,0,1)\ , \qquad \boldsymbol{w}_3 = (0,1,1)\ .$$

2. Let W be the vector space of polynomials in t of degree $\leq n$. The set $\{1, t, t^2, \ldots, t^n\}$ is linearly independent and spans W. Thus it is a basis of W and so $\dim W = n+1$. A different basis when, e.g., $n = 3$ is $\{1 + 2t^2, t + t^2, t^2 - 1\}$.

1.1.7 Coordinates

Given a basis $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ for V, then any vector $\boldsymbol{v} \in V$ can be expressed as

$$\boldsymbol{v} = x_1\boldsymbol{e}_1 + \ldots + x_n\boldsymbol{e}_n\ , \qquad x_i \in K\ .$$

Then the n-tuple $\boldsymbol{x} = (x_1, \ldots, x_n)$ are the **coordinates** of $\boldsymbol{v}$ with respect to the given basis. If we change the basis the coordinates will change but, of course, $\boldsymbol{v}$ is still the same vector:

$$\boldsymbol{v} = y_1\boldsymbol{w}_1 + \ldots + y_n\boldsymbol{w}_n\ , \qquad y_i \in K\ ,$$

with coordinates $\boldsymbol{y} = (y_1, \ldots, y_n)$.

Note that $\boldsymbol{x}, \boldsymbol{y}$ are themselves vectors since $\boldsymbol{x}, \boldsymbol{y} \in K^n$. K^n is a vector space over the field K defined earlier.

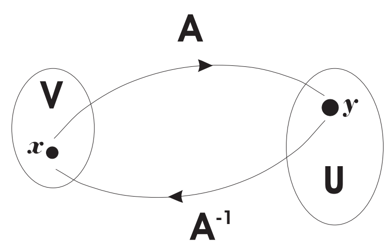
1.1.8 Linear maps

A **mapping** $\boldsymbol{A}$ of a vector space V into a vector space U assigns to any vector $\boldsymbol{x} \in V$ another vector $\boldsymbol{y} \in U$. We write either

$$\boldsymbol{A} : \boldsymbol{x} \rightarrow \boldsymbol{y}, \qquad \text{or} \qquad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}.$$

There might be an inverse (does not always exist), $\boldsymbol{A}^{-1}$ defined by

$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{y} \, .$$



The map is **linear** if it satisfies the following properties

(i) $\boldsymbol{A}(\boldsymbol{x}_1 + \boldsymbol{x}_2) = \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{A}\boldsymbol{x}_2$ for every $\boldsymbol{x}_1, \boldsymbol{x}_2 \in V$.

(ii) $\boldsymbol{A}(\alpha\boldsymbol{x}) = \alpha\boldsymbol{A}\boldsymbol{x}$ for every $\boldsymbol{x} \in V$ and every scalar α in K.

Examples of linear maps

- $\boldsymbol{A}\boldsymbol{x} = a\boldsymbol{x}$.
- For position vectors in 3D: $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{a} \wedge \boldsymbol{x}$, $\boldsymbol{a}$ a constant vector, (where "$\wedge$" is vector product).
- For position vectors in 3D: $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{a} \cdot \boldsymbol{x}$, $\boldsymbol{a}$ a constant vector, (where "$\cdot$" is scalar, or dot, product). Note that here $\boldsymbol{A}$ maps $V = \mathcal{R}^3$ into $U = \mathcal{R}$.

Examples of non-linear maps

(i) $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x} + \boldsymbol{a}$. ($|\boldsymbol{x}|$ is the length of $\boldsymbol{x}$.)

(ii) For position vectors in 3D: $\boldsymbol{A}\boldsymbol{x} = a|\boldsymbol{x}|\boldsymbol{x}$.

1.2 Matrices

A matrix is a rectangular array of real or complex numbers. We shall mainly use real numbers in this course but complex matrices are central to many applications. Examples are

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \qquad \begin{pmatrix} 3.6 & 2.4 \\ 9.3 & -4.5 \end{pmatrix} \qquad \begin{pmatrix} 1 \\ 3 \\ 7 \\ 9 \end{pmatrix} \qquad \begin{pmatrix} 1 & 3 & 7 & 9 \end{pmatrix} .$$

Notation

An $m \times n$ matrix has m rows and n columns. The matrix is usually denoted by a bold upper case letter, $\boldsymbol{A}$, say, and then a_{ij} will denote the jth entry in the ith row:

$$\boldsymbol{A} = (a_{ij}), \qquad (\boldsymbol{A})_{ij} = a_{ij}, \qquad \boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix}$$

Notes

i) An $m \times m$ matrix is called a **square** matrix.

ii) An $m \times 1$ matrix is a **column vector**:

$$\boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix} .$$

iii) A $1 \times n$ matrix is a **row vector**:

$$\boldsymbol{w} = \begin{pmatrix} w_1 & w_2 & \cdots & w_n \end{pmatrix} .$$

In what follows it is often useful to understand a general statement by working through the most simple non-trivial example. E.g., choose the smallest matrices to illustrate the point.

1.2.1 Algebra of matrices

For given n, m the set of all real (complex) $n \times m$ matrices form a vector space over $\mathcal{R}$ ($\mathcal{C}$). We need a rule of addition (+) and multiplication by a scalar which we make explicit in (a) and (b) below.

(a) **Addition of matrices**

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $n \times m$ matrices. Their sum $\boldsymbol{C}$ is an $n \times m$ matrix defined by

$$\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B} \quad \text{with} \quad c_{ij} = a_{ij} + b_{ij} \; .$$

E.g.,

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} 8 & 10 & 12 \\ 14 & 16 & 18 \end{pmatrix}$$

(b) **Multiplication by a scalar**

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $n \times m$ matrices.

$$\boldsymbol{B} = \lambda \boldsymbol{A} \quad \text{means} \quad b_{ij} = \lambda a_{ij} \quad i = 1 \ldots m, \; j = 1 \ldots n.$$

The statement of equality of matrices follows if we set $\lambda = 1$:

$$\boldsymbol{B} = \boldsymbol{A} \quad \text{means} \quad b_{ij} = a_{ij} \quad i = 1 \ldots m, \; j = 1 \ldots n.$$

(c) **Multiplication of matrices**

Matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ can only be multiplied if $\boldsymbol{A}$ is $m \times n$ and $\boldsymbol{B}$ is $n \times p$. Then

$$\boldsymbol{C} \;=\; \boldsymbol{AB}$$

is defined by

$$c_{ij} \;=\; \sum_{k=1}^{n} a_{ik} b_{kj} \qquad i = 1 \ldots m, \; j = 1 \ldots p,$$

and the product matrix $\boldsymbol{C}$ is $m \times p$.

An important fact is that the product of two square matrices does not **commute**. Suppose $\boldsymbol{A}$ and $\boldsymbol{B}$ are both $m \times m$. Then in general $\boldsymbol{AB} \neq \boldsymbol{BA}$. If this is the case we say that $\boldsymbol{A}$ and $\boldsymbol{B}$ do not **commute**. The **commutator** is defined by:

$$\boldsymbol{C} = [\boldsymbol{A}, \boldsymbol{B}] \equiv \boldsymbol{AB} - \boldsymbol{BA} \; .$$

Of course, $\boldsymbol{C}$ is also $m \times m$.

Examples of multiplication

$$\begin{pmatrix} 1 & 4 & -2 \end{pmatrix} \begin{pmatrix} 6 \\ 4 \\ 1 \end{pmatrix} = \begin{pmatrix} 20 \end{pmatrix}$$

$$\begin{pmatrix} 6 \\ 4 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 4 & -2 \end{pmatrix} = \begin{pmatrix} 6 & 24 & -12 \\ 4 & 16 & -8 \\ 1 & 4 & -2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix} , \qquad \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 23 & 34 \\ 31 & 46 \end{pmatrix}$$

Note the 2×2 matrices here do not commute.

$$\begin{pmatrix} 9 & 8 & 6 \\ 4 & 3 & 2 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 6 \\ -4 \\ 1 \end{pmatrix} = \begin{pmatrix} 28 \\ 14 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 6 & -4 & 1 \end{pmatrix} \begin{pmatrix} 9 & 8 & 6 \\ 4 & 3 & 2 \\ 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 39 & 38 & 31 \end{pmatrix}$$

1.2.2 Some definitions and properties

(a) **Transpose**

The **transpose** of an $m \times n$ matrix $\boldsymbol{M}$ is the $n \times m$ matrix denoted $\boldsymbol{M}^T$ given by the interchange of the rows and columns of $\boldsymbol{M}$:

$$(\boldsymbol{M}^T)_{ij} = (\boldsymbol{M})_{ji} \, , \qquad \text{for all} \;\; i, j.$$

Note that

(i) $(\boldsymbol{M}^T)^T = \boldsymbol{M}$.

