A short summary of linear algebra and matrix theory

Dr I M Jaimoukha

Contents

Vector spaces

- A **field** is a set of objects, called **scalars**, for which addition, subtraction, multiplication, and division, are defined and the usual axioms of arithmetic hold.

- The sets of real numbers $\mathcal{R}$ and complex numbers $\mathcal{C}$ are fields. The set of integers is *not* a field: why?

- A **vector space** V over a field $\mathcal{F}$ is a set V of objects, called **vectors**, for which two operations, **vector addition** $(+)$ and **scalar multiplication** $(\cdot)$ are defined such that for all $x, y, z \in V$ and all $\alpha, \beta \in \mathcal{F}$ the following axioms are satisfied:

  1. $x + y \in V$ (closure w.r.t. $+$)
  2. $\alpha \cdot x \in V$ (closure w.r.t. $\cdot$)
  3. $x + y = y + x$ ($+$ commutative)
  4. $(x + y) + z = x + (y + z)$ ($+$ associative)
  5. $\exists\, 0 \in V$ such that $x + 0 = x$ (zero vector)
  6. $\exists\, \bar{x} \in V$ s.t. $x + \bar{x} = 0$ (negatives)
  7. $\alpha \cdot (x+y) = \alpha \cdot x + \alpha \cdot y$ ($\cdot$ distributive w.r.t. $+$)
  8. $\alpha \cdot (\beta \cdot x) = (\alpha\beta) \cdot x$ ($\cdot$ associative)
  9. $(\alpha + \beta) \cdot x = \alpha \cdot x + \beta \cdot x$
  10. $1 \cdot x = x$

- A vector space (or **linear space**) V over the field $\mathcal{F}$ will be denoted by $V(\mathcal{F})$.

- A **real vector space** is a vector space over $\mathcal{R}$ and a **complex vector space** is a vector space over $\mathcal{C}$.

- The following are examples of vector spaces:

  1. The sets of ordered n-tuples of real and complex numbers:

$$\mathcal{R}^n(\mathcal{R}) = \{[x_1, \ldots, x_n]^T : x_i \in \mathcal{R}, \forall\, i\}$$
$$\mathcal{C}^n(\mathcal{C}) = \{[x_1, \ldots, x_n]^T : x_i \in \mathcal{C}, \forall\, i\}$$

  2. The set $\mathcal{P}^n$, of polynomial functions of degree n.
  3. The set $\mathcal{C}[0, 2\pi]$, of continuous functions over the interval $[0, 2\pi]$.
  4. The set $\mathcal{C}[-\infty, \infty]$.

- The following are *not* vector spaces:

  1. The set of ordered n-tuples of negative numbers.
  2. The set $\mathcal{R}^n(\mathcal{C})$.

- Let $\mathcal{F}$ be a field. Then $\mathcal{F}^n(\mathcal{F})$ is always a vector space and is denoted by $\mathcal{F}^n$. In this notation, $\mathcal{C}^n(\mathcal{C}) = \mathcal{C}^n$ and $\mathcal{R}^n(\mathcal{R}) = \mathcal{R}^n$.

Subspaces

- A subset S of $V(\mathcal{F})$ is called a **subspace** if S is itself a vector space over $\mathcal{F}$.

- A nonempty subset S of $V(\mathcal{F})$ is a subspace of $V(\mathcal{F})$ if it satisfies the closure axioms:

  1. $x, y \in S \Rightarrow x + y \in S$.
  2. $x \in S,\ \alpha \in \mathcal{F} \Rightarrow \alpha x \in S$.

- Every vector space $V(\mathcal{F})$ has two special subspaces:

  1. The zero subspace $\Phi = \{0\}$.
  2. The vector space $V(\mathcal{F})$ itself.

- Any other subspace of $V(\mathcal{F})$ is called a **proper** subspace.

- The set of vectors of the form $(x_1, 0)$ where $x_1 \in \mathcal{R}$ is a proper subspace of $\mathcal{R}^2$.

- The set of all vectors in the first quadrant is *not* a subspace of $\mathcal{R}^2$.

Linear independence and span

- A set of vectors $\{x_1, \ldots, x_n\} \subset V(\mathcal{F})$ is called **linearly dependent** if there exist $\alpha_1, \ldots, \alpha_n \in \mathcal{F}$, not all zero, such that: $\alpha_1 x_1 + \cdots + \alpha_n x_n = 0$.

- A set of vectors $\{x_1, \ldots, x_n\} \subset V(\mathcal{F})$ is called **linearly independent** if:

$$\alpha_1 x_1 + \cdots + \alpha_n x_n = 0 \Leftrightarrow \alpha_i = 0,\ \forall\, i.$$

- Let $S = \{x_1, \ldots, x_n\} \subset V(\mathcal{F})$. The expression,

$$\sum_{i=1}^{n} \alpha_i x_i,$$

  where $\alpha_1, \ldots, \alpha_n \in \mathcal{F}$, is called **a linear combination** of the vectors in S.

- The **span** of $S = \{x_1, \ldots, x_n\} \subset V(\mathcal{F})$ is the set of all linear combinations of the vectors in S:

$$\text{span}\ (S) = \{x = \sum_{i=1}^{n} \alpha_i x_i :\ \alpha_i \in \mathcal{F}\},$$

  and is always a subspace of $V(\mathcal{F})$.

