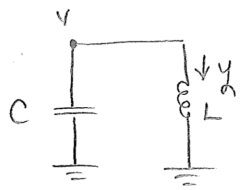
2017

EE501 LINEAR ALGEBRA

Example (where we don't need linear algebra)



system: $\begin{cases} C\dot{v} + y = 0 \\ L\dot{y} = v \end{cases}$ energy : $E = \frac{1}{2} C v^2 + \frac{1}{2} L y^2$

(2nd order) LC osc. $\Rightarrow C\ddot{v} + \dot{y} = 0 \Rightarrow C\ddot{v} + L^{-1} v = 0 \Rightarrow \ddot{v} + C^{-1} L^{-1} v = 0$

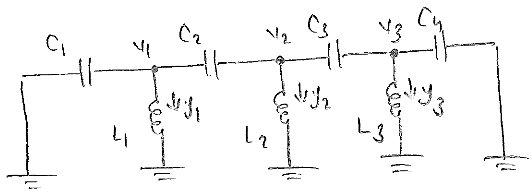
$v, y \in \mathbb{R}$

$C, L \in \mathbb{R}$

sol'n: $v(t) = \alpha \cos(\omega t + \phi)$ α, ϕ : determined by $v(0), y(0)$

$\omega = \sqrt{C^{-1} L^{-1}}$: natural freq.

Example (where linear algebra is useful)



system: $\begin{cases} C\dot{v} + y = 0 \\ L\dot{y} = v \end{cases}$

6th order oscillator

$$C = \begin{bmatrix} C_1 + C_2 & -C_2 & 0 \\ -C_2 & C_2 + C_3 & -C_3 \\ 0 & -C_3 & C_3 + C_4 \end{bmatrix}, \quad L = \begin{bmatrix} L_1 & 0 & 0 \\ 0 & L_2 & 0 \\ 0 & 0 & L_3 \end{bmatrix}$$

$$v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$\Rightarrow v, y \in \mathbb{R}^3, \quad C, L \in \mathbb{R}^{3\times 3}$ C, L : Hermitian pos. def. matrices

sol'n: $v_k(t) = \alpha_1 \cos(\omega_1 t + \phi_1) + \alpha_2 \cos(\omega_2 t + \phi_2) + \alpha_3 \cos(\omega_3 t + \phi_3)$

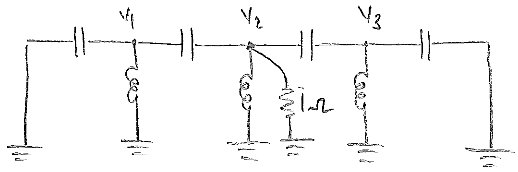
$\omega_1, \omega_2, \omega_3$: eigenvalues (roots of char. poly.) of $\sqrt{C^{-1} L^{-1}}$ ↝ square root of matrix

let $x = \begin{bmatrix} v \\ y \end{bmatrix}$ (state of the system, $x \in \mathbb{R}^6$)

energy : $E = \frac{1}{2} v^T C v + \frac{1}{2} y^T L y = \frac{1}{2} x^T \begin{bmatrix} C & 0 \\ 0 & L \end{bmatrix} x =: E(x)$

$\sqrt{E(x)}$ is a norm on $\mathbb{R}^6$.

How about ?



$\Rightarrow$ system: $\begin{cases} C\dot{v} + Gv + y = 0 \\ L\dot{y} = v \end{cases}$ where $\underset{\text{orth. projection matrix}}{G = e_2 e_2^T}$ with $e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

$$\dot{E}(x(t)) = \underbrace{-x(t)^T \begin{bmatrix} G & 0 \\ 0 & 0 \end{bmatrix} x(t)}_{\text{power dissipated on } 1\Omega} = -v_2^2$$

Question: $E(x(t)) \to 0$?

Answer: Let η_1, η_2, η_3 be the eigenvectors of $C^{-1}L^{-1}$. Then:

if $\eta_k \in$ null space (G) for some $k = 1, 2, 3$ then $E(x(t)) \not\to 0$.

otherwise $E(x(t)) \to 0$.