(ii)

$$(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$$

(i,j) element:

$$\sum_k a_{jk}b_{ki} = \sum_k b_{ki}a_{jk} = \sum_k (\boldsymbol{B}^T)_{ik}(\boldsymbol{A}^T)_{kj}$$

This result generalizes: $(\boldsymbol{ABC})^T = \boldsymbol{C}^T\boldsymbol{B}^T\boldsymbol{A}^T$ etc.

(b) **Symmetric and anti-symmetric matrices**

We define a **symmetric** matrix $\boldsymbol{S}$ to be a square matrix which satisfies $\boldsymbol{S}^T = \boldsymbol{S}$. Thus

$$s_{ij} = s_{ji}.$$

We define an **anti-symmetric** (or **skew-symmetric**) matrix $\boldsymbol{A}$ to be a square matrix which satisfies $\boldsymbol{A}^T = -\boldsymbol{A}$. Thus

$$a_{ij} = -a_{ji}.$$

Given an general $m \times m$ matrix $\boldsymbol{B}$ we can construct its symmetric and anti-symmetric parts given, respectively, by $\boldsymbol{S}$ and $\boldsymbol{A}$ to be

$$\boldsymbol{S} = \frac{1}{2}(\boldsymbol{B} + \boldsymbol{B}^T), \qquad \boldsymbol{A} = \frac{1}{2}(\boldsymbol{B} - \boldsymbol{B}^T).$$

Conversely, we may always decompose $\boldsymbol{B}$ as the sum of a symmetric matrix and an anti-symmetric matrix: $\boldsymbol{B} = \boldsymbol{S} + \boldsymbol{A}$.

(c) **Diagonal matrix**

A square matrix $\boldsymbol{A}$ with non-zero entries only on the diagonal: $a_{ij} = 0 \quad i \neq j$. E.g.,

$$\begin{pmatrix} 1.2 & 0 & 0 \\ 0 & -3.4 & 0 \\ 0 & 0 & 7.6 \end{pmatrix}$$

(d) **Unit matrix**

A diagonal matrix denoted $\mathbf{1}$ or $\boldsymbol{I}$ with elements denoted δ_{ij}, called the **Kroneka delta**, where $\delta_{ii} = 1, \delta_{ij} = 0 \quad i \neq j$. E.g., for $n = 3$

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

For any matrix $\boldsymbol{A}$ we have $\boldsymbol{IA} = \boldsymbol{AI} = \boldsymbol{A}$.

(e) **Orthogonal matrix**

A square matrix $\boldsymbol{O}$ which satisfies $\boldsymbol{O}\boldsymbol{O}^T = \boldsymbol{O}^T\boldsymbol{O} = \boldsymbol{I}$. E.g.,

$$\boldsymbol{R}(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

More on this soon.

(f) **Complex conjugation**

If $\boldsymbol{A} = (a_{ij})$ then the **complex conjugate** is $\boldsymbol{A}^* = (a_{ij}^*)$.

(g) **Hermitian conjugation**

If $\boldsymbol{A} = (a_{ij})$ then the **hermitian conjugate** is $\boldsymbol{A}^\dagger = (\boldsymbol{A}^T)^* = (\boldsymbol{A}^*)^T = (a_{ji}^*)$.

An **hermitian** matrix satisfies $\boldsymbol{A}^\dagger = \boldsymbol{A}$ (c.f. symmetric matrix) and is important in quantum mechanics.

(h) **Trace**

The **trace** of a matrix is defined for square matrices. For $\boldsymbol{A}$, $m \times m$, we have

$$\text{trace}(\boldsymbol{A}) = \sum_{i=1}^{m} a_{ii} \,.$$

It is the sum of the elements on the main diagonal of the matrix.

Some properties of **trace** are:

(i)

$$\begin{aligned} \text{trace}(\boldsymbol{AB}) &= \text{trace}(\boldsymbol{BA}) \\ \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ji} &= \sum_{j=1}^{n}\sum_{i=1}^{m} b_{ji}a_{ij} \end{aligned}$$

Sufficient that $\boldsymbol{A}$ is $n \times m$ and $\boldsymbol{B}$ is $m \times n$.

(ii)

$$\begin{aligned} \text{trace}(\boldsymbol{ABC}) &= \text{trace}(\boldsymbol{CAB}) \\ \sum_{ijk} a_{ij}b_{jk}c_{ki} &= \sum_{kij} c_{ki}a_{ij}b_{jk} \end{aligned}$$

(iii) This result can be generalized and holds for any **cyclic** permutation of the order of multiplication. For example

$$\text{trace}(\boldsymbol{A}_1\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{A}_4) = \text{trace}(\boldsymbol{A}_3\boldsymbol{A}_4\boldsymbol{A}_1\boldsymbol{A}_2) = \text{etc.}$$

A **cyclic** permutation shifts all elements by a given amount with those elements shifted off one end being inserted at the other. E.g., $12345 \to 45123$). (It's like moving the numbers around a clockface.)

1.2.3 Inner or scalar product

Can introduce a product of two vectors $\boldsymbol{x}, \boldsymbol{y}$ called the **inner** or **scalar** product. (It can be defined for many kinds of vector space but it not part of the axioms defining them; it is an extra optional property.) Give well-known examples:

- For column vectors $\boldsymbol{x}, \boldsymbol{y}$ real

$$\boldsymbol{x} \cdot \boldsymbol{y} \;=\; \sum_{i=1}^{n} x_i y_i \;=\; \boldsymbol{x}^T \boldsymbol{y} \; .$$

- For column vectors $\boldsymbol{x}, \boldsymbol{y}$ complex

$$\boldsymbol{x} \cdot \boldsymbol{y} \;=\; \sum_{i=1}^{n} x_i^* y_i \;=\; \boldsymbol{x}^\dagger \boldsymbol{y} \; .$$

  Note: in this case $\boldsymbol{x} \cdot \boldsymbol{y} = (\boldsymbol{y} \cdot \boldsymbol{x})^*$.

Then $\boldsymbol{x} \cdot \boldsymbol{x} = \sum_i |x_i|^2$ is real and positive.
We define the **magnitude** of $\boldsymbol{x}$ to be $||\boldsymbol{x}|| = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}}$.

(i) If $\boldsymbol{x} \cdot \boldsymbol{y} = 0$ then $\boldsymbol{x}$ and $\boldsymbol{y}$ are **orthogonal**.

(ii) A basis $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n$ which satisfies $\boldsymbol{e}_i \cdot \boldsymbol{e}_i = 1$, $\boldsymbol{e}_i \cdot \boldsymbol{e}_j = 0, \quad i \neq j$ is **orthonormal**. Write as

$$\boldsymbol{e}_i \cdot \boldsymbol{e}_j \;=\; \delta_{ij} \; .$$

E.g., in 2D

$$\boldsymbol{e}_1 = (1, 0), \quad \boldsymbol{e}_2 = (0, 1) \; .$$

1.2.4 Relevance to linear equations

The system of linear algebraic equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= y_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= y_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= y_m \end{aligned}$$

can be written compactly using matrix notation as

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y} \ .$$

1. The equations relate an n-dim column vector $\boldsymbol{x}$ to an m-dim column vector $\boldsymbol{y}$.

2. They may be viewed as defining a linear transformation from an n-dimensional vector space V_n to an m-dimensional vector space V_m.

   The problem of solving the equations can be viewed as finding the vector $\boldsymbol{x} \in V_n$ which is mapped under the transformation $\boldsymbol{A}$ to the vector $\boldsymbol{y} \in V_m$. This may not always be possible or there may not always be unique solution for $\boldsymbol{x}$. Usually $m = n$ but the interpretation applies more generally.

1.2.5 Summation convention

It is often convenient to employ the **summation convention** which is that the appearance of any repeated suffix in a formula automatically implies summation over that suffix. Thus,

$$\sum_{k=1}^{n} a_{ik}b_{kj} \quad \text{is written just as} \quad a_{ik}b_{kj} \ .$$

A more complicated example would be

$$Q_{ij}P_{jkkl}R_{lm} \equiv \sum_j \sum_k \sum_l Q_{ij}P_{jkkl}R_{lm} \ .$$

It is important that no suffix occurs more than twice.

This convention will *not* be used in these lectures so that, for example, a_{ii} will mean a_{11} or a_{22} etc., and not $(a_{11} + a_{22} + \cdots)$.

Whatever convention we use, we call the sum over pairs of indices a **contraction** of the indices.

