Matrices & Linear Algebra

Vector Spaces

- **Scalars** are the **elements** of a **number field** (for example, **R** and **C**), which
  - Is a set of elements on which the **operations** of **addition** and **multiplication** are defined, and satisfy the **usual laws of arithmetic** (**commutative**, **associative** and **distributive**).
  - Is **closed** under **addition** and **multiplication**
  - Includes **identity elements** for **addition** and **multiplication** (0 and 1).
  - Includes **inverses** (negatives and reciprocals) for **addition** and **multiplication**, **except 0**.
- **Vectors** are elements of a **linear vector space** defined **over a number field *F***. A vector space V
  - Is a **set of elements** on which the operations of **vector addition** and **scalar multiplication** are defined and **satisfy certain axioms**.
  - Is **closed** under these operations.
  - Includes an **identity element** (**0**) for vector addition.
- If the number field F over which the linear vector space is defined is real, then the vector space is real.
- Notes:
  - Vector multiplication is *not*, in general, defined for a vector space.
  - The basic example of a vector space is a list of n scalars, $\mathbf{R}^n$. Vector addition and scalar multiplication are defined component-wise.

- $\mathbf{R}^2$ is not exactly the same as $\mathbf{C}$, because $\mathbf{C}$ has a rule for multiplication.
- Similarly, $\mathbf{R}^3$ is not quite the same as physical space, because physical space has a rule (Pythagoras') for the distance between two points.

The Inner Product

- The **inner product** is used to give a meaning to lengths and angles in a vector space.
- It is a **scalar function**, $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$.
- An inner product must
  - Be **linear in the second argument**

$$\langle \boldsymbol{x}, \alpha \boldsymbol{y} \rangle = \alpha \langle \boldsymbol{x}, \boldsymbol{y} \rangle$$

$$\langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle$$

  - Have **Hermitian symmetry**

$$\langle \boldsymbol{y}, \boldsymbol{x} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle^*$$

  - Be **positive definite**

$$\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$$

  with equality if and only if $\boldsymbol{x} = \boldsymbol{0}$.

Notes:

- The inner product has an existence without reference to any basis.
- The hermitian symmetry is required so that $\langle \boldsymbol{x}, \boldsymbol{x} \rangle$ is real. It is ***not*** required in a **real** vector space.
- It follows from the above that the inner product is **antilinear** in the first argument:

$$\langle \alpha \boldsymbol{x}, \boldsymbol{y} \rangle = \alpha^* \langle \boldsymbol{x}, \boldsymbol{y} \rangle$$

$$\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$$

- In $\mathbf{C}^n$, the standard (Euclidean) inner product, the "dot product", is:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = x_i^* y_i$$

The complex conjugation is needed to maintain Hermitian symmetry, and to ensure that the product is linear in the **second** argument.

- The **Cauchy-Schwarz Inequality** states that

$$\left|\langle \boldsymbol{x}, \boldsymbol{y} \rangle\right|^2 \le \langle \boldsymbol{x}, \boldsymbol{x} \rangle \langle \boldsymbol{y}, \boldsymbol{y} \rangle$$

Or, equivalently:

$$\left|\langle \boldsymbol{x}, \boldsymbol{y} \rangle\right| \le |\boldsymbol{x}| |\boldsymbol{y}|$$

With equality if and only if $\boldsymbol{x} = \alpha \boldsymbol{y}$.

To prove, assume that $\boldsymbol{x}, \boldsymbol{y} \ne \boldsymbol{0}$ (in which case the inequality is trivial). We first say:

$$\begin{aligned}\langle \boldsymbol{x} - \alpha\boldsymbol{y} \mid \boldsymbol{x} - \alpha\boldsymbol{y} \rangle &= \langle \boldsymbol{x} - \alpha\boldsymbol{y} \mid \boldsymbol{x} \rangle - \alpha \langle \boldsymbol{x} - \alpha\boldsymbol{y} \mid \boldsymbol{y} \rangle \\ &= \langle \boldsymbol{x} \mid \boldsymbol{x} \rangle - \alpha^* \langle \boldsymbol{y} \mid \boldsymbol{x} \rangle - \alpha \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle + \alpha\alpha^* \langle \boldsymbol{y} \mid \boldsymbol{y} \rangle \\ &= \langle \boldsymbol{x} \mid \boldsymbol{x} \rangle - \alpha^* \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle^* - \alpha \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle + \alpha\alpha^* \langle \boldsymbol{y} \mid \boldsymbol{y} \rangle \\ &= |\boldsymbol{x}|^2 + |\alpha|^2 |\boldsymbol{y}|^2 - 2\,\mathrm{Re}\left(\alpha \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle\right)\end{aligned}$$

Now, this quantity **must** be positive, because of the positive definite property of the inner product. If we choose the **phase** of α (which is arbitrary) such that $\alpha \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle$ is real and non-negative, so that $\alpha \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle = |\alpha| \left|\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle\right|$, we then have that:

$$|\boldsymbol{x}|^2 + |\alpha|^2 |\boldsymbol{y}|^2 - 2|\alpha| \left|\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle\right| \ge 0$$

$$\left(|\boldsymbol{x}| - |\alpha||\boldsymbol{y}|\right)^2 + 2|\alpha||\boldsymbol{x}||\boldsymbol{y}| - 2|\alpha| \left|\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle\right| \ge 0$$

And now, if we choose $|\alpha| = |\boldsymbol{x}| / |\boldsymbol{y}|$, we get:

$$|\boldsymbol{x}||\boldsymbol{y}| \ge \left|\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle\right|$$

As required.