Bases and dimension

- A **basis** for $V(\mathcal{F})$ is a set of linearly independent vectors $S = \{x_1, \ldots, x_n\} \subset V(\mathcal{F})$ that spans $V(\mathcal{F})$.

- Every vector space has a basis.

- Let $\mathcal{F}$ be a field. The set $S = \{e_1, \ldots, e_n\}$, where,

$$\begin{aligned} e_1 &= [1, 0, 0, \ldots, 0, 0]^T \\ e_2 &= [0, 1, 0, \ldots, 0, 0]^T \\ &\vdots \\ e_n &= [0, 0, 0, \ldots, 0, 1]^T, \end{aligned}$$

  is a basis for $\mathcal{F}^n$, called the **natural basis.**

- The choice of basis is not unique. However, all bases for $V(\mathcal{F})$ have the same number of vectors called the **dimension** of $V(\mathcal{F})$, written dim $V(\mathcal{F})$.

- Examples:
  1. $\mathcal{F}^n$ has basis $\{e_1, \ldots, e_n\}$ and dimension n,
  2. $\mathcal{P}^{n-1}$ has basis $\{1, t, \ldots, t^{n-1}\}$ and dimension n,
  3. $\mathcal{C}[0, 2\pi]$ has basis $\{\exp(jkt) : k = \ldots, -1, 0, 1, \ldots\}$ and an infinite (but countable) dimension,
  4. $\mathcal{C}[-\infty, \infty]$ has an uncountable dimension.

Isomorphism and coordinate representation

- Let U and V be vector spaces over the same field $\mathcal{F}$. A function,

$$f : U \to V,$$

  is called an **isomorphism** if,
  1. f is invertible (one-to-one and onto),
  2. $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$, for all $x, y \in U$, and for all $\alpha, \beta \in \mathcal{F}$ (linear).

- Two vector spaces U and V over the same field $\mathcal{F}$ are called **isomorphic** if there exists an isomorphism $f : U \to V$.

- Any two isomorphic vector spaces have the same 'structure' since,
  - every vector in one is represented by a unique vector in the other (1),
  - every linear relation in one is represented by a corresponding linear relation in the other (2).

- Two finite dimensional vector spaces $U(\mathcal{F})$ and $V(\mathcal{F})$ are isomorphic if and only if dim $U(\mathcal{F})$ = dim $V(\mathcal{F})$.

- If $S = \{x_1, \ldots, x_n\}$ is a basis for $V(\mathcal{F})$, then every vector x in $V(\mathcal{F})$ can be expressed as,

$$x = \sum_{i=1}^{n} \alpha_i x_i,$$

for some unique vector,

$$\hat{x} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathcal{F}^n,$$

called the **coordinate representation** of x (with respect to the basis S).

- Let $V(\mathcal{F})$ be an n-dimensional vector space. The coordinate representation with respect to any basis defines an isomorphism from $V(\mathcal{F})$ to $\mathcal{F}^n$.

- Let $x = [\alpha_1, \ldots, \alpha_n]^T$ where each α_i belongs to the field $\mathcal{F}$. We can consider x as either:

  1. an element of the vector space $\mathcal{F}^n$, or,
  2. as the coordinate representation of an element in some vector space $V(\mathcal{F})$ (w.r.t. some basis).

- It follows that we can confine our attention to coordinate vector spaces such as $\mathcal{C}^n$ (for complex vector spaces) and $\mathcal{R}^n$ (for real vector spaces).

Inner product and norm

- An **inner product** on a complex vector space V is any function from $V \times V$ to $\mathcal{C}$ which satisfies:

  1. $\overline{<x, y>} = <y, x>$
  2. $<\alpha x + \beta y, z> \; = \alpha <x, z> + \beta <y, z>, \forall \alpha, \beta \in \mathcal{C}$
  3. $<x, x> \; > 0, \;\; \forall \, x \neq 0$

- An inner product on a complex vector space is:

  - **Hermitian** (1),
  - linear in the 1st argument (2), and **conjugate linear** in the 2nd argument (1,2):

    $$<x, \alpha y + \beta z> \; = \bar{\alpha} <x, y> + \bar{\beta} <x, z>, \forall \alpha, \beta \in \mathcal{C}$$

  - **positive definite** (3).

- The **standard inner product** on $\mathcal{C}^n$ is given by $<x, y> = y'x$ where $y' = [\bar{y}_1, \ldots, \bar{y}_n]$.

- A vector space on which an inner product is defined is called an **inner-product space.**

- $\mathcal{C}[0, 2\pi]$ is an inner product space with,

$$<f, g> = \frac{1}{2\pi} \int_0^{2\pi} \bar{g}(t) f(t) dt.$$

- A **norm** on a complex vector space V is a function from V to $\mathcal{R}$ which satisfies:

  1. $\|x\| > 0, \quad \forall\, x \neq 0$
  2. $\|\alpha x\| = |\alpha|\ \|x\|, \ \forall \alpha \in \mathcal{C}$ or $\mathcal{R}$
  3. $\|x + y\| \leq \|x\| + \|y\|$

- A norm on a complex vector space:
  - is positive definite (1),
  - is **homogeneous** (2),
  - satisfies the **triangle inequality** (3).