1.3 Determinants

1.3.1 Definition

The solution of the linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= y_1 \\ a_{21}x_1 + a_{22}x_2 &= y_2 \end{aligned}$$

can be written

$$\frac{x_1}{y_1a_{22} - y_2a_{12}} = \frac{x_2}{y_2a_{11} - y_1a_{21}} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}}$$

(provided no denominator vanishes), or more neatly as

$$\frac{x_1}{\begin{vmatrix} y_1 & a_{12} \\ y_2 & a_{22} \end{vmatrix}} = \frac{x_2}{\begin{vmatrix} a_{11} & y_1 \\ a_{21} & y_2 \end{vmatrix}} = \frac{1}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} ,$$

where the **determinant** is defined as

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc .$$

Similarly, for 3 equations in 3 unknowns:

$$\frac{x_1}{\begin{vmatrix} y_1 & a_{12} & a_{13} \\ y_2 & a_{22} & a_{23} \\ y_3 & a_{23} & a_{33} \end{vmatrix}} = \frac{x_2}{\begin{vmatrix} a_{11} & y_1 & a_{13} \\ a_{21} & y_2 & a_{23} \\ a_{13} & y_3 & a_{33} \end{vmatrix}} = \frac{x_3}{\begin{vmatrix} a_{11} & a_{12} & y_1 \\ a_{21} & a_{22} & y_2 \\ a_{31} & a_{32} & y_3 \end{vmatrix}} = \frac{1}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{23} & a_{33} \end{vmatrix}} .$$

Spot the rule. The denominator of x_i is the determinant got by replacing the i-th column of (a_{ij}) by $\boldsymbol{y}$.

The 3×3 determinant is defined by

$$\begin{vmatrix} a & b & c \\ c & d & e \\ f & g & h \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} .$$

The general rule is defined recursively and to do this we first define **minors** and **cofactors**.

1.3.2 Minors and cofactors

Consider the square $n \times n$ matrix $\boldsymbol{A} = (a_{ij})$. Let $\boldsymbol{M}_{ij}$ be the $(n-1) \times (n-1)$ submatrix of $\boldsymbol{A}$ obtained by deleting its ith row and jth column.

The determinant $|\boldsymbol{M}_{ij}|$ is called the **minor** of the element a_{ij} of $\boldsymbol{A}$.

The **cofactor** of a_{ij}, denoted A_{ij} is the "signed" minor:

$$A_{ij} = (-1)^{i+j} |\boldsymbol{M}_{ij}| \; .$$

The "signs" $(-1)^{i+j}$ form chess-board pattern with +'s on the main diagonal:

$$\begin{pmatrix} + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ + & - & + & - & \cdots \\ & & \textit{etc.} & & \end{pmatrix}$$

The matrix with the cofactors as its elements is called the **classical adjoint** of $\boldsymbol{A}$ denoted adj$\boldsymbol{A}$. It is defined by $(\mathrm{adj}\boldsymbol{A})_{ij} = A_{ji}$:

$$\mathrm{adj}\boldsymbol{A} = \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{j1} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{j2} & \cdots & A_{n2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{1i} & A_{2i} & \cdots & A_{ji} & \cdots & A_{ni} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{1m} & A_{2m} & \cdots & A_{jm} & \cdots & A_{nm} \end{pmatrix}$$

1.3.3 General rule for calculating a determinant

Given the square $n \times n$ matrix $\boldsymbol{A}$ then the **determinant** of $\boldsymbol{A}$, denoted $|\boldsymbol{A}|$ or det $\boldsymbol{A}$, is defined by

$$|\boldsymbol{A}| = \sum_{j=1}^{n} a_{ij} A_{ij} \qquad \text{for } \textbf{any} \text{ fixed value of } \; i,$$

or

$$|\boldsymbol{A}| = \sum_{i=1}^{n} a_{ij} A_{ij} \qquad \text{for } \textbf{any} \text{ fixed value of } \; j.$$

Examples

Let

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ -1 & 4 & 5 \\ 2 & 7 & 8 \end{pmatrix} \; .$$

Choosing to fix $i = 1$ then

$$
\begin{aligned}
|\boldsymbol{A}| &= a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13} \\
&= 1 \cdot \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} - 2 \cdot \begin{vmatrix} -1 & 5 \\ 2 & 8 \end{vmatrix} + 3 \cdot \begin{vmatrix} -1 & 4 \\ 2 & 7 \end{vmatrix} \\
&= 1[32 - 35] - 2[-8 - 10] + 3[-7 - 8] \;=\; -12
\end{aligned}
$$

Or fixing $j = 2$ get

$$
\begin{aligned}
|\boldsymbol{A}| &= a_{12}A_{12} + a_{22}A_{22} + a_{32}A_{32} \\
&= -2 \cdot \begin{vmatrix} -1 & 5 \\ 2 & 8 \end{vmatrix} + 4 \cdot \begin{vmatrix} 1 & 3 \\ 2 & 8 \end{vmatrix} - 7 \cdot \begin{vmatrix} 1 & 3 \\ -1 & 5 \end{vmatrix} \\
&= -2[-8 - 10] + 4[8 - 6] - 7[5 + 3] \;=\; -12
\end{aligned}
$$

The neat way is to pick a row or column with the most zeros.

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

Then choose $j = 2$ and get

$$
|\boldsymbol{A}| = -7 \cdot \begin{vmatrix} 1 & 3 \\ -1 & 5 \end{vmatrix} = -56 .
$$

More work if you choose $j = 3$ for instance.

A notion central to understanding determinants is the idea of a **permutation**.

1.3.4 Permutations and determinants

A **permutation** of the numbers $\{1, 2, 3, \cdots, n\}$ is a rearrangement or a sorting of the numbers into a different order. So

$$
123456 \to 562341
$$

is a permutation. We can denote this permutation as

$$
\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 6 & 2 & 3 & 4 & 1 \end{pmatrix} \qquad \text{or simply} \qquad \sigma = 562341 .
$$

In general, the notation is

$$\sigma = j_1 j_2 \cdots j_n \quad \text{with} \quad j_i = \sigma(i) \ .$$

There are $n!$ different permutations of n numbers or objects; the permutations simply specify the different orders in which they can be laid out.

Consider a given permutation $\sigma = j_1 j_2 \cdots j_n$. We say σ is **even** or **odd** according to whether there is an even or odd number of pairs (i, k) for which

$$i > k \quad \text{but} \quad i \ \textbf{precedes} \ k \quad \text{in} \quad \sigma \ .$$

Then define the **parity** of σ to be

$$P_\sigma = \begin{cases} 1 & \text{if } \sigma \text{ is even} \\ -1 & \text{if } \sigma \text{ is odd} \end{cases}$$

So for $\sigma = 562341$ the pairs that satisfy the criterion above are

$$(5,2) \ (5,3) \ (5,4) \ (5,1) \ (6,2) \ (6,3) \ (6,4) \ (6,1) \ (2,1) \ (3,1) \ (4,1)$$

There are 11 pairs and so $P_\sigma = -1$.

Another and very useful way to understand the meaning of P_σ is to count the number of pairwise interchanges of neighbours that get you back to $12345 \cdots n$. If this is even(odd) $P_\sigma = 1(-1)$. In our example,

$$\begin{array}{ccccccc} 562341 & \longrightarrow & 156234 & \longrightarrow & 152346 & \longrightarrow & 123456 \\ & 5 & & 3 & & 3 & \end{array}$$

There are 11 pairwise interchanges (cannot be fewer than other method), and so $P_\sigma = -1$.

In particular, P_σ clearly **changes sign** upon interchange of any pair of neighbours but also under interchange of **any** pair of j's:

$$P_{123} = 1 \quad P_{132} = -1, \quad P_{321} = -1 \ .$$

We now define an important object. This called variously the **Levi-Cevita tensor** or the **epsilon tensor**. It is defined to be

$$\varepsilon_{j_1 j_2 \cdots j_n} = \begin{cases} 0 & \text{if any pair of } j_1 j_2 \cdots j_n \text{ are equal} \\ \\ P_{j_1 j_2 \cdots j_n} & \text{otherwise} \end{cases}$$

Thus

$$\varepsilon_{123} = 1, \quad \varepsilon_{321} = -1, \quad \varepsilon_{112} = 0, \quad \varepsilon_{111} = 0$$

$$\varepsilon_{1234} = 1, \quad \varepsilon_{2143} = 1, \quad \varepsilon_{2413} = -1, \quad \varepsilon_{1232} = 0\ .$$

The important result is the following.

$$\begin{aligned} |\boldsymbol{A}| &= \sum_{\text{all permutations}} P_{j_1 j_2 \cdots j_n} a_{1j_1} a_{2j_2} \cdots a_{nj_n} \\ &= \sum_{j_1 j_2 \cdots j_n} \varepsilon_{j_1 j_2 \cdots j_n} a_{1j_1} a_{2j_2} \cdots a_{nj_n} \\ &= \sum_{i_1 i_2 \cdots i_n} \varepsilon_{i_1 i_2 \cdots i_n} a_{i_1 1} a_{i_2 2} \cdots a_{i_n n}\ . \end{aligned}$$

The second sum is over $j_i = 1, 2, \cdots n$ for each j_i and in the third sum similarly over the i's.

Remarks

- The result is easily checked for $n = 2, 3$ and the general result can be established by induction.
- The sum on RHS consists of $n!$ terms, corresponding to the number of permutations, each of which is a product of n elements from (a_{ij}); each term has exactly one element from each row and column.

To get a feel for this expression we illustrate with $n = 3$.