- We can use the **Cauchy-Schwarz inequality** to prove the **triangle inequality** ($|\boldsymbol{x} + \boldsymbol{y}| \le |\boldsymbol{x}| + |\boldsymbol{y}|$), by writing $|\boldsymbol{x} + \boldsymbol{y}|^2$ in terms of the inner product, expanding, using the inequality, and factorising.
- The **Cauchy-Schwarz Inequality** allows us to define, in **real vector space**, the angle θ between two vectors, though

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = |\boldsymbol{x}| |\boldsymbol{y}| \cos\theta$$

[This is possible because, by the Cauchy Schwartz, $\left|\frac{\langle x,y\rangle}{|x||y|}\right| \le 1$. If $\langle \boldsymbol{x},\boldsymbol{y}\rangle = 0$, in any vector space, $\boldsymbol{x}$ and $\boldsymbol{y}$ are said to be **orthogonal**.

- A knowledge of the inner product of the basis vectors is sufficient to determine the inner product of any two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. Let:

  $$\left\langle \boldsymbol{e}_i \mid \boldsymbol{e}_j \right\rangle = G_{ij}$$

  Then

  $$\langle \boldsymbol{x},\boldsymbol{y}\rangle = G_{ij} x_i^* y_j$$

  Where the G_{ij} are the **metric coefficients** of the basis. The Hermetian Symmetry of the inner product implies that

  $$G_{ij} = G_{ji}^*$$

  The matrix G is **hermitian**.

- For an **orthonormal basis**, in which $\left\langle \boldsymbol{e}_i \mid \boldsymbol{e}_j \right\rangle = \delta_{ij}$, we have that

  $$\langle \boldsymbol{x} \mid \boldsymbol{y}\rangle = x_i^* y_i$$

Bases

- Let $S = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_m\}$ be a subset of the vectors in V.
- A **linear combination** of S is any vector of the form $x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \cdots + x_m\boldsymbol{e}_m$, where the x are scalars.
- The **span** of S is the set of **all vectors** that are **linear combinations** of S. If the span of S is the entire vector space V, then S is said to **span** $\boldsymbol{V}$.
- The vectors of S are said to be **linearly independent** if

  $$x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \cdots + x_m\boldsymbol{e}_m = \boldsymbol{0} \quad \Leftrightarrow \quad x_1, x_2, \cdots, x_m = 0$$

  If, on the other hand, such an equation holds for non-trivial values of the coefficients, one of the vectors is **redundant**, and can be written as a linear combination of the other vectors.

- If an **additional** vector is added to a **spanning set**, it **remains** a spanning set. If a vector is **removed** from a

linearly independent set, it **remains** a linearly independent set.

- A **basis** for a vector space V is the **subset of vectors** S **that spans** V **and is also linearly independent**. The properties of basis sets are:
    - **All bases** of V have the **same number of elements**, n, which is called the **dimension** of V.
    - **Any n linearly independent vectors** in V form a **basis** of V.
    - **Any** vector $\boldsymbol{x} \in V$ can be written [prove by considering $V \cup \{\boldsymbol{x}\}$ as a linearly dependent set] in a **unique way** [prove by contradiction] as a **linear combination** of the vectors in a basis. The relevant scalars are called the **components** of $\boldsymbol{x}$ with respect to that particular basis.
- The **same vector** (a geometrical identity) has **different components** with respect to **different bases**. To see how we can change from one to the other, consider **two bases** of V – $S = \{\boldsymbol{e}_i\}$ and $S' = \{\boldsymbol{e}'_i\}$.
    - Because both are bases, the elements of one basis can be written in terms of the other:

    $$\boxed{\boldsymbol{e}_j = \boldsymbol{e}'_i R_{ij}}$$

    Where R_{ij} is the **transformation matrix** between the two bases
    - Now, consider a vector $\boldsymbol{x} \in V$. The representation of the vector in each basis is:

    $$\boldsymbol{x} = \boldsymbol{e}_j x_j = \boldsymbol{e}'_i x'_i$$

    However, using our result from above, we can write this as:

    $$\boldsymbol{x} = \boldsymbol{e}'_i x'_i = \boldsymbol{e}_j x_j = \boldsymbol{e}'_i R_{ij} x_j$$

    - From this, we can deduce the **transformation law for vector components**:

    $$\boxed{x'_i = R_{ij} x_j}$$

Note that:

- The law is the "reverse" of that for basis vector transformation. This is to ensure that, overall, the vector $\boldsymbol{x}$ stays unchanged by transformation.
- The first suffix of R corresponds to the same basis in both relations.