- A vector space on which a norm is defined is called a **normed space.**

- The **standard norm** on $\mathcal{C}^n$ is $\|x\| = \sqrt{x'x}$, and the standard norm on $\mathcal{C}[0, 2\pi]$ is,

$$\|f\| = \sqrt{<f, f>} = \frac{1}{\sqrt{2\pi}}\sqrt{\int_0^{2\pi} |f(t)|^2 dt}.$$

- We can define a norm on a finite-dimensional inner-product space V, called the **Euclidean norm,** by, $\|x\| = \sqrt{<x, x>}$. Such a vector space is called a **Euclidean space.**

Angles and orthogonality

- Let $x, y \in \mathcal{C}^n$. Then,

$$|x'y| \leq \|x\|\|y\| \quad \textbf{(Schwarz inequality)}$$

- The **angle** between nonzero $x, y \in \mathcal{C}^n$ is defined as,

$$\theta_{xy} = \cos^{-1}\left(\frac{|x'y|}{\|x\|\|y\|}\right)$$

  x and y are said to be **orthogonal** if $x'y = 0$.

- A set $S = \{x_1, \ldots, x_m\} \subset \mathcal{C}^n$ is called orthogonal if $x_i'x_j = 0, \ \forall i \neq j$. It is called **orthonormal** if, in addition, $\|x_i\| = 1, \ \forall i$.

- Every orthogonal set in $\mathcal{C}^n$ is linearly independent.

- Two subspaces U, V in $\mathcal{C}^n$ are said to be orthogonal if, $u'v = 0, \ \forall\, u \in U, \ \forall\, v \in V$.

- Let U be a subspace of $\mathcal{C}^n$. The subspace,

$$U^\perp = \{x \in \mathcal{C}^n : x'u = 0, \forall u \in U\},$$

  is called the **orthogonal complement** of U (in $\mathcal{C}^n$).

Gram-Schmidt orthogonalisation process

- Let $S = \{x_1, \ldots, x_m\} \subset \mathcal{C}^n$ be a given linearly independent set. The following procedure produces an orthogonal set $T = \{y_1, \ldots, y_m\} \subset \mathcal{C}^n$ such that span $(S) =$ span (T):

$$
\begin{aligned}
y_1 &= x_1 \\
y_2 &= x_2 - \frac{y_1' x_2}{y_1' y_1} y_1 \\
y_3 &= x_3 - \frac{y_1' x_3}{y_1' y_1} y_1 - \frac{y_2' x_3}{y_2' y_2} y_2 \\
&\vdots \\
y_m &= x_m - \frac{y_1' x_m}{y_1' y_1} y_1 - \cdots - \frac{y_{m-1}' x_m}{y_{m-1}' y_{m-1}} y_{m-1}
\end{aligned}
$$

- To obtain an orthonormal T, simply divide each y_i by its norm. Alternatively, use the following modified procedure:

$$
\begin{aligned}
&y_1 = x_1, \quad y_1 := y_1/\|y_1\| \\
&y_2 = x_2 - (y_1' x_2) y_1, \quad y_2 := y_2/\|y_2\| \\
&y_3 = x_3 - (y_1' x_3) y_1 - (y_2' x_3) y_2, \quad y_3 := y_3/\|y_3\| \\
&\vdots \\
&y_m = x_m - (y_1' x_m) y_1 - \cdots - (y_{m-1}' x_m) y_{m-1}, \\
&\qquad y_m := y_m/\|y_m\|
\end{aligned}
$$

- Every Euclidean space has an **orthonormal basis**.

- The standard bases in $\mathcal{C}^n$ and $\mathcal{R}^n$ are orthonormal (w.r.t. standard norm and inner product).

- The basis $\{\exp(jkt) : k = \ldots, -1, 0, 1, \ldots\}$ for the space $\mathcal{C}[0, 2\pi]$ is orthonormal (w.r.t. standard inner product and norm).

- The basis $\{x_1, \ldots, x_n\} = \{1, \ldots, t^{n-1}\}$ for the space $\mathcal{P}^{n-1}[-1, 1]$ is not orthogonal w.r.t. inner product:

$$<f, g> = \int_{-1}^{1} \bar{g}(t) f(t) dt.$$

To obtain an orthogonal basis, apply the Gram-Schmidt orthogonalisation procedure:

$$
\begin{aligned}
y_1 &= 1 \\
y_2 &= t - \frac{<t, 1>}{<1, 1>} 1 = t \\
y_3 &= t^2 - \frac{<t^2, 1>}{<1, 1>} 1 - \frac{<t, t^2>}{<t, t>} t = t^2 - \frac{1}{3} \\
&\vdots
\end{aligned}
$$

The polynomials $\{y_1, \ldots, y_n\}$ constructed in this way are called the **Legendre polynomials**.