1. let $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ be 3-dimensional vectors. Then the well-known **vector product** given by $\boldsymbol{a} = \boldsymbol{b} \wedge \boldsymbol{c}$ (written also as $\boldsymbol{b} \times \boldsymbol{c}$) has elements

$$\begin{aligned} a_i &= \varepsilon_{ijk} b_j c_k\ : \\ a_1 &= b_2 c_3 - b_3 c_2\ , \quad a_2 = b_3 c_1 - b_1 c_3\ , \quad a_3 = b_1 c_2 - b_2 c_1\ . \end{aligned}$$

   Then by construction we clearly have

$$\boldsymbol{b} \cdot (\boldsymbol{b} \wedge \boldsymbol{c}) = \sum_i b_i \left(\sum_{jk} \varepsilon_{ijk} b_j c_k \right) = \sum_{ijk} \varepsilon_{ijk} b_i b_j c_k = 0\ .$$

   The last result follows because $\varepsilon_{ijk} = -\varepsilon_{jik}$; it is anti-symmetric under $i \leftrightarrow j$ whereas $b_i b_j$ is obviously symmetric under this interchange. Similarly, $\boldsymbol{c} \cdot (\boldsymbol{b} \wedge \boldsymbol{c}) = 0$.

3. Consider

$$\boldsymbol{A} = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}$$

Then

$$|\boldsymbol{A}| = \sum_{ijk} \varepsilon_{ijk} a_i b_j c_k = \boldsymbol{a} \cdot (\boldsymbol{b} \wedge \boldsymbol{c}) ,$$

the determinant of a 3×3 matrix is the scalar triple product of its rows (or columns) treated as vectors.

4. The important general result is that if the **same** vector occurs twice anywhere in the sum involving the ε-tensor (i.e., in the **contraction** of vectors with ε) then the answer is zero.

$$\sum_{j_k j_l} \varepsilon_{j_1 j_2 \cdots j_n} v_{j_k} v_{j_l} = 0 .$$

The result follows because the permutation needed for $j_k \leftrightarrow j_l$ is always odd; $\varepsilon_{j_1 \cdots j_n}$ is **anti-symmetric** under interchange of **any** pair of indices.

It also follows immediately that if a matrix has any two rows (or columns) equal then its determinant is **zero**.

5. Consider the 3×3 matrix $\boldsymbol{A} = (a_{ij})$. Then the cofactors A_{ij} are given by

$$\begin{aligned} A_{1j} &= \sum_{j_2 j_3} \varepsilon_{j j_2 j_3} a_{2 j_2} a_{3 j_3} = (\boldsymbol{a}_2 \wedge \boldsymbol{a}_3)_j , \\ A_{2j} &= \sum_{j_1 j_3} \varepsilon_{j_1 j j_3} a_{1 j_1} a_{3 j_3} = (\boldsymbol{a}_3 \wedge \boldsymbol{a}_1)_j , \\ A_{3j} &= \sum_{j_1 j_2} \varepsilon_{j_1 j_2 j} a_{1 j_1} a_{2 j_2} = (\boldsymbol{a}_1 \wedge \boldsymbol{a}_2)_j . \end{aligned}$$

Here $\boldsymbol{a}_i = (a_{i1}, a_{i2}, a_{i3})$ – the i-th row of $\boldsymbol{A}$ written as a (row) vector. It is easy to verify that for each j these are the correct "signed" sub-determinants. Also, we see that, for example,

$$|\boldsymbol{A}| = \sum_{j_1 j_2 j_3} \varepsilon_{j_1 j_2 j_3} a_{1 j_1} a_{2 j_2} a_{3 j_3} = \sum_{j_1} a_{1 j_1} A_{1 j_1} .$$

which recovers our earlier expression for $|\boldsymbol{A}|$.

[Not lectured but for completeness the general expression for A_{ij} is

$$A_{ij} = \sum_{j_1 \cdots j_{i-1} j_{i+1} \cdots j_n} \varepsilon_{j_1 \cdots j_{i-1} j j_{i+1} \cdots j_n} a_{1 j_1} \cdots \underbrace{[a_{i j_i}]}_{\text{omit this term}} \cdots a_{n j_n}$$

]

We see also that

$$\sum_k a_{2k} A_{1k} = \sum_k a_{3k} A_{1k} = 0.$$

This follows because

(a) It is the scalar triple product with two vectors the same.

(b) It is the determinant of a matrix with two rows the same.

(c) When we unpack the sums we see that the same vector (either $\boldsymbol{a}_2$ or $\boldsymbol{a}_3$) occurs twice in the contraction with ε.

The general result for arbitrary n is that

$$\sum_k a_{ik} A_{jk} = \begin{cases} |\boldsymbol{A}| & i = j \\ 0 & i \neq j \end{cases}$$

As in (b) above, for $i \neq j$ this is the determinant of a matrix with two rows the same.

In matrix notation we have

$$\sum_k a_{ik} A_{jk} = \delta_{ij} \, |\boldsymbol{A}| \quad \text{or}$$

$$\boldsymbol{A}(\text{adj}\boldsymbol{A}) = (\det \boldsymbol{A})\boldsymbol{I}$$

1.3.5 Properties of determinants

We collect here properties derived above and a few extra ones with examples.

1. Interchanging any two rows or columns of a matrix changes the sign of its determinant.

2. $|\boldsymbol{A}| = 0$ if any two rows or columns are the same.

3. The matrix obtained by multiplying all the elements of any one row (or column) of $\boldsymbol{A}$ by λ has determinant $\lambda|\boldsymbol{A}|$.

4. Adding a multiple of one row (column) to another row (column) leaves the determinant unchanged. This is a useful way of reducing the calculation of $|\boldsymbol{A}|$.

E.g., our 3×3 example from before:

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

$R2 \rightarrow R2 + R1$:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 6 & 8 \\ 2 & 7 & 8 \end{pmatrix} .$$

$R3 \rightarrow R3 - 2 * R1$:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 6 & 8 \\ 0 & 3 & 2 \end{pmatrix} .$$

$R3 \rightarrow R3 - 1/2 * R1$:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 6 & 8 \\ 0 & 0 & -2 \end{pmatrix} .$$

Then easily find $|\boldsymbol{A}| = 1 * 6 * (-2) = -12$.

We have reduced the matrix to **upper triangular form** by performing **row operations** . Can similarly define **lower triangular form** and **column operations**. The determinant is then just the product of the elements on the main diagonal.

This is a much faster method for large matrices. The original definition requires $O(n!)$ mathematical operations $(\times, \ +)$, whereas this new method of reduction to upper (lower) triangular form requires only $O(n^3)$ operations. (Computationally, there can be issues with accuracy depending on the values of the matrix elements.)

5. det $\boldsymbol{AB} = (\det \boldsymbol{A})(\det \boldsymbol{B})$. This follows directly from the definition in terms of the ε-tensor but is fiddly to show. It relies on a useful result that I state for $n = 3$ but is easily generalized:

$$\sum_{j_1 j_2 j_3} \varepsilon_{j_1 j_2 j_3} a_{i_1 j_1} a_{i_2 j_2} a_{i_3 j_3} \ = \ |\boldsymbol{A}| \varepsilon_{i_1 i_2 i_3} \ .$$

Show this by interchanging two i's on both sides and noting that this is equivalent to interchanging the associated pair of j's on LHS together with multiplying by (-1) because the ε-tensor is antisymmetric under interchange of j's.

6. $|\boldsymbol{A}| = |\boldsymbol{A}^T|$. Using rows or columns in the formula are equivalent.

7. For ordinary 3D vectors in standard notation:

$$\boldsymbol{u} \wedge \boldsymbol{v} \;=\; \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}, \qquad \text{curl } \boldsymbol{v} \;=\; \nabla \wedge \boldsymbol{v} \;=\; \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ \partial/\partial x_1 & \partial/\partial x_2 & \partial/\partial x_3 \\ v_1 & v_2 & v_3 \end{vmatrix}.$$

8. The (signed) volume of the parallelepiped in 3D with sides $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ is $V(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \boldsymbol{a} \cdot (\boldsymbol{b} \wedge \boldsymbol{c})$. Thus

$$V(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) \;=\; \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

The general result, which can be proved by induction, is that the (signed) volume of a parallelepiped in n-dimensions with sides $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots \boldsymbol{a}_n$ is

$$V(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots \boldsymbol{a}_n) = |\boldsymbol{A}| \quad \text{where} \quad \boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1 & \longrightarrow \\ \boldsymbol{a}_2 & \longrightarrow \\ \vdots & \vdots \\ \boldsymbol{a}_n & \longrightarrow \end{pmatrix}$$

Of course, here $\boldsymbol{A} = (a_{ij})$ as usual. In all examples can use columns instead of rows.