- We defined R_{ij}, above, by

$$\boldsymbol{e}_j = \boldsymbol{e}_i' R_{ij}$$

The condition of the basis $\{\boldsymbol{e}_j\}$ to be orthonormal is

$$\boldsymbol{e}_i^\dagger \boldsymbol{e}_j = \delta_{ij}$$

$$\left(\boldsymbol{e}_k' R_{ki}\right)^\dagger \boldsymbol{e}_l' R_{lj} = \delta_{ij}$$

$$R_{ki}^* R_{lj} \boldsymbol{e}_k'^\dagger \boldsymbol{e}_l' = \delta_{ij}$$

If the second basis is also orthonormal, this becomes:

$$R_{ki}^* R_{kj} = \delta_{ij}$$

$$\boldsymbol{R}^\dagger \boldsymbol{R} = \boldsymbol{I}$$

In other words, transformations between orthonormal bases is described by **unitary matrices**. In real vector space, an **orthogonal matrix** does this – in $\mathbf{R}^2$ and $\mathbf{R}^3$, this corresponds to a **rotation** and/or **reflection**.

- Given any $\boldsymbol{m}$ vectors $\boldsymbol{u}_1 \cdots \boldsymbol{u}_m$ that span an $\boldsymbol{n}$-dimensional vector space ($\boldsymbol{m} \geq \boldsymbol{n}$), it is possible to construct an **orthogonal basis** $\boldsymbol{e}_1 \cdots \boldsymbol{e}_n$ using the **Gram-Schmidt procedure**:

$$\boldsymbol{e}_1 = \boldsymbol{u}_1$$

$$\boldsymbol{e}_r = \boldsymbol{u}_r - \sum_{s=1}^{r-1} \frac{\boldsymbol{e}_s \cdot \boldsymbol{u}_r}{\boldsymbol{e}_s \cdot \boldsymbol{e}_s} \boldsymbol{e}_s$$

What we are effectively doing is taking each vector $\boldsymbol{u}$ and "removing" any "bits" of vectors we've already added to the basis from it, to leave us with a final vector that is orthogonal to all others already added…

We can prove, by **induction**, that this works:

Inductive step

Assume that vectors $\boldsymbol{e}_1 \cdots \boldsymbol{e}_t$ have already been added to the orthogonal basis (such that $\boldsymbol{e}_i \cdot \boldsymbol{e}_j = 0 \quad \forall\, i \neq j$), and now consider the vector $\boldsymbol{e}_{t+1}$ that we're about to add:

$$\boldsymbol{e}_{t+1} = \boldsymbol{u}_{t+1} - \sum_{s=1}^{r-1} \frac{\boldsymbol{e}_s \cdot \boldsymbol{u}_{t+1}}{\boldsymbol{e}_s \cdot \boldsymbol{e}_s} \boldsymbol{e}_s$$

And now, consider dotting it with $\boldsymbol{e}_v$ ($v \leq t$), any of the vectors already in the basis:

$$\begin{aligned}
\boldsymbol{e}_{t+1} \cdot \boldsymbol{e}_v &= \boldsymbol{u}_{t+1} \cdot \boldsymbol{e}_v - \sum_{s=1}^{r-1} \frac{\boldsymbol{e}_s \cdot \boldsymbol{u}_{t+1}}{\boldsymbol{e}_s \cdot \boldsymbol{e}_s} \underbrace{\left[\boldsymbol{e}_s \cdot \boldsymbol{e}_v\right]}_{=0 \text{ if } v \neq s} \\
&= \boldsymbol{u}_{t+1} \cdot \boldsymbol{e}_v - \frac{\boldsymbol{e}_v \cdot \boldsymbol{u}_{t+1}}{\boldsymbol{e}_s \cdot \boldsymbol{e}_s} \left[\boldsymbol{e}_s \cdot \boldsymbol{e}_s\right] \\
&= \boldsymbol{u}_{t+1} \cdot \boldsymbol{e}_v - \boldsymbol{e}_v \cdot \boldsymbol{u}_{t+1} \\
&= 0
\end{aligned}$$

So the new vector is indeed orthogonal to all the vectors already in the set.

"Starting off" step

Consider $\boldsymbol{e}_1 \cdot \boldsymbol{e}_2$:

$$\begin{aligned}
\boldsymbol{e}_1 \cdot \boldsymbol{e}_2 &= \boldsymbol{u}_1 \cdot \left(\boldsymbol{u}_2 - \frac{\boldsymbol{e}_1 \cdot \boldsymbol{u}_2}{\boldsymbol{e}_1 \cdot \boldsymbol{e}_1} \boldsymbol{e}_1 \right) \\
&= \boldsymbol{u}_1 \cdot \boldsymbol{u}_2 - \frac{\boldsymbol{u}_1 \cdot \boldsymbol{u}_2}{\boldsymbol{e}_1 \cdot \boldsymbol{e}_1} \left[\boldsymbol{u}_1 \cdot \boldsymbol{e}_1\right] \\
&= \boldsymbol{u}_1 \cdot \boldsymbol{u}_2 - \frac{\boldsymbol{u}_1 \cdot \boldsymbol{u}_2}{\boldsymbol{e}_1 \cdot \boldsymbol{e}_1} \left[\boldsymbol{e}_1 \cdot \boldsymbol{e}_1\right] \\
&= 0
\end{aligned}$$

So the first two vectors are, indeed, orthogonal.