- **Proposition**. Let $E(\mathcal{C})$ be an n-dimensional Euclidean space and let $S = \{u_1, \ldots, u_n\}$ be an orthonormal basis for $E(\mathcal{C})$. Let $x, y \in E(\mathcal{C})$ have a coordinate vectors $\hat{x}$ and $\hat{y}$ respectively. Then,

$$\hat{x} = \begin{bmatrix} <x, u_1> \\ <x, u_2> \\ \vdots \\ <x, u_n> \end{bmatrix} \in \mathcal{C}^n.$$

Furthermore,

$$<x, y> = \hat{y}'\hat{x}.$$

In particular,

$$<x, x> = \hat{x}'\hat{x} \qquad \textbf{(Parseval's Theorem)}$$

- **Proof**. Let $x = \alpha_1 u_1 + \cdots + \alpha_i u_i + \cdots + \alpha_n u_n$. Then,

$$\begin{aligned} <x, u_i> &= <\alpha_1 u_1 + \cdots + \alpha_i u_i + \cdots + \alpha_n u_n, u_i> \\ &= \alpha_i . \end{aligned}$$

This proves the first result. The second result (and hence Parseval's Theorem) is proved by expanding $<x, y>$ and using the fact that S is orthonormal.

- The result continues to hold for more general inner-product spaces.

Orthogonal projection and best approximation

- Let U be a subspace of $\mathcal{C}^n$ and let $\{u_1, \ldots, u_m\}$ be an orthonormal basis for U.

- Let $U^\perp$ denote the orthogonal complement of U in $\mathcal{C}^n$.

- Then, every vector y in $\mathcal{C}^n$ can be written uniquely as,

$$y = \hat{y} + z,$$

where,

$$\hat{y} = <y, u_1> u_1 + \cdots + <y, u_m> u_m,$$

is in U and z is in $U^\perp$.

- Furthermore, Pythagoras' Theorem implies that,

$$<y, y> \;=\; <\hat{y}, \hat{y}> + <z, z> .$$

- These results admit straightforward generalisations to more general inner-product spaces, and can be proved along the same lines as the previous proposition.

- $\hat{y}$ is called the **orthogonal projection** of y onto the subspace U.

- It is the (unique) closest point in U to y (or the **best approximation** of y in U), in the sense that,

$$\|y - \hat{y}\| < \|y - u\|,$$

  for all u in U distinct from $\hat{y}$.

- Consider the problem of approximating $y = t^5$ by a degree one polynomial over the interval $[-1, 1]$:
  - An orthonormal basis for $P^1[-1, 1]$ consists of the first two normalised Legendre polynomials,

$$\sqrt{\frac{1}{2}}, \quad \sqrt{\frac{3}{2}}t.$$

  - Hence, the best degree one approximation of t^5 in $[-1, 1]$ is given by,

$$\begin{aligned}\hat{y} &= <t^5, \frac{1}{\sqrt{2}}> \frac{1}{\sqrt{2}} + <t^5, \sqrt{\frac{3}{2}}t> \sqrt{\frac{3}{2}}t \\ &= \frac{3}{7}t.\end{aligned}$$

Linear transformations

- Let U & V be vector spaces over the field $\mathcal{F}$, and let $\mathcal{L}: U \to V$ be a **transformation** from U to V.

- $\mathcal{L}$ is called a **linear transformation** if

$$\mathcal{L}(\alpha x + \beta y) = \alpha\mathcal{L}(x) + \beta\mathcal{L}(y),$$

  for all x and y in U and all α and β in $\mathcal{F}$.

- Notation:
  - U is called the **domain** of $\mathcal{L}$ and V is called the **target** of $\mathcal{L}$.
  - If u is in U, then $\mathcal{L}(u)$ is called the **image** of u under $\mathcal{L}$.
  - If $S = \{u_1, \ldots, u_n\}$ is a subset of V, then the set $\mathcal{L}(S) = \{\mathcal{L}(u_1), \ldots, \mathcal{L}(u_n)\} \subset V$ is called the image of S under $\mathcal{L}$.

- The following are linear transformations:
  1. Let $\mathcal{L}_\theta : \mathcal{R}^2 \to \mathcal{R}^2$ denote anticlockwise rotation by an angle $\theta \in [0, \pi]$.
  2. Let $\mathcal{L} : \mathcal{R}^2 \to \mathcal{R}^2$ denote reflection about the x-axis.
  3. Let $\mathcal{L}_D : \mathcal{P}^n \to \mathcal{P}^{n-1}$ denote differentiation.

- The **range** of $\mathcal{L}$ is defined by,

$$\mathcal{R}(\mathcal{L}) = \{\mathcal{L}(x) : \ x \in U\} = \mathcal{L}(U),$$

and is always a subspace of V.

- The **kernel** (or **nullspace**) of $\mathcal{L}$ is defined by,

$$\mathcal{N}(\mathcal{L}) = \{x \in U : \ \mathcal{L}(x) = 0\}.$$

and is always a subspace of U.

- The **rank** of $\mathcal{L}$ is the dimension of the range of $\mathcal{L}$, $\rho(\mathcal{L}) = \dim\{\mathcal{R}(\mathcal{L})\}$.

- The **nullity** of $\mathcal{L}$, denoted by $\nu(\mathcal{L})$, is the dimension of the nullspace of $\mathcal{L}$, $\nu(\mathcal{L}) = \dim\{\mathcal{N}(\mathcal{L})\}$.