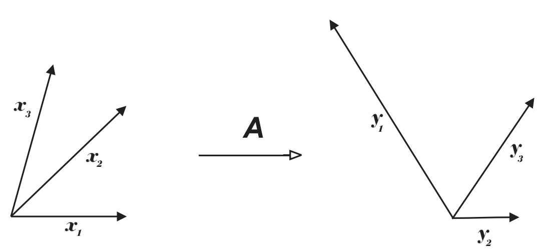
9. A result that is proved using $\det \boldsymbol{AB} = (\det \boldsymbol{A})(\det \boldsymbol{B})$ can be illustrated in 3D. Given two parallelepipeds defined by $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ and $(\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3)$ which are related by

$$\boldsymbol{A}\boldsymbol{x}_i \;=\; \boldsymbol{y}_i, \qquad i = 1, 2, 3$$

(treating $\boldsymbol{x}$ and $\boldsymbol{y}$ as column vectors), then

$$\frac{V(\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3)}{V(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)} \;=\; |\boldsymbol{A}|\,.$$

Certainly true if $\boldsymbol{A}$ is diagonal.



1.4 Inverse of a matrix

We consider only square matrices from now on.

Suppose we can find a matrix $\boldsymbol{A}^{-1}$ such that

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I} .$$

We then can find the solution to the system of linear algebraic equations

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y} ,$$

be premultiplying both sides by $\boldsymbol{A}^{-1}$ to give

$$\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{A}^{-1}\boldsymbol{A})\boldsymbol{x} = \boldsymbol{I}\boldsymbol{x} = \boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{y} ,$$

and hence we determine $\boldsymbol{x}$.

The question is whether given $\boldsymbol{A}$ that $\boldsymbol{A}^{-1}$ exists and whether it is unique.

1.4.1 Uniqueness of inverse

If $\boldsymbol{A}^{-1}$ exists, it is unique and is **both** the **left** and **right** inverse. By this we mean

If $\boldsymbol{L}\boldsymbol{A} = \boldsymbol{I}$ then $\boldsymbol{L}$ is the **left inverse** of $\boldsymbol{A}$

If $\boldsymbol{A}\boldsymbol{R} = \boldsymbol{I}$ then $\boldsymbol{R}$ is the **right inverse** of $\boldsymbol{A}$.

Suppose that $\boldsymbol{L}$ is not unique, i.e., $\boldsymbol{L}_1\boldsymbol{A} = \boldsymbol{I}$ and $\boldsymbol{L}_2\boldsymbol{A} = \boldsymbol{I}$. Then

$$\begin{aligned} \boldsymbol{L}_1 - \boldsymbol{L}_2 &= (\boldsymbol{L}_1 - \boldsymbol{L}_2)\boldsymbol{I} = (\boldsymbol{L}_1 - \boldsymbol{L}_2)\boldsymbol{A}\boldsymbol{R} \\ &= (\boldsymbol{L}_1\boldsymbol{A} - \boldsymbol{L}_2\boldsymbol{A})\boldsymbol{R} = (\boldsymbol{I} - \boldsymbol{I})\boldsymbol{R} = 0 . \end{aligned}$$

Hence, $\boldsymbol{L}_1 = \boldsymbol{L}_2$ and so $\boldsymbol{L}$ (and likewise $\boldsymbol{R}$) is unique.

Now

$$\boldsymbol{L} = \boldsymbol{L}\boldsymbol{I} = \boldsymbol{L}\boldsymbol{A}\boldsymbol{R} = \boldsymbol{I}\boldsymbol{R} = \boldsymbol{R} ,$$

and so the left and right inverses are the same.

1.4.2 Existence and construction of inverse

Earlier in this course we derived the important result that

$$\boldsymbol{A}(\mathrm{adj}\boldsymbol{A}) = (\det \boldsymbol{A})\boldsymbol{I} .$$

Thus, if $\boldsymbol{A}^{-1}$ exists, we have a ready-made construction of the right inverse of $\boldsymbol{A}$ and hence of $\boldsymbol{A}^{-1}$, namely

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

The 3×3 case is familiar. Suppose

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a} & \longrightarrow \\ \boldsymbol{b} & \longrightarrow \\ \boldsymbol{c} & \longrightarrow \end{pmatrix} ,$$

then

$$\boldsymbol{A}^{-1} = \frac{1}{\boldsymbol{a}\cdot(\boldsymbol{b}\wedge\boldsymbol{c})} \begin{pmatrix} \boldsymbol{b}\wedge\boldsymbol{c} & \boldsymbol{c}\wedge\boldsymbol{a} & \boldsymbol{a}\wedge\boldsymbol{b} \\ \downarrow & \downarrow & \downarrow \end{pmatrix} .$$

This works because

$$\begin{pmatrix} \boldsymbol{a} & \longrightarrow \\ \boldsymbol{b} & \longrightarrow \\ \boldsymbol{c} & \longrightarrow \end{pmatrix} \begin{pmatrix} \boldsymbol{b}\wedge\boldsymbol{c} & \boldsymbol{c}\wedge\boldsymbol{a} & \boldsymbol{a}\wedge\boldsymbol{b} \\ \downarrow & \downarrow & \downarrow \end{pmatrix} = \begin{pmatrix} \boldsymbol{a}\cdot(\boldsymbol{b}\wedge\boldsymbol{c}) & 0 & 0 \\ 0 & \boldsymbol{b}\cdot(\boldsymbol{c}\wedge\boldsymbol{a}) & 0 \\ 0 & 0 & \boldsymbol{c}\cdot(\boldsymbol{a}\wedge\boldsymbol{b}) \end{pmatrix} .$$

If $|\boldsymbol{A}| = 0$ then $\boldsymbol{A}^{-1}$ does not exist and we say that $\boldsymbol{A}$ is a **singular matrix**. This is the matrix generalization of the statement that $x * 0 = 1$ has no solution for x.

However, a matrix whose determinant is zero is still not trivial. Some examples are

$$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} , \qquad \begin{pmatrix} 5 & 7 & 9 \\ 6 & 9 & 12 \\ 4 & 5 & 6 \end{pmatrix} .$$

These matrices are equivalent under row operations. If $|\boldsymbol{A}| = 0$ then at least one row (column) can be reduced to all zeros by row (column) operations.

$$\begin{pmatrix} 0 & 4 & 0 & 7 \\ 0 & 3 & 0 & 5 \\ 0 & -1 & 0 & 9 \\ 0 & 1 & 0 & 0 \end{pmatrix} , \qquad \begin{pmatrix} 15 & 4 & 18 & 7 \\ 11 & 3 & 13 & 5 \\ 7 & -1 & 16 & 9 \\ 2 & 1 & 1 & 0 \end{pmatrix} .$$

Equivalent under column operations $C1 \to 2 * C2 + C4,\ C3 \to C2 + 2 * C4$.

1.4.3 Orthogonal matrices

A square matrix $\boldsymbol{O}$ which satisfies $\boldsymbol{O}\boldsymbol{O}^T = \boldsymbol{O}^T\boldsymbol{O} = \boldsymbol{I}$. Thus, $\boldsymbol{O}^{-1} = \boldsymbol{O}^T$. We have

$$|\boldsymbol{O}^T\boldsymbol{O}| \;=\; |\boldsymbol{O}|^2 \;=\; |\boldsymbol{I}| \;=\; 1, \quad \Longrightarrow \quad |\boldsymbol{O}| \;=\; \pm 1\,.$$

(i) **Rotations**

$$\boldsymbol{R}(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

The rotation of a vector in $\mathcal{R}^n$ is a linear map given by an orthogonal matrix; $n = 2$ (2D) example given above. A rotation of column vector $\boldsymbol{x}$ through angle θ gives vector $\boldsymbol{y}$ where

$$\boldsymbol{y} = \boldsymbol{R}(\theta)\boldsymbol{x}.$$

Rotations preserve the length of the vector and so

$$\boldsymbol{x}^T\boldsymbol{x} = \boldsymbol{y}^T\boldsymbol{y} = \boldsymbol{x}^T(\boldsymbol{R}^T\boldsymbol{R})\boldsymbol{x}\,.$$

True for all $\boldsymbol{x}$ and hence we deduce that $\boldsymbol{R}^T\boldsymbol{R} = \boldsymbol{I}$. Prove $\boldsymbol{R}\boldsymbol{R}^T = \boldsymbol{I}$ by noting that $\boldsymbol{x} = \boldsymbol{R}^T\boldsymbol{y}$ and repeating argument. For rotations $|\boldsymbol{R}| = 1$.

(ii) **Reflections**

The vector $\boldsymbol{x}'$ obtained by reflecting $\boldsymbol{x}$ in a plane with unit normal $\boldsymbol{n}$ is $\boldsymbol{x}' = \boldsymbol{x} - 2(\boldsymbol{x}\cdot\boldsymbol{n})\boldsymbol{n}$. In matrix notation, writing $\boldsymbol{n}$ as a column vector:

$$\boldsymbol{x}' \;=\; \boldsymbol{O}\boldsymbol{x}, \qquad \boldsymbol{O} \;=\; \boldsymbol{I} - 2\boldsymbol{n}\boldsymbol{n}^T, \qquad (\boldsymbol{O})_{ij} \;=\; \delta_{ij} - 2n_in_j\,.$$

Use $\boldsymbol{n}^T\boldsymbol{n} = 1$ to show $\boldsymbol{O}^T\boldsymbol{O} = \boldsymbol{I}$. For reflection $|\boldsymbol{O}| = -1$. Check this (without loss of generality) by choosing $\boldsymbol{n}^T = (0, 0, 1)$. Then $\boldsymbol{O} = \mathrm{diag}(1, 1, -1)$ (diagonal matrix with these elements on diagonal).