Matrices

- **Array viewpoint**
    - Matrices can be regarded, simply, as an **array of numbers**, R_{ij}.
    - The rule for **multiplying** a **matrix** by a **vector** is then

$$(\boldsymbol{A}\boldsymbol{x})_i = \boldsymbol{A}_{ij}\boldsymbol{x}_j$$

- The rules for matrix **addition** and **multiplication** are

$$(\boldsymbol{A}+\boldsymbol{B})_{ij} = \boldsymbol{A}_{ij} + \boldsymbol{B}_{ij}$$
$$(\boldsymbol{AB})_{ij} = \boldsymbol{A}_{ik}\boldsymbol{B}_{kj}$$

- **Linear Operator Viewpoint**
  - A **linear operator A** acts on a vector space V to produce other elements of V.
  - The property of **linearity** means that:

$$\boldsymbol{L}(\alpha \boldsymbol{x}) = \alpha \boldsymbol{L}(\boldsymbol{x})$$
$$\boldsymbol{A}(\boldsymbol{x}+\boldsymbol{y}) = \boldsymbol{A}(\boldsymbol{x}) + \boldsymbol{A}(\boldsymbol{y})$$

  - A linear operator can exist without reference to any basis. It can be thought of as a **linear transformation** or **mapping** of the space V. [Some linear operators can even transform between different bases].
  - The **components** of **A** with respect to a basis $\{\boldsymbol{e}_i\}$ is defined by the **action** of **A** on the **basis vectors**:

$$\boldsymbol{A}\boldsymbol{e}_j = A_{ij}\boldsymbol{e}_i$$

    The components form a **square matrix**. [In other words, the j^{th} *column* of A contains the compoenents of the result of **A** acting on e_j].
  - We now know enough to determine the action of **A** only any $\boldsymbol{x}$:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}(x_j\boldsymbol{e}_j) = x_j\boldsymbol{A}\boldsymbol{e}_j = x_jA_{ij}\boldsymbol{e}_i = A_{ij}x_j\boldsymbol{e}_i$$

    So:

$$(\boldsymbol{A}\boldsymbol{x})_i = A_{ij}x_j$$

    This corresponds to the rule for multiplying a **matrix** by a **vector**.
  - Furthermore, the **sum** of two linear operators is defined by

$$(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{x} = \boldsymbol{e}_{\boldsymbol{i}}\left(A_{ij}+B_{ij}\right)x_j$$

  - And the **product** of two linear operators is defined by

$$(\boldsymbol{AB})\boldsymbol{x} = \boldsymbol{A}(\boldsymbol{Bx}) = \boldsymbol{A}(\boldsymbol{e}_i B_{ij} x_j) = (\boldsymbol{A}\boldsymbol{e}_k) B_{kj} x_j = \boldsymbol{e}_i A_{ik} B_{kj} x_j$$

- Both these operations satisfy the rules of matrix addition and multiplication and action on a vector. As such, a **linear operator** can be represented as a **matrix**.

- **Back to Change of Basis**
  - Above, we wrote one set of basis vectors in terms of the other:

$$\boldsymbol{e}_j = \boldsymbol{e}_i' R_{ij}$$

But we could also have written

$$\boldsymbol{e}_j' = \boldsymbol{e}_i S_{ij}$$

Substituting one into the other, we have

$$\boldsymbol{e}_j = \boldsymbol{e}_k S_{ki} R_{ij}$$

$$\boldsymbol{e}_j' = \boldsymbol{e}_k' R_{ki} S_{ij}$$

But this can only be true if

$$S_{ki} R_{ij} = R_{ki} S_{ij} = \delta_{kj}$$

Which implies that

$$RS = SR = 1$$

$$\Rightarrow R = S^{-1}$$

  - We noted, above, that the transformation laws for vector components could be written

$$x_i' = R_{ij} x_j$$

We can write this in matrix form, as

$$x' = Rx$$

With the inverse relation

$$x' = R^{-1} x$$

- **Linear Operators – Change of Basis**
  - To find how the components of a linear operator $\boldsymbol{A}$ transform under a change of basis, we note that we require

$$\boldsymbol{Ax} = \boldsymbol{e}_i A_{ij} x_j = \boldsymbol{e}_i' A_{ij}' x_j'$$

Using $\boldsymbol{e}_j = R_{ij} \boldsymbol{e}_i'$, we have that:

$$\boldsymbol{e}'_k R_{ki} A_{ij} x_j = \boldsymbol{e}'_k A'_{kj} x'_j$$
$$R_{ki} A_{ij} x_j = A'_{kj} x'_j$$
$$RAx = A'x'$$
$$RA\left(R^{-1}x'\right) = A'x'$$