- Let $\mathcal{L}: U \to V$ be a given linear transformation and let $y \in V$ be given. The equation $\mathcal{L}(x) = y$ has,
  - no solutions if $y \notin \mathcal{R}(\mathcal{L})$,
  - exactly 1 solution if $y \in \mathcal{R}(\mathcal{L})$ and $\mathcal{N}(\mathcal{L}) = \{0\}$,
  - infinite solutions if $y \in \mathcal{R}(\mathcal{L})$ and $\dim\{\mathcal{N}(\mathcal{L})\} > 0$. For if $\mathcal{L}(x) = y$, then

$$\mathcal{L}(x + z) = \mathcal{L}(x) + \mathcal{L}(z) = y, \ \forall z \in \mathcal{N}(\mathcal{L}).$$

Matrix representation of linear transformations

- **Proposition.** Let $U(\mathcal{F})$ and $V(\mathcal{F})$ be n- and m- dimensional Euclidean spaces, respectively. Any linear transformation $\mathcal{L} : U(\mathcal{F}) \to V(\mathcal{F})$ may be represented by a matrix via choosing bases as follows:
  - Let $B_u = \{u_1, \ldots, u_n\}$ be a basis for $U(\mathcal{F})$.
  - Let $B_v = \{v_1, \ldots, v_m\}$ be a basis for $V(\mathcal{F})$.
  - Define the matrix $L \in \mathcal{F}^{m \times n}$ as follows: the ith column of L is the coordinate vector of $\mathcal{L}(u_i)$.

  Let $x \in U(\mathcal{F})$ have a coordinate vector $\hat{x}$. Then,

$$y = \mathcal{L}(x),$$

has the coordinate vector,

$$\hat{y} = L\hat{x}.$$

• **Proof**. Let,

$$L = \begin{bmatrix} l_{11} & \cdots & l_{1n} \\ \vdots & \ddots & \vdots \\ l_{m1} & \cdots & l_{mn} \end{bmatrix} \in \mathcal{F}^{m\times n},$$

and,

$$\hat{x} = \begin{bmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_n \end{bmatrix} \in \mathcal{F}^n.$$

Then,

$$\begin{aligned} y &= \mathcal{L}(x) \\ &= \mathcal{L}(\sum_i \hat{x}_i u_i) \\ &= \sum_i \hat{x}_i \mathcal{L}(u_i) \\ &= \sum_i \hat{x}_i (\sum_j l_{ji} v_j) \\ &= \sum_j (\sum_i l_{ji} \hat{x}_i) v_j. \end{aligned}$$

So the jth coordinate of $\mathcal{L}(x)$ is $\Sigma_i l_{ji}\hat{x}_i$, and the result follows.

• Let $\mathcal{L}_\theta$ denote anticlockwise rotation by an angle θ of vectors in $\mathcal{R}^2$:

$$\mathcal{L}_\theta(e_1) = e_1 \cos\theta + e_2 \sin\theta \Rightarrow \text{coordinate:} \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$$

$$\mathcal{L}_\theta(e_2) = -e_1 \sin\theta + e_2 \cos\theta \Rightarrow \text{coordinate:} \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$$

$$\Rightarrow L = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

• Let $\mathcal{L}$ denote reflection about the x-axis in $\mathcal{R}^2$:

$$\mathcal{L}(e_1) = e_1, \quad \mathcal{L}(e_2) = -e_2, \quad \Rightarrow \quad L = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

• Let $\mathcal{L}_D : \mathcal{P}^4 \to \mathcal{P}^3$ denote differentiation:

$$\mathcal{L}_D(1) = 0 \Rightarrow \text{coord.} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathcal{L}_D(t) = 1 \Rightarrow \text{coord.} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

$$\mathcal{L}_D(t^2) = 2t \Rightarrow \text{coord.} \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}, \quad \mathcal{L}_D(t^3) = 3t^2 \Rightarrow \text{coord.} \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix},$$

$$\Rightarrow L = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

- Conversely, any $m \times n$ matrix defines a linear transformation (via choosing bases) between n and m-dimensional Euclidean spaces $U(\mathcal{F})$ and $V(\mathcal{F})$.

- A matrix representation of a linear transformation is not unique (since bases are non-unique).

- Let $U(\mathcal{F})$ and $V(\mathcal{F})$ be, respectively, n- and m-dimensional Euclidean spaces with given bases. Since:

  1. The coordinate representation sets up isomorphisms between $U(\mathcal{F})$ and $\mathcal{F}^n$ and between $V(\mathcal{F})$ and $\mathcal{F}^m$,
  2. Any linear transformation $\mathcal{L} : U(\mathcal{F}) \to V(\mathcal{F})$ has a matrix representation $L \in \mathcal{F}^{m\times n}$.

  Then, we can confine our attention to spaces such as $\mathcal{F}^n$ and $\mathcal{F}^m$ and matrices such as $L \in \mathcal{F}^{m\times n}$.

- Let $L \in \mathcal{F}^{m\times n}$. We can consider L as either,

  1. a linear transformation from $\mathcal{F}^n$ to $\mathcal{F}^m$, or,
  2. the matrix representation of a linear transformation between n and m-dimensional Euclidean spaces $U(\mathcal{F})$ and $V(\mathcal{F})$, respectively (with respect to some bases for $U(\mathcal{F})$ and $V(\mathcal{F})$).

