Contents

• Vector spaces
• Subspaces
• Linear independence and span
• Bases and dimension
• Isomorphism and coordinate representation
• Inner product and norm
• Angles and orthogonality
• Gram-Schmidt orthogonalisation process
• Orthogonal projection and best approximation
• Linear transformations
• Matrix representation of linear transformations
• Change of basis and similarity
• Matrices as linear transformations
• Eigenvalues and eigenvectors
• The characteristic polynomial
• Matrix diagonalisation
• Matrix exponential function

A short summary of linear algebra and matrix theory

Dr I M Jaimoukha
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 of objects, called vectors, for which two operations, vector addition (+) and scalar multiplication (·) 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. ·)
  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. \( \alpha \cdot (x + y) = \alpha \cdot x + \alpha \cdot y \) (· distributive w.r.t. +)
  8. \( \alpha \cdot (\beta \cdot x) = (\alpha \beta) \cdot x \) (· 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:
     \[ \mathcal{R}^n(\mathcal{R}) = \{ [x_1, \ldots, x_n] \mid x_i \in \mathcal{R}, \forall i \} \]
     \[ \mathcal{C}^n(\mathcal{C}) = \{ [x_1, \ldots, x_n] \mid x_i \in \mathcal{C}, \forall i \} \]
  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 $V(F)$ is called a subspace if $S$ is itself a vector space over $F$.

- A nonempty subset $S$ of $V(F)$ is a subspace of $V(F)$ if it satisfies the closure axioms:
  1. $x, y \in S \implies x + y \in S$.
  2. $x \in S, \alpha \in F \implies \alpha x \in S$.

- Every vector space $V(F)$ has two special subspaces:
  1. The zero subspace $\Phi = \{0\}$.
  2. The vector space $V(F)$ itself.

- Any other subspace of $V(F)$ is called a proper subspace.

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

- The set of all vectors in the first quadrant is not a subspace of $R^2$.

Linear independence and span

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

- A set of vectors $\{x_1, \ldots, x_n\} \subset V(F)$ is called linearly independent if:
  $$\alpha_1 x_1 + \cdots + \alpha_n x_n = 0 \iff \alpha_i = 0, \forall i.$$

- Let $S = \{x_1, \ldots, x_n\} \subset V(F)$. The expression,
  $$\sum_{i=1}^{n} \alpha_i x_i,$$
  where $\alpha_1, \ldots, \alpha_n \in F$, is called a linear combination of the vectors in $S$.

- The span of $S = \{x_1, \ldots, x_n\} \subset V(F)$ is the set of all linear combinations of the vectors in $S$:
  $$\text{span}(S) = \{x = \sum_{i=1}^{n} \alpha_i x_i : \alpha_i \in F\},$$
  and is always a subspace of $V(F)$.  

Bases and dimension

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

- Every vector space has a basis.

- Let $F$ be a field. The set $S = \{e_1, \ldots, e_n\}$, where,
  
  $e_1 = [1, 0, 0, \ldots, 0, 0]^T$

  $e_2 = [0, 1, 0, \ldots, 0, 0]^T$

  $\vdots$

  $e_n = [0, 0, 0, \ldots, 0, 1]^T,$

is a basis for $F^n$, called the natural basis.

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

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

Isomorphism and coordinate representation

- Let $U$ and $V$ be vector spaces over the same field $F$. 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 $\alpha, \beta \in F$ (linear).

- Two vector spaces $U$ and $V$ over the same field $F$ 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 $U(F)$ and $V(F)$ are isomorphic if and only if $\dim U(F) = \dim V(F)$.
• 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 $\alpha_i$ 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. $\langle x, y \rangle = \langle y, x \rangle$
2. $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$, $\forall \alpha, \beta \in \mathcal{C}$
3. $\langle x, x \rangle > 0$, $\forall x \neq 0$

• 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):

$$\langle x, \alpha y + \beta z \rangle = \overline{\alpha} \langle x, y \rangle + \overline{\beta} \langle x, z \rangle, \forall \alpha, \beta \in \mathcal{C}$$

– positive definite (3).