Which means that

$$A' = RAR^{-1}$$

- **Matrix Multiplication**
  - Matrix multiplication does not commute. But it does distribute, so, with a bit of care, normal rules of algebra can be applied. For example:

$$(\boldsymbol{1} - \boldsymbol{W})(\boldsymbol{1} + \boldsymbol{W})$$
$$= \boldsymbol{1}(\boldsymbol{1} + \boldsymbol{W}) - \boldsymbol{W}(\boldsymbol{1} + \boldsymbol{W})$$
$$= \boldsymbol{1} + \boldsymbol{1}\boldsymbol{W} - \boldsymbol{W}\boldsymbol{1} - \boldsymbol{W}^2$$
$$= \boldsymbol{1} - \boldsymbol{W}^2$$
$$= (\boldsymbol{1} + \boldsymbol{W})(\boldsymbol{1} - \boldsymbol{W})$$

Hermitian Conjugate

- We define the **Hermitian conjugate** of a matrix as follows:

$$A^\dagger = (A^T)^*$$
$$\left(A^\dagger\right)_{ij} = A^*_{ji}$$

- Importantly:

$$(AB)^\dagger = B^\dagger A^\dagger$$

  (Note the reversal of the order).
- We can also write the inner product as:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = x^\dagger G y$$

  Where G is the matrix of metric coefficients. This preserve the hermitian symmetry of the inner product as long as the matrix is hermitian – $G = G^\dagger$.
- The **adjoint** of a linear operator $\boldsymbol{A}$ with respect to a given inner product is a linear operator $\boldsymbol{A}^\dagger$ satisfying

$$\left\langle \boldsymbol{A}^\dagger \boldsymbol{x} \mid \boldsymbol{y} \right\rangle = \langle \boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{y} \rangle$$

With the standard inner product, we find that the matrix defining $\boldsymbol{A}^\dagger$ is, indeed, the hermitian conjugate of $\boldsymbol{A}$.

Special Matrices

- **Symmetry**
  - A **symmetric matrix** is equal to its transpose
$$\boldsymbol{A} = \boldsymbol{A}^T$$
  - An **hermitian matrix** is equal to its hermitian conjugate.
$$\boldsymbol{A} = \boldsymbol{A}^\dagger$$
  - An **antisymmetric** (or **skew-symmetric**) matrix satisfies
$$\boldsymbol{A}^T = -\boldsymbol{A}$$
  - An **anti-hermitian** (or **skew-hermitian**) matrix satisfies
$$\boldsymbol{A}^\dagger = -\boldsymbol{A}$$
- **Orthogonality**
  - An **orthogonal matrix** is one whose **transpose** is equal to its **inverse**
$$\boldsymbol{A}^T = \boldsymbol{A}^{-1}$$
  - A **unitary matrix** is one whose **hermitian conjugate** is equal to its **inverse**
$$\boldsymbol{A}^\dagger = \boldsymbol{A}^{-1}$$
    We note that if $\boldsymbol{U}$ is a unitary matrix, then $\boldsymbol{A}^\dagger \boldsymbol{A} = \boldsymbol{1}$. This implies that the **columns** of $\boldsymbol{A}$ are **orthonormal vectors**.
- A **normal** matrix is one that **commutes with its Hermitian conjugate**:
$$\boldsymbol{A}\boldsymbol{A}^\dagger = \boldsymbol{A}^\dagger \boldsymbol{A}$$
  It is easy to verify that **hermitian**, **anti-hermitian** and **unitary** matrices are all **normal**.
- **Relationships**

- If $\boldsymbol{A}$ is Hermitian, then $\boldsymbol{A}i$ is **anti-hermitian**, and **vice-versa**.
- If $\boldsymbol{A}$ is Hermitian, then
$$\exp(\boldsymbol{A}i) = \sum_{n=0}^{\infty} \frac{(\boldsymbol{A}i)^n}{n!}$$
Is **unitary**.
- *[This can be remembered by bearing in mind that if* $\boldsymbol{z}$ *is a real number,* $\boldsymbol{iz}$ *is imaginary, and if* $\boldsymbol{z}$ *is a real number, then* e^{iz} *has unit modulus (see below when talking about eigenvalues of normal matrices)]*

- To prove that a matrix is a certain type of special matrix, find an expression for the determining property. For example, to prove it's unitary, find $\boldsymbol{UU}^\dagger$.

Eigenvalues and Eigenvectors

- An **eigenvector** of a linear operator $\boldsymbol{A}$ is a non-zero vector $\boldsymbol{x}$ satisfying
$$\boldsymbol{Ax} = \lambda \boldsymbol{x}$$
$$(\boldsymbol{A} - \lambda \boldsymbol{1})\boldsymbol{x} = \boldsymbol{0}$$
For some scalar λ, called the **eigenvalue**.
- The equation (in its second form) effectively says that a **linear combination** of the **columns** of the matrix $(\boldsymbol{A} - \lambda \boldsymbol{I})$ is equal to 0. This is equivalent to the statement
$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0$$
Which is called the **characteristic equation** of the matrix.
- There are two possibilities in terms of roots:
  - If there are $\boldsymbol{n}$ **distinct solutions** to the characteristic equation, then there are $\boldsymbol{n}$ **linearly independent eigenvectors**. We prove this as follows. Assume that

$$\sum a_\alpha \boldsymbol{e}_\alpha = \boldsymbol{0}$$

We can multiply both sides by whatever we want, so:

$$(\boldsymbol{A} - \lambda_1 \boldsymbol{I})\sum a_\alpha \boldsymbol{e}_\alpha = \boldsymbol{0}$$

$$\boldsymbol{y} = (\lambda_2 - \lambda_1)a_2\boldsymbol{e}_2 + (\lambda_3 - \lambda_1)a_3\boldsymbol{e}_3 + \cdots + (\lambda_n - \lambda_1)a_n\boldsymbol{e}_n = \boldsymbol{0}$$