1.5 Linear equations

1.5.1 Cramer's rule

If $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ and $|\boldsymbol{A}| \neq 0$, then

$$\boldsymbol{x} \;=\; \boldsymbol{A}^{-1}\boldsymbol{y} \;=\; \frac{(\mathrm{adj}\boldsymbol{A})\boldsymbol{y}}{|\boldsymbol{A}|}\,.$$

Then

$$x_i \;=\; \frac{1}{|\boldsymbol{A}|}\sum_k A_{ki}y_k\,.$$

We can rewrite the RHS and we get

$$x_i = \frac{1}{|\boldsymbol{A}|} \begin{vmatrix} a_{11} & \cdots & y_1 & \cdots \\ a_{21} & \cdots & y_2 & \cdots \\ \vdots & & \vdots & \\ a_{n1} & \cdots & y_n & \cdots \end{vmatrix}$$

where the y's replace the i-th column in $\boldsymbol{A}$. Thus, we get **Cramer's rule**.

1.5.2 Uniqueness of solutions

Consider the set of equations

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y} ,$$

where $\boldsymbol{A}$ is $m \times n$, $\boldsymbol{x}$ is $n \times 1$ and $\boldsymbol{y}$ is $m \times 1$ (i.e., column vectors). Given $\boldsymbol{y}$, we wish to investigate the possible solutions to these m equations for the n unknowns $x_1, x_2, \cdots, x_n$.

- We may have **redundant** equations in this set. A redundant equation is some linear combination of the others and should be omitted. If there are redundant equations the equations will be **linearly dependent**.

- There may be **inconsistent** equations in the set. Best seen by example:

$$\begin{pmatrix} 4 & 3 \\ 4 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 9 \end{pmatrix} \qquad \begin{pmatrix} 2 & 3 & 4 \\ -4 & 6 & 5 \\ 0 & 12 & 13 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 7 \end{pmatrix} .$$

  First example is obvious. In second, on LHS $R3 = 2 * R1 + R2$ but $7 \neq 2 * 1 + 3$.

  Inconsistent if the LHS is **linearly dependent** but the corresponding y-values on RHS do not obey the **same** linear relationship. Then no solution exists.

We can first check linear dependence on the LHS by inspecting the entries in $\boldsymbol{A}$ and then, if necessary, inspect the entries in $\boldsymbol{y}$ to check for **redundancy** or **inconsistency**.

(1) If $m < n$ the system is **underdetermined**; there is not enough information to fix all the x's. However, unless the equations are inconsistent, it is possible to express some of the x's in terms of the others. That is, to find a **family** of solutions.

E.g., $m = 1,\ n = 2$:

$$a_{11}x_1 + a_{12}x_2 \ = \ y_1 \ .$$

This defines a straight line in the 2D space of (x_1, x_2).

In general, the family of solutions will lie in an $(n - m)$ dimensional subspace (or larger if there are redundant equations) of the n-dim space in which $\boldsymbol{x}$ lies. E.g., in 3D (assuming no redundancy) $m = 2,\ n = 3$ is a line, $m = 1,\ n = 3$ is a plane.

(2) If $m > n$ then the LHS of the equations **must** be linearly dependent since the vectors $\boldsymbol{a}_1 \cdots \boldsymbol{a}_m$ lie in an n-dim space. Then different cases are

(i) The equations are **inconsistent** and so there is **no** solution. In this case we say that the system is overdetermined. E.g., 3×2 case

$$\begin{pmatrix} 3 & 1 \\ 5 & 2 \\ 13 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \ = \ \begin{pmatrix} 1 \\ 4 \\ 12 \end{pmatrix} \ .$$

On LHS $R3 = R1 + 2 * R2$ but on RHS $12 \neq 1 + 2 * 4$.

(ii) There are redundant equations and we can discard them and so reduce m.

- $m > n$: then still overdetermined as in (2)(i);
- $m < n$: then underdetermined as in (1);
- $m = n$: important case.

(c) The $n \times n$ case.

(i) $|\boldsymbol{A}| \neq 0$. In this case the rows (and columns) of $\boldsymbol{A}$ are linearly independent and so the equations are neither redundant nor inconsistent. The inverse $\boldsymbol{A}^{-1}$ exists and is unique. The system of equations has the solution

$$\boldsymbol{x} \ = \ \boldsymbol{A}^{-1}\boldsymbol{y} \ .$$

In the special case $\boldsymbol{y} = 0$ the **only** solution is $\boldsymbol{x} = 0$. Thus,

$$\boldsymbol{A}\boldsymbol{x} \ = \ 0 \quad \textbf{and} \quad |\boldsymbol{A}| \ \neq \ 0 \quad \Longrightarrow \quad \boldsymbol{x} \ = \ 0 \ .$$

Note, that since the columns of $\boldsymbol{A}$ treated as vectors $\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_n$ are linearly independent they form a **basis** for $\mathcal{R}^n$. Thus

$$\boldsymbol{A} \ = \ \begin{pmatrix} \boldsymbol{c}_1 & \boldsymbol{c}_2 & \cdots & \boldsymbol{c}_n \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \ .$$

The equations are then just

$$\boldsymbol{y} \;=\; x_1\boldsymbol{c}_1 + x_2\boldsymbol{c}_2 + \cdots + x_n\boldsymbol{c}_n \;;$$

the x's are the **coordinates** of $\boldsymbol{y}$ in this basis. Hence, if $\boldsymbol{y} = \mathbf{0}$, the zero vector, we expect all coordinates $x_i = 0, \forall\, i$.

- In general we can solve the equations $\boldsymbol{Ax} = \boldsymbol{y}$ by performing row operations to both sides (i.e., on $\boldsymbol{A}$ **and** $\boldsymbol{y}$) chosen to reduce $\boldsymbol{A}$ to upper triangular form. The equations then solve iteratively.

$$\begin{pmatrix} 1 & 4 & 3 \\ 2 & 9 & 5 \\ -1 & -1 & -4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 9 \\ 17 \\ -8 \end{pmatrix}$$

Then operations $R2 \rightarrow (R2 - 2 * R1),\;\; R3 \rightarrow (R3 + R1),\;\; R3 \rightarrow (R3 - 3 * R2)$ give

$$\begin{pmatrix} 1 & 4 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 9 \\ -1 \\ 4 \end{pmatrix}$$

We can now solve, in order, $x_3 = 2,\; x_2 = 1,\; x_1 = -1$. Also, $|\boldsymbol{A}| = 2$.
This is an example of **Gaussian elimination**.

(ii) $|\boldsymbol{A}| = 0,\; \boldsymbol{y} = \mathbf{0}$. We seek solutions of the **homogeneous** equations

$$\boldsymbol{Ax} \;=\; \mathbf{0}\;.$$

It is now convenient to think of the rows of $\boldsymbol{A}$ being vectors $\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_n$:

$$\boldsymbol{A} \;=\; \begin{pmatrix} \mathbf{r}_1 & \longrightarrow \\ \mathbf{r}_2 & \longrightarrow \\ \vdots & \vdots \\ \mathbf{r}_n & \longrightarrow \end{pmatrix}\;.$$

The $\mathbf{r}$'s are linearly dependent, and again let the greatest number of independent vectors be k. So the $\mathbf{r}$'s span a subspace S_r of $\boldsymbol{R}^n$, with $\dim S_r = k$.

The equations are now written

$$\boldsymbol{x} \cdot \mathbf{r}_i \;=\; 0, \qquad i = 1, 2, \cdots, n\;.$$

Example with $n = 3$. Suppose $k = 2$ and so there are two linearly independent vectors in $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$. Choose these to be $\mathbf{r}_1, \mathbf{r}_2$. Then they form a basis for the 2D space S_r. The equations to be solved are

$$\boldsymbol{x} \cdot \mathbf{r}_1 \;=\; 0, \quad \boldsymbol{x} \cdot \mathbf{r}_2 \;=\; 0, \qquad \boldsymbol{x} \cdot \mathbf{r}_3 \;=\; 0\,.$$

The trick is to find a vector $\boldsymbol{z}$ that does **not** lie in S_r, and $\boldsymbol{z} = \mathbf{r}_1 \wedge \mathbf{r}_2$ is the obvious choice. By construction $\boldsymbol{z} \cdot \mathbf{r}_i = 0, \quad i = 1, 2, 3$. Then clearly

$$\boldsymbol{A}\boldsymbol{z} \;=\; \mathbf{0}\,,$$

and hence we deduce the solution for $\boldsymbol{x}$ to be

$$\boldsymbol{x} \;=\; \lambda \boldsymbol{z} \;\equiv\; \lambda(\mathbf{r}_1 \wedge \mathbf{r}_2)\,,$$

for any value of λ.