• The standard inner product on $\mathcal{C}^n$ is given by $\langle x, y \rangle = y'x$ where $y' = [y_1, \ldots, y_n]$.

• 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,

$$\langle f, g \rangle = \frac{1}{2\pi} \int_{0}^{2\pi} \overline{g(t)} f(t) dt.$$
A norm on a complex vector space $V$ is a function from $V$ to $\mathbb{C}$ which satisfies:
1. $\|x\| > 0, \quad \forall x \neq 0$
2. $\|\alpha x\| = |\alpha| \|x\|, \quad \forall \alpha \in \mathbb{C}$ or $\mathbb{R}$
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 $\mathbb{C}^n$ is $\|x\| = \sqrt{x', x}$, and the standard norm on $\mathbb{C}[0, 2\pi]$ is,
$$\|f\| = \sqrt{\langle f, f \rangle} = \frac{1}{\sqrt{2\pi}} \sqrt{\int_0^{2\pi} |f(t)|^2 \, dt}.$$

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 $x, y \in \mathbb{C}^n$. Then,
  $$|x'y| \leq \|x\| \|y\|$$  \text{(Schwarz inequality)}

- The angle between nonzero $x, y \in \mathbb{C}^n$ 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 $S = \{x_1, \ldots, x_m\} \subset \mathbb{C}^n$ 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 $\mathbb{C}^n$ is linearly independent.

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

- Let $U$ be a subspace of $\mathbb{C}^n$. The subspace,
  $$U^\perp = \{x \in \mathbb{C}^n : x'u = 0, \forall u \in U\},$$
  is called the orthogonal complement of $U$ (in $\mathbb{C}^n$).
Gram-Schmidt orthogonalisation process

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

\[
\begin{align*}
y_1 &= x_1 \\
y_2 &= x_2 - \frac{y_1^* x_2}{y_1^* y_1} y_1 \\
y_3 &= x_3 - \frac{y_1^* x_3}{y_1^* y_1} y_1 - \frac{y_2^* x_3}{y_2^* y_2} y_2 \\
& \quad \vdots \\
y_m &= x_m - \frac{y_1^* x_m}{y_1^* y_1} y_1 - \cdots - \frac{y_{m-1}^* x_m}{y_{m-1}^* y_{m-1}} y_{m-1}
\end{align*}
\]

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

\[
\begin{align*}
y_1 &= x_1, \quad y_1 := y_1 / \|y_1\| \\
y_2 &= x_2 - (y_1^* x_2) y_1, \quad y_2 := y_2 / \|y_2\| \\
y_3 &= x_3 - (y_1^* x_3) y_1 - (y_2^* x_3) y_2, \quad y_3 := y_3 / \|y_3\| \\
& \quad \vdots \\
y_m &= x_m - (y_1^* x_m) y_1 - \cdots - (y_{m-1}^* x_m) y_{m-1}, \quad y_m := y_m / \|y_m\|
\end{align*}
\]

- Every Euclidean space has an orthonormal basis.

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

- The basis \( \{\exp(jkt) : k = \ldots, -1, 0, 1, \ldots\} \) for the space \( \mathbb{C}[0, 2\pi] \) 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:

\[
\langle f, g \rangle = \int_{-1}^{1} g(t) f(t) \, dt.
\]

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

\[
\begin{align*}
y_1 &= 1 \\
y_2 &= t - \frac{\langle t, 1 \rangle}{\langle 1, 1 \rangle} 1 = t \\
y_3 &= t^2 - \frac{\langle t^2, 1 \rangle}{\langle 1, 1 \rangle} 1 - \frac{\langle t^2, t \rangle}{\langle t, t \rangle} t = t^2 - 1 \\
& \quad \vdots
\end{align*}
\]

The polynomials \( \{y_1, \ldots, y_n\} \) constructed in this way are called the Legendre polynomials.
• **Proposition.** Let \( E(C) \) be an \( n \)-dimensional Euclidean space and let \( S = \{ u_1, \ldots, u_n \} \) be an orthonormal basis for \( E(C) \). Let \( x, y \in E(C) \) have a coordinate vectors \( \hat{x} \) and \( \hat{y} \) respectively. Then,