We can do the same again with $\boldsymbol{y}$:

$$(\boldsymbol{A} - \lambda_2 \boldsymbol{I})\boldsymbol{y} = \boldsymbol{0}$$

$$(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)a_3\boldsymbol{e}_3 + \cdots + (\lambda_n - \lambda_1)(\lambda_n - \lambda_2)a_n\boldsymbol{e}_n = \boldsymbol{0}$$

We can then repeat this until we obtain:

$$\underbrace{(\lambda_n - \lambda_1)(\lambda_n - \lambda_2)\cdots(\lambda_n - \lambda_{n-2})(\lambda_n - \lambda_{n-1})}a_n\boldsymbol{e}_n = \boldsymbol{0}$$

Now, if all the λ are distinct, the expression enclosed by a brace is non-zero. Therefore, a_n **must** be 0. Removing the last vector and repeatedly applying this method shows us that *all* the a_n must be 0. Therefore,

$$\sum a_\alpha \boldsymbol{e}_\alpha = \boldsymbol{0}$$

is only true if all the a_α are 0. Therefore, the vectors are linearly independent.

- If the roots are **not all distinct**, then the repeated values are said to be **degenerate**. If a value λ occurs **m times**, there may be **any number** between **1 and m** of **linearly independent** eigenvectors. Any **linear combination** of these is also an eigenvector.

A **defective matrix** is one who **vector space** is not **spanned** by its **eigenvectors**. Such a matrix cannot be **diagonalised** by a change of basis.

- It can be shown that a **normal matrix** is **never defective**. In fact, an **orthonormal basis** can always be constructed from the **eigenvectors** of a matrix, **if and only if** the matrix is **normal**.
- Some interesting properties can be derived regarding the properties of the **eigenvectors** and **eigenvalues** of **normal matrices**:

- The **eigenvectors** corresponding to **distinct eigenvalues** are **orthogonal**.
- The **eigenvalues** are
  - **Real** for **hermitian matrices**.
  - **Imaginary** for **anti-Hermitian matrices.**
  - **Of unit modulus** for **unitary matrices**.

  A good way to remember these properties is to consider that a 1×1 matrix is just a number λ, and to be Hermitian, imaginary or unitary, it must satisfy

  $$\lambda = \lambda^* \qquad \lambda = -\lambda^* \qquad \lambda^* = \lambda^{-1}$$

  Which are precisely the conditions for λ being real, imaginary or of unit modulus.

The method to prove these results is, in general, as follows:

- Choose two arbitrary **eigenvectors** and write the eigenvector equations:

  $$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x} \qquad \boldsymbol{A}\boldsymbol{y} = \mu\boldsymbol{y}$$

- Take one of these equations, and find the **hermitian conjugate**.
- Then
  - For a **hermitian matrix**, construct two expressions for $\boldsymbol{y}^\dagger\boldsymbol{A}\boldsymbol{x}$.
  - For a **unitary matrix**, multiply both sides by the other eigenvector equation that hadn't be used.
- Re-arrange in the form ***something = 0***.
- Assume that $\boldsymbol{x} = \boldsymbol{y}$, and using the fact that $\boldsymbol{x}, \boldsymbol{y} \neq \boldsymbol{0}$, deduce something about the eigenvalues.
- Now, assume that $\boldsymbol{x} \neq \boldsymbol{y}$ and deduce that $\boldsymbol{y}^\dagger\boldsymbol{x} = 0$ as long as $\lambda \neq \mu$, proving that the vectors are **orthogonal**.

- Matrices are given particular names:

- If all **eigenvalues** are < 0 (> 0), the matrix is **negative** (**positive**) **definite**.
- If all **eigenvalues** are ≤ 0 (≥ 0), the matrix is **negative** (**positive**) **semi-definite**.
- A matrix is **definite** if it is **either positive definite** or **negative definite**.

Diagonalization

- Two **square** matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are said to be **similar** if they are related by

$$\boldsymbol{B} = \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$$

In other words, if they are **representations** of the **same linear transformation** in **different bases**. $\boldsymbol{S}$ is called the **similarity matrix**.

- A matrix is said to be **diagonalisable** if it is **similar** to a **diagonal matrix** – in other words, if

$$\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S} = \Lambda$$

Where Λ is a diagonal matrix.