Change of basis and similarity

- Suppose that the linear transformation $\mathcal{L} : V \to V$ has a matrix representation $L_1 \in \mathcal{C}^{n\times n}$ w.r.t. a basis B_1 and let $x \in V$ have a coordinate vector $\hat{x}_1$ w.r.t. B_1. Then $y = \mathcal{L}(x)$ has a coordinate vector,

$$\hat{y}_1 = L_1 \hat{x}_1.$$

- Let B_2 be another basis. Then the coordinate vectors of x and y w.r.t. B_2 are given by,

$$\hat{x}_2 = T\hat{x}_1, \qquad \hat{y}_2 = T\hat{y}_1,$$

  respectively, for some nonsingular $T \in \mathcal{C}^{n\times n}$ (in fact, the ith column of T is the coordinate vector of the ith basis vector in B_1 w.r.t. B_2). Hence,

$$\begin{aligned} \hat{y}_2 &= T\hat{y}_1 = TL_1\hat{x}_1 \\ &= TL_1T^{-1}\hat{x}_2, \end{aligned}$$

  and so the matrix representation of $\mathcal{L}$ w.r.t. B_2 is,

$$L_2 = TL_1T^{-1}.$$

- If $T \in \mathcal{C}^{n\times n}$ is nonsingular, then L and TLT^{-1} are said to be **similar**. Similar matrices represent the same linear transformation (w.r.t. different bases).

Matrices as linear transformations

- Let $L \in \mathcal{C}^{m\times n}$. If l_{ij} is the (i,j)th entry of L, we write $L = [l_{ij}]$.

- The **transpose** of $L = [l_{ij}] \in \mathcal{C}^{m\times n}$, denoted by L^T, is that matrix in $\mathcal{C}^{n\times m}$ whose entries are l_{ji}. That is, $L^T = [l_{ji}] \in \mathcal{C}^{n\times m}$.

- The **Hermitian adjoint** (sometimes called the **conjugate transpose**) of $L = [l_{ij}] \in \mathcal{C}^{m\times n}$, denoted by L', is defined as $L' = [\bar{l}_{ji}] \in \mathcal{C}^{n\times m}$.

- The transpose, the Hermitian adjoint and the inverse all obey the **reverse order law**:

$$\begin{aligned}(AB)^T &= B^T A^T,\\ (AB)' &= B'A',\\ (AB)^{-1} &= B^{-1}A^{-1},\end{aligned}$$

(whenever the respective inverses exist).

- Any vector $x \in \mathcal{C}^n$ can be regarded as a matrix $x \in \mathcal{C}^{n\times 1}$. A scalar $\alpha \in \mathcal{C}$ can be regarded as a matrix $\alpha \in \mathcal{C}^{1\times 1}$.

- The **range** of a matrix $L \in \mathcal{C}^{m\times n}$ is defined by,

$$\mathcal{R}(L) = \{Lx : x \in \mathcal{C}^n\},$$

and is always a subspace of $\mathcal{C}^m$.

- The **rank** of a matrix $L \in \mathcal{C}^{m\times n}$, denoted by $\rho(L)$, is the dimension of the range of L.

- Let $L = [c_1 \ \dots \ c_n] \in \mathcal{C}^{m\times n}$, where $c_i \in \mathcal{C}^m$ is the ith column of L. Then,

$$\begin{aligned}\mathcal{R}(L) &= \{Lx : x \in \mathcal{C}^n\}\\ &= \{[c_1 \ \dots \ c_n]\begin{bmatrix} x_1\\ \vdots \\ x_n\end{bmatrix} : x_i \in \mathcal{C}\}\\ &= \{\sum_{i=1}^{n} c_i x_i : x_i \in \mathcal{C}\}\\ &= \text{span}\{c_1, \dots, c_n\}\end{aligned}$$

Hence, the rank of L is the number of linearly independent columns of L.

- The **row rank** of $L \in \mathcal{C}^{n\times m}$ is the number of linearly independent rows of L. However, row rank = column rank, or,

$$\rho(L) = \rho(L^T).$$

- The **kernel** (or **nullspace**) of $L \in \mathcal{C}^{m\times n}$ is defined by,

$$\mathcal{N}(L) = \{x \in \mathcal{C}^n : \ Lx = 0\}$$

and is always a subspace of $\mathcal{C}^n$.

- The **nullity** of L, denoted by $\nu(L)$, is the dimension of the kernel of L.

- Since $\mathcal{N}(L)$ consists of all those vectors which are orthogonal to the rows of L (or to the columns of L^T), we have,

$$\mathcal{N}(L) = [\mathcal{R}(L^T)]^{\perp}.$$

It follows that $\rho(L^T) + \nu(L) = n$, and so,

$$\rho(L) + \nu(L) = n.$$

- Let $T \in \mathcal{C}^{n\times n}$ be nonsingular. Then $L \in \mathcal{C}^{n\times n}$ and $T^{-1}LT$ are similar, and so represent the same linear transformation. Hence,

$$\begin{aligned} \rho(L) &= \rho(T^{-1}LT), \\ \nu(L) &= \nu(T^{-1}LT). \end{aligned}$$

- Let $L \in \mathcal{C}^{m\times n}$ and $y \in \mathcal{C}^m$ be given. The equation,

$$y = Lx,$$

has,

  - no solutions if $y \notin \mathcal{R}(L)$,
  - exactly 1 solution if $y \in \mathcal{R}(L)$ and $\mathcal{N}(L) = 0$,
  - infinite solutions if $y \in \mathcal{R}(L)$ and $\dim\{\mathcal{N}(L)\} > 0$. (For if $Lx = y$, then $L(x+z) = Lx + Lz = y$, $\forall z \in \mathcal{N}(L)$).