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

Furthermore,

\[<x, y> = \hat{y}^t \hat{x}.\]

In particular,

\[<x, x> = \hat{x}^t \hat{x} \quad \text{(Parseval's Theorem)}\]

• **Proof.** Let \( x = \alpha_1 u_1 + \cdots + \alpha_i u_i + \cdots + \alpha_n u_n \). Then,

\[
<x, u_i> = <\alpha_1 u_1 + \cdots + \alpha_i u_i + \cdots + \alpha_n u_n, u_i> = \alpha_i.
\]

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 \( C^n \) and let \( \{ u_1, \ldots, u_n \} \) be an orthonormal basis for \( U \).

• Let \( U^\perp \) denote the orthogonal complement of \( U \) in \( C^n \).

• Then, every vector \( y \) in \( C^n \) can be written uniquely as,

\[
y = \hat{y} + z,
\]

where,

\[
\hat{y} = <y, u_1> u_1 + \cdots + <y, u_n> u_n,
\]

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,

$$\begin{align*}
\frac{1}{\sqrt{2}}, & \quad \frac{3}{\sqrt{2}} t.
\end{align*}$$

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

$$\hat{y} = \langle t^5, \frac{1}{\sqrt{2}} \rangle \frac{1}{\sqrt{2}} + \langle t^5, \frac{3}{\sqrt{2}} t \rangle \frac{3}{\sqrt{2}} t\left(\frac{3}{\sqrt{2}} t\right)$$

$$= \frac{3}{4} t^5.$$

---

**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 $\alpha$ and $\beta$ 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 \( \mathcal{L} : U \rightarrow V \) 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 \( \mathcal{L} : U(\mathcal{F}) \rightarrow V(\mathcal{F}) \) 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 \( i \)th 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,

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

and,

\[ \hat{x} = \begin{bmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_n \end{bmatrix} \in \mathbb{F}^n. \]

Then,

\[ y = L(x) = \mathcal{L}(\sum_i \hat{x}_i u_i) = \sum_i \hat{x}_i \mathcal{L}(u_i) = \sum_i \hat{x}_i (\sum_j l_{ji} v_j) = \sum_j (\sum_i l_{ji} \hat{x}_i) v_j. \]

So the \( j \)th coordinate of \( \mathcal{L}(x) \) is \( \sum_i l_{ji} \hat{x}_i \), and the result follows.

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

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

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

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

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

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

• Let \( \mathcal{L}_D : \mathcal{P}^4 \rightarrow \mathcal{P}^3 \) denote differentiation:

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

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

\[ \Rightarrow L = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \]
• Conversely, any $m \times n$ matrix defines a linear transformation (via choosing bases) between $n$ and $m$-dimensional Euclidean spaces $U(F)$ and $V(F)$.

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

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

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

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

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

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

Change of basis and similarity

• Suppose that the linear transformation $L : V \to V$ has a matrix representation $L_1 \in C^{n \times n}$ 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 $y = L(x)$ 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,

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

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

  $$\hat{y}_2 = T \hat{y}_1 = TL_1 \hat{x}_1$$

and so the matrix representation of $L$ w.r.t. $B_2$ is,

  $$L_2 = TL_1 T^{-1}.$$ 

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

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

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

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

\[
(AB)^T = B^T A^T, \\
(AB)' = B' A', \\
(AB)^{-1} = B^{-1} A^{-1},
\]

(whenever the respective inverses exist).

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

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

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

and is always a subspace of $C^m$.

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

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

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

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

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

\[
\rho(L) = \rho(L^T).
\]
- The kernel (or nullspace) of $L \in C^{m \times n}$ is defined by,
  \[ \mathcal{N}(L) = \{ x \in C^n : Lx = 0 \} \]
  and is always a subspace of $C^n$.