- Consider a matrix $\boldsymbol{S}$ whose columns are the eigenvectors of the matrix $\boldsymbol{A}$:

$$\begin{aligned}\boldsymbol{A}\boldsymbol{S} &= \boldsymbol{A}\begin{bmatrix} \boxed{\boldsymbol{e}_1} & \boxed{\boldsymbol{e}_2} & \cdots & \boxed{\boldsymbol{e}_n} \end{bmatrix} \\ &= \begin{bmatrix} \boxed{\lambda_1\boldsymbol{e}_1} & \boxed{\lambda_2\boldsymbol{e}_2} & \cdots & \boxed{\lambda_3\boldsymbol{e}_n} \end{bmatrix} \\ &= \begin{bmatrix} \boxed{\boldsymbol{e}_1} & \boxed{\boldsymbol{e}_2} & \cdots & \boxed{\boldsymbol{e}_n} \end{bmatrix}\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & 0 & \\ & 0 & \ddots & \\ & & & \lambda_n \end{bmatrix} \\ &= \boldsymbol{S}\Lambda\end{aligned}$$

We can therefore say that

$$\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S} = \Lambda$$

Provided that $\boldsymbol{S}$ is invertible – ie:, provided that the columns of $\boldsymbol{S}$ are linearly independent – ie: provided that the eigenvectors of $\boldsymbol{A}$ are linearly independent.

- Notes:
  - We notice that $\boldsymbol{S}$ is the transformation matrix to the eigenvector basis. Therefore, diagonalisation is the process of expressing a matrix in its **simplest form** by transforming to its **eigenvector basis**.
  - An $n \times n$ matrix is diagonalisable if and only if it has n linearly independent eigenvectors. That is to say, only if it is **normal**. Furthermore, if the eigenvectors are chosen to be **orthonormal**, then the columns of $\boldsymbol{S}$ are orthonormal and $\boldsymbol{S}$ is therefore **unitary** (= a matrix whose columns are orthonormal vectors).
- Diagonalisation is rather useful in carrying out certain operations on matrices:

$$\boldsymbol{A}^m = (\boldsymbol{S}\Lambda\boldsymbol{S}^{-1})(\boldsymbol{S}\Lambda\boldsymbol{S}^{-1})\cdots(\boldsymbol{S}\Lambda\boldsymbol{S}^{-1}) = \boldsymbol{S}\Lambda^m\boldsymbol{S}^{-1}$$

$$\det(\boldsymbol{A}) = \det(\boldsymbol{S}\Lambda\boldsymbol{S}^{-1}) = \det(\boldsymbol{S})\det(\Lambda)\det(\boldsymbol{S}^{-1}) = \det(\Lambda)$$

$$\mathrm{tr}(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{S}\Lambda\boldsymbol{S}^{-1}) = \mathrm{tr}(\Lambda\boldsymbol{S}\boldsymbol{S}^{-1}) = \mathrm{tr}(\Lambda)$$

$$\mathrm{tr}(\boldsymbol{A}^m) = \mathrm{tr}(\Lambda^m)$$

  Where we have used the following properties of determinants and traces:

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

$$\det(\boldsymbol{S})\det(\boldsymbol{S}^{-1}) = 1$$

$$\mathrm{tr}(\boldsymbol{AB}) = (\boldsymbol{AB})_{ii} = A_{ij}B_{ji} = B_{ji}A_{ij} = (\boldsymbol{BA})_{jj} = \mathrm{tr}(\boldsymbol{BA})$$

- Note that in general, for any matrix $\boldsymbol{A}$

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

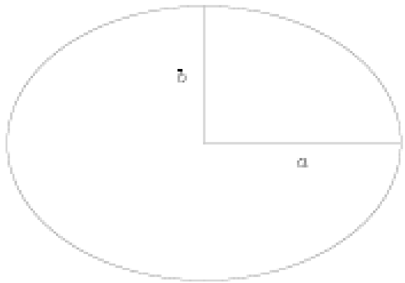
$$\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i$$

Quadratic & Hermitian Forms

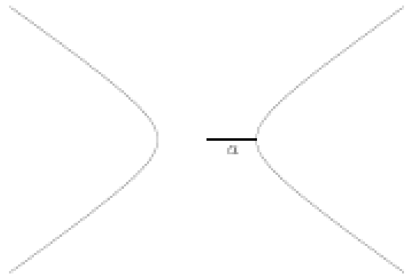
- The **quadratic form** associated with a **real symmetric matrix** $\boldsymbol{A}$ is
$$Q(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} = A_{ij}x_i x_j$$
Q is a **homogeneous quadratic function** – ie: $Q(\alpha\boldsymbol{x}) = \alpha^2 Q(\boldsymbol{x})$.
- In fact, *any* homogenous quadratic equation is the quadratic form of a symmetric matrix:
$$Q = ax^2 + 2bxy + cy^2 = \begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$$
- In fact, $\boldsymbol{A}$ can be diagonalised by a **real orthogonal** transformation:
$$\boldsymbol{S}^T\boldsymbol{A}\boldsymbol{S} = \Lambda \qquad\qquad (\boldsymbol{S}^T = \boldsymbol{S}^{-1})$$
And the vector $\boldsymbol{x}$ transforms according to $\boldsymbol{x} = \boldsymbol{S}\boldsymbol{x}'$, so
$$Q = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = \left(\boldsymbol{x}'^T\boldsymbol{S}^T\right)\left(\boldsymbol{S}\Lambda\boldsymbol{S}^T\right)\left(\boldsymbol{S}\boldsymbol{x}'\right) = \boldsymbol{x}'^T\Lambda\boldsymbol{x}'$$
The quadric form can therefore be reduced to:
$$Q = \sum_{i=1}^{n} \lambda_i x_i'$$
Where the x_i' are given by:
$$\boldsymbol{x}' = \boldsymbol{S}^{-1}\boldsymbol{x} = \boldsymbol{S}^T\boldsymbol{x}$$
We have effectively **rotated** the coordinates to reduce the quadric form to its simplest form.
- The **quadric surfaces** (or **quadrics**) are the family of surfaces
$$Q(\boldsymbol{x}) = k = \text{constant}$$
In the **eigenvector basis**, this simplifies to
$$\lambda_1 x_1'^2 + \lambda_2 x_2'^2 + \lambda_3 x_3'^2 = k$$
- The conic and quadric surfaces that can result are depicted on the next page. The relevant semi-axes are given by $1/\sqrt{\lambda}$. If $\lambda \to 0$, the shape "comes apart".
- A few special cases:
    - If $\lambda_1 = \lambda_2 = \lambda_3$, we have a **sphere**.
    - If (for example), $\lambda_1 = \lambda_2$, we have a surface of revolution about the third axis, whatever it might be.