- Equivalently, the equation has,

  - no solutions if $\rho(L) < \rho([L \ \ y])$,
  - exactly 1 solution if $\rho(L) = \rho([L \ \ y])$ and the only solution to $Lx = 0$ is $x = 0$,
  - an infinite number of solutions if $\rho(L) = \rho([L \ \ y])$ and there exists a nonzero solution to $Lx = 0$.

- Notice that these are the only possibilities, e.g., the equation $y = Lx$ cannot have only two solutions.

- Let $L \in \mathcal{C}^{m\times n}$ be given. The equation, $y = Lx$, has a solution for every y if and only if $\mathcal{R}(L) = \mathcal{C}^m$, or equivalently, if and only if $\rho(L) = m$.

- $L \in C^{n \times n}$ is called **nonsingular** if $\rho(L) = n$. It follows that $Lx = y$ has a unique solution $x = L^{-1}y$ for every y.
- Let $L \in C^{n \times n}$. The following are equivalent.
  1. $Lx = y$ has a unique solution for every y.
  2. $Lx = 0$ if and only if $x = 0$.
  3. $\rho(L) = n$.
  4. L is nonsingular.
  5. $\det(L) \neq 0$.
- Let $A, B \in C^{n \times n}$. Then,
  1. In general,
  $$\det(A + B) \neq \det(A) + \det(B).$$
  2. $\det(AB) = \det(A)\det(B)$,
  3. $\det(I + AB) = \det(I + BA)$,
  4. $\det(A) = (\det(A'))'$,
  5. $\det(kA) = k^n \det(A)$,
  6. If A is nonsingular,
  $$\det(A^{-1}) = \frac{1}{\det(A)}.$$

Eigenvalues and eigenvectors

- A scalar $\lambda \in C$ is called an **eigenvalue** of $A \in C^{n \times n}$ if there exists a vector $x \in C^n$ such that,
$$Ax = \lambda x, \qquad x \neq 0.$$
x is called an **eigenvector** of A associated with λ.
- Remarks:
  1. Eigenvalues are only defined for square matrices.
  2. An eigenvector cannot be the zero vector.
  3. If x is an eigenvector associated with λ, then so is αx for any nonzero scalar α.
  4. When dealing with eigenvalues, we have to work with complex vector spaces since the eigenvalues of a real matrix may be complex.
- The set of all eigenvalues of $A \in C^{n \times n}$ is called the **spectrum** of A and is denoted by $\sigma(A)$.
- Properties:
  - $A \in C^{n \times n}$ is singular if and only if $0 \in \sigma(A)$.
  - If $\lambda \in \sigma(A)$, then $\lambda^k \in \sigma(A^k)$, for any $k > 1$.
  - If $T \in C^{n \times n}$ is nonsingular, $\sigma(A) = \sigma(T^{-1}AT)$.

The characteristic polynomial

- We can write the eigenvalue-eigenvector equation as,

$$(\lambda I - A)x = 0, \qquad x \neq 0.$$

Thus, $\lambda \in \sigma(A)$ if and only if $\lambda I - A$ is singular, that is,

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

- The polynomial $p(s) = \det(sI - A)$ is called the **characteristic polynomial** of A. The leading coefficient of $p(s)$ is $+1$ and so $p(s)$ has degree n.

- The eigenvalues of A are precisely the zeros of the characteristic polynomial $p(s)$.

- The (algebraic) **multiplicity** of an eigenvalue λ is the multiplicity of λ as a zero of $p(s)$. An eigenvalue λ is called **simple** if the multiplicity of λ is one.

- Any matrix $A \in C^{n\times n}$ has n eigenvalues (counting multiplicities). In fact, if $\sigma(A) = \{\lambda_1, \ldots, \lambda_n\}$, where we repeat eigenvalues according to multiplicity, then,

$$p(s) = (s - \lambda_1)\cdots(s - \lambda_n).$$

Matrix diagonalisation

- Let $A \in C^{n\times n}$ have eigenvalues $\lambda_1, \ldots, \lambda_n$ and let $t_1, \ldots, t_n$ be the corresponding eigenvectors. Define,

$$T = \begin{bmatrix} t_1 & t_2 & \cdots & t_n \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Then, $AT = T\Lambda$. Furthermore, if $t_1, \ldots, t_n$ are linearly independent, then,

$$T^{-1}AT = \Lambda,$$

and A is said to be diagonalisable.

- $A \in C^{n\times n}$ is diagonalisable if and only if it has n linearly independent eigenvectors.

- Eigenvectors associated with distinct eigenvalues are linearly independent.

- Suppose that $A \in C^{n\times n}$ has n distinct eigenvalues. Then the corresponding n eigenvectors are linearly independent, and A is diagonalisable.

- The case of repeated eigenvalues is more difficult, and is explained in more detail in introductory textbooks on linear algebra.