The result for the general case stated above is that there will be $(n-k)$ independent vectors $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_{n-k}$ that do not lie in S_r so that for any $\mathbf{s} \in S_r$ then $\mathbf{s} \cdot \boldsymbol{z}_i = 0, \forall\, i$. The solution for $\boldsymbol{x}$ is then of the form

$$\text{Since} \quad \boldsymbol{A}\boldsymbol{z}_i = \mathbf{0}, \quad i = 1, 2, \cdots, (n-k), \quad \text{we get solution} \quad \boldsymbol{x} = \sum_{i=1}^{n-k} \alpha_i \boldsymbol{z}_i\,,$$

for any $\alpha_i, \quad i = 1, 2 \cdots, (n-k)$.

(The space spanned by $\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_{n-k}$ is called the **kernel** of $\boldsymbol{A}$.)

(iii) $|\boldsymbol{A}| = 0, \;\boldsymbol{y} \neq \mathbf{0}$.

The column vectors of $\boldsymbol{A}$, denoted $\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_n$, referred to in (i) are **linearly dependent** and so **do not** form a basis for $\mathcal{R}^n$ but rather only **span** a subspace $S_c \in \mathcal{R}^n$. If the greatest number of independent vectors is k, then S_c has dim $S_c = k$. (Note, that although S_c and S_c have the same dimension they are generally **not** the same subspace.)

Look again at the equation in the form

$$\boldsymbol{y} \;=\; x_1\boldsymbol{c}_1 + x_2\boldsymbol{c}_2 + \cdots + x_n\boldsymbol{c}_n\,.$$

- If $\boldsymbol{y}$ does **not** lie in the subspace S_c ($\boldsymbol{y} \notin S_c$), there can be **no** solution for $\boldsymbol{x}$. (S_c is called the **image** of $\boldsymbol{A}$ since $\boldsymbol{A}$ must map every vector $\boldsymbol{x} \in \mathcal{R}^n$ into S_c.)
  Example with $n = 3$.

$$\begin{pmatrix} 1 & -1 & 1 \\ 2 & 5 & 9 \\ 1 & 3 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \;=\; \begin{pmatrix} -1 \\ 12 \\ 7 \end{pmatrix}, \quad \begin{pmatrix} 2 \\ -2 \\ 8 \end{pmatrix}.$$

The columns are linearly dependent: $\boldsymbol{c}_3 = 2 * \boldsymbol{c}_1 + \boldsymbol{c}_2$. I can choose the basis for the 2D space S_c to be

$$\boldsymbol{c}_1 \;=\; \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \quad \boldsymbol{c}_2 \;=\; \begin{pmatrix} -1 \\ 5 \\ 3 \end{pmatrix}.$$

In manner similar to before, consider $\boldsymbol{w} = \boldsymbol{c}_1 \wedge \boldsymbol{c}_2$. Here

$$\boldsymbol{c}_1 \wedge \boldsymbol{c}_2 \;=\; \begin{pmatrix} 1 \\ -4 \\ 7 \end{pmatrix}.$$

Since $\boldsymbol{c}_1,\ \boldsymbol{c}_2, \boldsymbol{c}_1 \wedge \boldsymbol{c}_2$ **do** form a basis for $\mathcal{R}^3$ we can write

$$\boldsymbol{y} \;=\; p_1\boldsymbol{c}_1 + p_2\boldsymbol{c}_2 + p_3\boldsymbol{c}_1 \wedge \boldsymbol{c}_2\ .$$

The point is that if $p_3 \neq 0$ then $\boldsymbol{y}$ does not lie in S_c and there is no solution. The condition for $\boldsymbol{y}$ to lie in S_c is

$$\boldsymbol{y} \cdot \boldsymbol{c}_1 \wedge \boldsymbol{c}_2 \;=\; 0.$$

This is satisfied by the first example but not the second.

The result for the general case stated above is that there will be $(n-k)$ independent vectors $\boldsymbol{w}_1, \cdots, \boldsymbol{w}_{n-k}$ that **do not** lie in S_c so that for any $\mathbf{s} \in S_c$ then $\mathbf{s} \cdot \boldsymbol{w}_i = 0, \forall\ i$. The conditions for a solution for $\boldsymbol{x}$ to exist is then that

$$\boldsymbol{y} \cdot \boldsymbol{w}_i \;= 0 \qquad i = 1, 2, \cdots, (n-k)\ .$$

Suppose these conditions are satisfied and we find a solution $\boldsymbol{x}_0$: $\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{y}$. This solution is **not unique** since we clearly also have

$$\boldsymbol{A}(\boldsymbol{x}_0 + \alpha\boldsymbol{z}) \;=\; \boldsymbol{y}, \quad \text{where} \quad \boldsymbol{A}\boldsymbol{z} \;=\; \mathbf{0}\ .$$

Thus, the most general solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ in this case is

$$\boldsymbol{x} \;=\; \boldsymbol{x}_0 + \sum_{i=1}^{n-k} \alpha_i \boldsymbol{z}_i\ ,$$

where the $\boldsymbol{z}_i$ satisfy $\boldsymbol{A}\boldsymbol{z}_i = \mathbf{0}$, $\ i = 1, 2, \cdots, (n-k)$ as explained in (ii).

1.6 Eigenvalues and eigenvectors

If

$$\boldsymbol{A}\boldsymbol{v} \;=\; \lambda \boldsymbol{v}\;,$$

where λ is a scalar and $\boldsymbol{v} \neq \boldsymbol{0}$, then

- λ is an **eigenvalue** of $\boldsymbol{A}$
- $\boldsymbol{v}$ is the **eigenvector** of $\boldsymbol{A}$ corresponding to the eigenvalue λ.

(i) Acting (or operating) on $\boldsymbol{v}$ with $\boldsymbol{A}$ scales it by λ leaving the **direction** unchanged.

(ii) If $\boldsymbol{v}$ is an eigenvector then so is $\alpha\boldsymbol{v}$.

We can then write

$$\boldsymbol{A}\boldsymbol{v} \;=\; \lambda \boldsymbol{I}\boldsymbol{v}\;, \quad \Longrightarrow \quad (\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{v} \;=\; 0\;.$$

The only solution is $\boldsymbol{v} = \boldsymbol{0}$ **except** for special values of λ for which $\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = 0$. Thus, we seek solutions for λ to

$$\det(\boldsymbol{A} - \lambda\boldsymbol{I}) \;=\; \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} \;\equiv\; P_A(\lambda) \;=\; 0\;.$$

This determinant is a polynomial of degree n in λ and is called the **characteristic polynomial** $P_A(\lambda)$ of $\boldsymbol{A}$. It is degree n since

- each term in $P_A(\lambda)$ is n-th order in the a's and contains one element from each row and column.
- The product of a's on the diagonal is one such term and this contains λ^n.

$P_A(\lambda)$ has n roots and these are the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$. The set of eigenvalues is the **spectrum** of $\boldsymbol{A}$. They may be **complex** even if the entries in $\boldsymbol{A}$ are real. E.g., consider

$$\boldsymbol{A} \;=\; \begin{pmatrix} -2 & 6 \\ 6 & 7 \end{pmatrix}\;. \qquad P_A(\lambda) \;=\; \begin{vmatrix} -2 - \lambda & 6 \\ 6 & 7 - \lambda \end{vmatrix}$$

$$=\; (-2 - \lambda)(7 - \lambda) - 6^2 \;=\; \lambda^2 - 5\lambda - 50 \;=\; 0 \quad \Longrightarrow \quad \lambda \;=\; 10, -5\;.$$

Here $\boldsymbol{A}$ is symmetric.

$$\boldsymbol{A} \;=\; \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} . \qquad P_A(\lambda) \;=\; \begin{vmatrix} \cos\theta-\lambda & \sin\theta \\ -\sin\theta & \cos\theta-\lambda \end{vmatrix}$$

$$= \;(\cos\theta-\lambda)^2+\sin^2\theta \;=\; \lambda^2-2\lambda\cos\theta+1 \;=\; 0 \;\implies\; \lambda \;=\; e^{\pm i\theta} .$$

Here $\boldsymbol{A}$ is anti-symmetric and also orthogonal and a rotation matrix: the eigenvalues give the angle of rotation.

(i)

$$\det(\boldsymbol{A}-\lambda\boldsymbol{I}) \;=\; P_A(\lambda) \;=\; (\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n) .$$

Evaluate with $\lambda = 0$ and find important result

$$\det\boldsymbol{A} \;=\; (-1)^n\prod_{i=1}^{n}\lambda_i .$$

(ii) If $|\boldsymbol{A}| = 0$ then at least one eigenvalue is zero and the corresponding eigenvectors satisfy

$$\boldsymbol{A}\boldsymbol{v} \;=\; \boldsymbol{0} .$$

This **homogeneous** equation was discussed earlier, and we can see that the set of eigenvectors with $\lambda = 0$ will span the **kernel** of $\boldsymbol{A}$; the space of vectors annihilated by $\boldsymbol{A}$.