- If (for example), $\lambda_3 = 0$, we have the translation of a conic section along the relevant axis (an **elliptic** or **hyperbolic cylinder**).

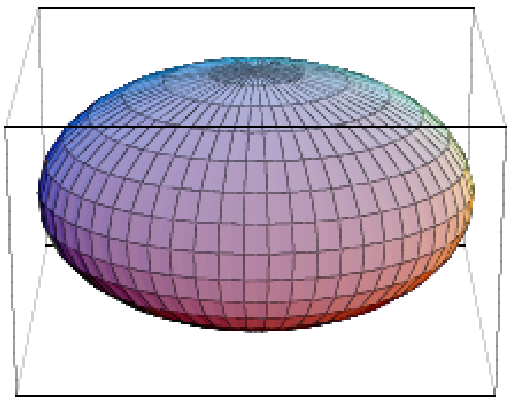
Conic sections and quadric surfaces



ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$

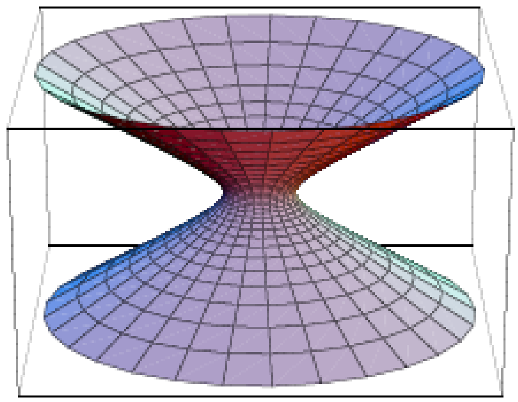


hyperbola $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$



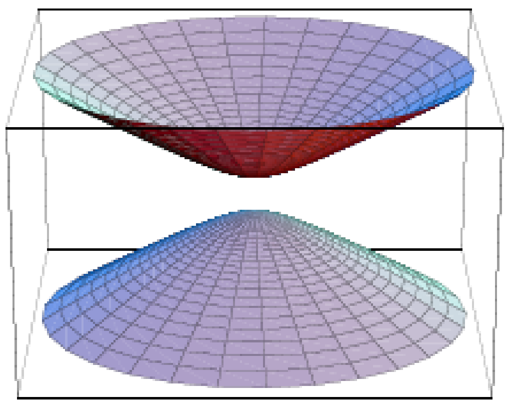
ellipsoid

$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$



hyperboloid of one sheet

$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$



hyperboloid of two sheets $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1$

- In a **complex vector space**, the **Hermitian form** associated with an **Hermitian matrix $\boldsymbol{A}$** is:

$$H(\boldsymbol{x}) = \boldsymbol{x}^{\dagger}\boldsymbol{A}\boldsymbol{x} = x_i^* A_{ij} x_j$$

H is a real scalar, because

$$H^*(\boldsymbol{x}) = \left(x_i^* A_{ij} x_j\right)^* = x_j^* A_{ij}^* x_i = x_j^* A_{ji} x_i = H(\boldsymbol{x})$$

We also know that $\boldsymbol{A}$ can be diagonalised by a **unitary transformation**

$$\boldsymbol{U}^{\dagger}\boldsymbol{A}\boldsymbol{U} = \Lambda \qquad\qquad \boldsymbol{U}^{\dagger} = \boldsymbol{U}^{-1}$$

And therefore:

$$H(\boldsymbol{x}) = \boldsymbol{x}^{\dagger}\left(\boldsymbol{U}\Lambda\boldsymbol{U}^{\dagger}\right)\boldsymbol{x} = \left(\boldsymbol{U}^{\dagger}\boldsymbol{x}\right)^{\dagger}\Lambda\left(\boldsymbol{U}^{\dagger}\boldsymbol{x}\right) = \boldsymbol{x}'^{\dagger}\Lambda\boldsymbol{x}' = \sum_{i=1}^{n}\lambda_n x_i'^2$$

Therefore, a hermitian form can be reduced to a real quadratic form by transforming to the eigenvector basis.