- $A \in \mathcal{C}^{n\times n}$ is called a **normal matrix** if,

$$AA' = A'A.$$

- Normal matrices include the following special cases:

  1. Diagonal matrices.
  2. Hermitian matrices: $A = A'$. When A is real, this becomes $A = A^T$ and A is called symmetric.
  3. Skew Hermitian matrices: $A = -A'$. When A is real, this becomes $A = -A^T$ and A is called skew symmetric.
  4. Unitary matrices: non-singular matrices such that $A^{-1} = A'$. When A is real, this becomes $A^{-1} = A^T$ and A is said to be orthogonal.

- Properties of normal matrices:

  1. Any normal matrix A can be diagonalised by a unitary matrix T: $\Lambda = T'AT$.
  2. The eigenvalues of a Hermitian matrix are all real.
  3. The eigenvalues of a skew Hermitian matrix are all imaginary.
  4. The eigenvalues of a unitary matrix all lie on the unit circle in $\mathcal{C}$.

Matrix exponential function

- **Cayley-Hamilton Theorem.**
  Let $A \in \mathcal{C}^{n\times n}$ and let $p(s) = \det(sI - A)$ be the characteristic polynomial. Then,

$$p(A) = 0.$$

- One consequence is that

$$\begin{aligned} A^n &= \beta_{1,0} I + \beta_{2,0} A + \cdots + \beta_{n-1,0} A^{n-1}, \\ A^{n+1} &= \beta_{1,1} I + \beta_{2,1} A + \cdots + \beta_{n-1,1} A^{n-1}, \\ &\vdots \\ A^{n+j} &= \beta_{1,j} I + \beta_{2,j} A + \cdots + \beta_{n-1,j} A^{n-1}, \end{aligned}$$

  for some scalars $\beta_{i,j}$.

- For any $A \in \mathcal{C}^{n\times n}$, the series,

$$\sum_{i=0}^{\infty} \frac{A^i}{i!},$$

  converges and the limit is defined as the **matrix exponential**:

$$\exp(A) = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots .$$

- Suppose that $A \in C^{n\times n}$ and let $T \in C^{n\times n}$ be any nonsingular matrix. Then,

$$(T^{-1}AT)^i = T^{-1}A^iT, \qquad i = 0, 1, \ldots.$$

- Suppose that $A \in C^{n\times n}$ is diagonalisable so that,

$$A = T\Lambda T^{-1},$$

for nonsingular $T \in C^{n\times n}$ and diagonal $\Lambda \in C^{n\times n}$. Then,

$$\begin{aligned} \exp(A) &= I + T\Lambda T^{-1} + \frac{T\Lambda^2T^{-1}}{2!} + \cdots \\ &= T\exp(\Lambda)T^{-1}. \end{aligned}$$

- Some useful facts about the matrix exponential:

1. $\exp(A)\exp(B) \neq \exp(A+B)$ (unless $AB = BA$).
2. $\exp(-A) = [\exp(A)]^{-1}$,
3. $A\exp(A) = \exp(A)A$.
4. $\frac{d}{dt}\exp(At) = A\exp(At) = \exp(At)A$.

Linear Optimal Control - Tutorial 1

Linear algebra

1. Compute the Legendre polynomials of order 0, 1, 2, 3, 4 and 5 in the interval $[-1, 1]$. Compute their norm and modify the polynomial to obtain an orthonormal set of polynomials.

2. Let $\mathcal{L}$ be a linear transformation. Show, using the very definition of subspace, that $\mathcal{R}(\mathcal{L})$ and $\mathcal{N}(\mathcal{L})$ are indeed subspaces.

3. Let ${\mathcal{L}_\mathcal{D}}^2 : P^5[-1,1] \to P^3[-1,1]$ be the second order derivative operator. Compute the matrix representation L of this operator with respect to the basis composed of the Legendre polynomials of order 0, 1, 2, 3, 4 and 5. Compute range and kernel of this operator.

4. Repeat Exercise 3 using as a new basis the set of polynomials 1, t, t^2, t^3, t^4 and t^5. Let $\tilde{L}$ be the matrix representation of the operator ${\mathcal{L}_\mathcal{D}}^2 : P^5[-1,1] \to P^3[-1,1]$ with respect to the new basis. Show that L (in Exercise 3) and $\tilde{L}$ are similar matrices.

5. Let $L \in I\!R^{n\times n}$. Prove that eigenvectors of L corresponding to distinct eigenvalues are linearly independent.

6. The claim: two matrices $L \in I\!R^{n\times n}$ and $\tilde{L} \in I\!R^{n\times n}$ are similar if and only if they have the same spectrum, is true or false? Motivate your answer.

7. Let $A \in I\!R^{n\times n}$. Using Cayley-Hamilton Theorem write A^{-1} as a function of $A^0 = I, A, ..., A^{n-1}$.

8. Let $A \in I\!R^{n\times n}$ and $p(s) = \det(sI - A) = s^n + a_{n-1}s^{n-1} + \ldots + a_1 s + a_0$. Show that $|a_{n-1}| = \text{trace}(A)$ and $|a_0| = \det(A)$. The trace(A) is the sum of all diagonal elements of the matrix A.

9. Compute the exponential of the following matrices

$$\begin{bmatrix} 0 & -\omega \\ \omega & 0 \end{bmatrix}, \quad \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}.$$