(iii) By inspecting the definition of $P_A(\lambda)$ and coeff. of λ^0 term also can show that

$$\mathrm{trace}(\boldsymbol{A}) \;=\; \sum_{i=1}^{n}\lambda_i .$$

(iv) For real matrices the coefficients in $P_A(\lambda)$ are real and so if any λ are complex then they must come in complex-conjugate pairs. The number of real eigenvalues (and eigenvectors) can therefore be less than n; there are none in one of the 2×2 examples above when $\theta \neq 0, \pi$.

Each eigenvalue λ_a has its corresponding eigenvector $\boldsymbol{v}_a$:

$$\boldsymbol{A}\boldsymbol{v}_a \;=\; \lambda_a\boldsymbol{v}_a, \quad a = 1, 2, \cdots, n.$$

Since $\alpha\boldsymbol{v}$ will also do, we can choose α so that $\boldsymbol{v}$ is **normalized**, usually to 1. Using the inner (or scalar) product we can choose the eigenvectors so that $\boldsymbol{v}_a\cdot\boldsymbol{v}_a = 1$.

1.6.1 Real symmetric matrices

Defined by $\boldsymbol{A} = \boldsymbol{A}^* = \boldsymbol{A}^T$.

1. A real symmetric matrix has **real eigenvalues**

$$\begin{aligned}
\boldsymbol{Av} &= \lambda\boldsymbol{v} && && (1)\\
\boldsymbol{Av}^* &= \lambda^*\boldsymbol{v}^* && \text{complex conjugate} &&\\
(\boldsymbol{v}^*)^T\boldsymbol{A} &= \lambda^*(\boldsymbol{v}^*)^T && \text{transpose} && (2)\\
(\boldsymbol{v}^*)^T\boldsymbol{Av} &= \lambda(\boldsymbol{v}^*)^T\boldsymbol{v} && \text{left multiply (1) by } (\boldsymbol{v}^*)^T &&\\
(\boldsymbol{v}^*)^T\boldsymbol{Av} &= \lambda^*(\boldsymbol{v}^*)^T\boldsymbol{v} && \text{right multiply (2) by } \boldsymbol{v} &&\\
(\lambda-\lambda^*)(\boldsymbol{v}^*)^T\boldsymbol{v} &= 0 && \text{subtract} &&
\end{aligned}$$

Since $(\boldsymbol{v}^*)^T\boldsymbol{v} = 1$ we deduce that $(\lambda - \lambda^*) = 0$ and hence that λ is **real**.

The eigenvector $\boldsymbol{v}$ is therefore **real** since it solves real equations with real coefficients.

2. The eigenvectors corresponding to different eigenvalues of a symmetric matrix are **orthogonal**. Similar procedure to above.

$$\begin{aligned}
\boldsymbol{Av}_1 &= \lambda_1\boldsymbol{v}_1 && &&\\
\boldsymbol{v}_1^T\boldsymbol{A} &= \lambda_1\boldsymbol{v}_1^T && \text{transpose} &&\\
\boldsymbol{v}_1^T\boldsymbol{Av}_2 &= \lambda_1\boldsymbol{v}_1^T\boldsymbol{v}_2 && \text{right multiply by } \boldsymbol{v}_2 && (1)
\end{aligned}$$

Similarly,

$$\begin{aligned}
\boldsymbol{Av}_2 &= \lambda_2\boldsymbol{v}_2 && &&\\
\boldsymbol{v}_1^T\boldsymbol{Av}_2 &= \lambda_2\boldsymbol{v}_1^T\boldsymbol{v}_2 && \text{left multiply by } \boldsymbol{v}_1^T && (2)\\
(\lambda_1-\lambda_2)\boldsymbol{v}_1^T\boldsymbol{v}_2 &= 0 && (1) - (2) &&
\end{aligned}$$

Since $\lambda_1 \neq \lambda_2$ we deduce that $\boldsymbol{v}_1^T\boldsymbol{v}_2 = 0$.

(i) If some of the λ's coincide ("degeneracy") there are still n linearly independent eigenvectors which can be made to be orthogonal. This is done by choice of suitable linear combinations of those $\boldsymbol{v}$'s corresponding to the degenerate eigenvalues.

(ii) Normalize each $\boldsymbol{v}_i$ to unit magnitude. The eigenvectors then comprise an **orthonormal basis** which we now denote $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n$. So

$$\boldsymbol{Ae}_a = \lambda_i\boldsymbol{e}_a, \qquad \boldsymbol{e}_a \cdot \boldsymbol{e}_b \;=\; \boldsymbol{e}_a^T\boldsymbol{e}_b \;=\; \delta_{ab}\,.$$

We learn that if we want to choose a nice basis when working with $\boldsymbol{A}$ we should choose the basis given by its **orthonormal** eigenvectors.

1.6.2 Diagonalization of real symmetric matrices

Consider the $n \times n$ matrix $\boldsymbol{X}$ whose i-th column is $\boldsymbol{e}_i$:

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_n \\ \downarrow & \downarrow & \downarrow & \downarrow \end{pmatrix}$$

Now

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{pmatrix} \boldsymbol{e}_1 & \longrightarrow \\ \boldsymbol{e}_2 & \longrightarrow \\ \vdots & \vdots \\ \boldsymbol{e}_n & \longrightarrow \end{pmatrix} \begin{pmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_n \\ \downarrow & \downarrow & \downarrow & \downarrow \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

So

$$\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{I}.$$

Thus

- $\boldsymbol{X}^{-1} = \boldsymbol{X}^T$.
- $\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{X}\boldsymbol{X}^T = \boldsymbol{I}$.
- $\boldsymbol{X}$ is an **orthogonal** matrix.
- $\det \boldsymbol{X} = 1$.

Now,

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{A}\begin{pmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_n \\ \downarrow & \downarrow & \downarrow & \downarrow \end{pmatrix} = \begin{pmatrix} \lambda_1\boldsymbol{e}_1 & \lambda_2\boldsymbol{e}_2 & \cdots & \lambda_n\boldsymbol{e}_n \\ \downarrow & \downarrow & \downarrow & \downarrow \end{pmatrix}.$$

Thus

$$\boldsymbol{A}' = \boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

is a diagonal matrix with diagonal elements given by the eigenvalues of $\boldsymbol{A}$.

Clearly, $\boldsymbol{A}'$ has the **same** eigenvalues as $\boldsymbol{A}$ but the eigenvectors are the **canonical** basis:

$$\boldsymbol{e}_1' = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \boldsymbol{e}_2' = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \text{etc.}$$

1. Given $\boldsymbol{Ax} = \boldsymbol{y}$ we can write

$$\begin{aligned} \boldsymbol{A}(\boldsymbol{X}\boldsymbol{X}^T)\boldsymbol{x} &= y \\ (\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X})\boldsymbol{X}^T\boldsymbol{x} &= \boldsymbol{X}^T\boldsymbol{y} \\ \boldsymbol{A}'\boldsymbol{x}' &= \boldsymbol{y}', \end{aligned}$$

   with $\boldsymbol{x}' = \boldsymbol{X}^T\boldsymbol{x}, \quad \boldsymbol{y}' = \boldsymbol{X}^T\boldsymbol{y}$. Clearly, in the special case of the eigenvectors:

$$\boldsymbol{e}_i' = \boldsymbol{X}^T\boldsymbol{e}_i \; .$$

2. What are the **coordinates** of $\boldsymbol{x}$ in basis of the $\boldsymbol{e}$'s?

$$\boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{e}_i \; .$$

   Then

$$\boldsymbol{x}' = \boldsymbol{X}^T\boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{X}^T\boldsymbol{e}_i = \sum_{i=1}^{n} x_i \boldsymbol{e}_i' = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} .$$

   The entries in $\boldsymbol{x}'$ are simply the coordinates of $\boldsymbol{x}$ in the basis of eigenvectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$.

3. $$|\boldsymbol{A}'| = |\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}| = |\boldsymbol{X}^T||\boldsymbol{A}||\boldsymbol{X}| = |\boldsymbol{A}| \; ,$$

   since $|\boldsymbol{X}| = 1$. Then have $|\boldsymbol{A}'| = \prod_{i=1}^{n} \lambda_i = |\boldsymbol{A}|$. Result derived earlier.

Look at my earlier example.

$$\boldsymbol{A} = \begin{pmatrix} -2 & 6 \\ 6 & 7 \end{pmatrix}, \qquad \lambda_1 = 10, \; \lambda_2 = -5.$$

Look for $\boldsymbol{e}_1$. This satisfies

$$\begin{pmatrix} -2-\lambda_1 & 6 \\ 6 & 7-\lambda_1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \equiv \begin{pmatrix} -12 & 6 \\ 6 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0}\,.$$

These two equations are multiples of each other (by construction). Then

$$-12x + 6y = 0 \implies y = 2x.$$

The normalized vector is

$$\boldsymbol{e}_1 = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}.$$

$\boldsymbol{e}_2$ can be derived similarly but we also know it is orthogonal to $\boldsymbol{e}_1$. Hence,

$$\boldsymbol{e}_2 = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} \end{pmatrix}.$$