- 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 $T \in C^{n \times n}$ be nonsingular. Then $L \in C^{n \times n}$ and $T^{-1}LT$ are similar, and so represent the same linear transformation. Hence,
  \[ \rho(L) = \rho(T^{-1}LT), \]
  \[ \nu(L) = \nu(T^{-1}LT). \]

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

\[ \text{26} \]
• $L \in C^{n \times n}$ 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,
  \[
  Ax = \lambda x, \quad x \neq 0.
  \]

  $x$ is called an eigenvector of $A$ associated with $\lambda$.

• 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 $\lambda$, then so is $\alpha x$ for any nonzero scalar $\alpha$.
  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,
  \((\lambda I - A)x = 0, \quad x \neq 0.\)

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 \(\lambda\) is the multiplicity of \(\lambda\) as a zero of \(p(s)\). An eigenvalue \(\lambda\) is called simple if the multiplicity of \(\lambda\) is one.

- Any matrix \(A \in \mathbb{C}^{n \times n}\) 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 \(A \in \mathbb{C}^{n \times n}\) 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, \(AT = TA\). Furthermore, if \(t_1, \ldots, t_n\) are linearly independent, then,
  \[
  T^{-1}AT = \Lambda,
  \]
and \(A\) is said to be diagonalisable.

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

- Eigenvectors associated with distinct eigenvalues are linearly independent.

- Suppose that \(A \in \mathbb{C}^{n \times n}\) 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 ∈ C^{n×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 = AT and A is called symmetric.
  3. Skew Hermitian matrices: A = −A'. When A is real, this becomes A = −AT 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 = AT and A is said to be orthogonal.

• Properties of normal matrices:
  1. Any normal matrix A can be diagonalised by a unitary matrix T: Λ = 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 C.

---

Matrix exponential function

• Cayley-Hamilton Theorem.
  Let A ∈ C^{n×n} and let p(s) = det(sI − A) be the characteristic polynomial. Then,
  \[ p(A) = 0. \]

• One consequence is that
  \[
  A^n = \beta_{1,0}I + \beta_{2,0}A + \cdots + \beta_{n-1,0}A^{n-1},
  \]
  \[
  A^{n+1} = \beta_{1,1}I + \beta_{2,1}A + \cdots + \beta_{n-1,1}A^{n-1},
  \]
  \[ \vdots \]
  \[
  A^{n+j} = \beta_{1,j}I + \beta_{2,j}A + \cdots + \beta_{n-1,j}A^{n-1},
  \]
  for some scalars \( \beta_{i,j} \).

• For any A ∈ C^{n×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,
\[
(T^{-1}AT)^i = T^{-1}A^iT, \quad i = 0, 1, \ldots
\]

• 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,
\[
\exp(A) = I + T\Lambda T^{-1} + \frac{T\Lambda^2 T^{-1}}{2!} + \cdots = T \exp(\Lambda) T^{-1}.
\]

• 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. \( \frac{d}{dt} \exp(At) = A \exp(At) = \exp(At)A \).
9. Compute the exponential of the following matrices.

Show that $\det(A) = \det(A')$ and $\text{trace}(A) = \text{trace}(A')$. The trace is the sum of the diagonal elements of the matrix A.

Let $A \in \mathbb{R}^{n \times n}$ and $p(s) = (s-A)^{-1}$.

8. Let $A \in \mathbb{R}^{n \times n}$ and $\det(s) = (s-f(s))^n$. Use Cayley-Hamilton theorem with $A^{-1}$ as a function of $A$.

7. Let $A \in \mathbb{R}^{n \times n}$. Find $A^{-1}$.

6. The claim: Two matrices $I \in \mathbb{R}^{n \times n}$ and $J \in \mathbb{R}^{n \times n}$ are similar if and only if their values are linearly independent.

5. Let $I \in \mathbb{R}^{n \times n}$. Prove that eigenvectors of $I$ corresponding to distinct eigenvalues are similar matrices.


3. Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Compute the matrix representation of the operator $A$ that transforms polynomials to polynomials of degree $n$.

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

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

Linear Algebra

Linear Optimal Control - Tutorial 1