— ○ —

Keywords appeared in our simple example:

- → Hermitian pos. def.
- → square root (function of a matrix)
- → eigenvalue
- → char. poly.
- → eigenvector
- → orthogonal proj.
- → norm
- → null space

2017

<u>Definition</u> A <u>field</u> is a set F together with two mappings of the kind $F \times F \to F$ called addition & multiplication, denoted $(a,b) \mapsto a+b$ and $(a,b) \mapsto ab$, respectively, with the following properties:

(A1) $a+b = b+a$ for all $a,b \in F$ (commutativity)

(A2) $a+(b+c) = (a+b)+c$ $\forall\ a,b,c \in F$ (associativity)

(A3) There is an element in F, denoted by 0_F, such that

$a + 0_F = a$ $\forall a \in F$ (additive identity)

(A4) For each $a \in F$ there is an element in F, denoted by $-a$, such that $a + (-a) = 0_F$ (additive inverse)

(M1) $ab = ba$ $\forall\ a,b \in F$ (commutativity)

(M2) $a(bc) = (ab)c$ $\forall\ a,b,c \in F$ (associativity)

(M3) There is an element in F, denoted by 1_F, such that

$a 1_F = a$ $\forall a \in F$ (multiplicative identity)

(M4) For each $a \neq 0_F$ there is an element in F, denoted by a^{-1}, such that $aa^{-1} = 1_F$ (multiplicative inverse)

(D) $a(b+c) = ab + ac$ $\forall\ a,b,c \in F$

— o —

<u>Example</u> set of binary numbers with modulo 2 addition & mult.

$F = \{0,1\}$

⊕	0	1
0	0	1
1	1	0

⊙	0	1
0	0	0
1	0	1

Note that $0_F = 0$ & $1_F = 1$

Example Set of real numbers $F=\mathbb{R}$ with standard addition & mult.

Example Let $F=\mathbb{R}\times\mathbb{R}$. Let us define $\oplus$ & $\odot$ as:

Given $x=(x_1,x_2)\in F$, $y=(y_1,y_2)\in F$

$$x\oplus y := (x_1+y_1,\ x_2+y_2)$$
$$x\odot y := (x_1y_1-x_2y_2,\ x_1y_2+x_2y_1)$$

Note that this is no other than complex number field $\mathbb{C}$.

Then $0_F=(0,0)$ & $1_F=(1,0)$

Exercise Let $F=(0,\infty)=:\mathbb{R}_+$ (set of positive real numbers)

Define $x\oplus y := xy$, $x\odot y := e^{\ln x \ln y}$

Show that F satisfies the axioms of field. Find 1_F & 0_F

Linear Space (Vector space)

Definition A linear space V is a set whose elements are called vectors, associated with a field F, whose elements are called scalars. Two vectors can be added (and yield another vector). A vector can be multiplied by a scalar (yielding another vector). Axioms:

(A1) $x+y=y+x$ $\quad x,y\in V$ (commutativity)

(A2) $(x+y)+z=x+(y+z)$ (associativity)

(A3) there is an element 0 in V (called the zero vector) such that

$x+0=x$ (additive identity)

(A4) For each $x\in V$ there is a unique element $-x$ in V such that

$x+(-x)=0$ (additive inverse)

(M1) $a(bx) = (ab)x \qquad a, b \in F, \; x \in V$ (associativity)

(M2) $a(x+y) = ax + ay$ (distributivity)

(M3) $(a+b)x = ax + bx$ (distributivity)

(M4) $1x = x$ (1 is the multiplicative unit in F)

<u>Exercise</u> for a given linear space V show that its zero vector is unique. (pf: $0_1 \overset{(A3)}{=} 0_1 + 0_2 \overset{(A1)}{=} 0_2 + 0_1 \overset{(A3)}{=} 0_2$)

<u>Example</u> Show that $0_F x = 0_V$

Scalar zero in F — any vector in V — zero in V

<u>Proof</u>

$$
\begin{aligned}
0x &\overset{(A3)}{=} 0x + 0 && \text{(A4)}\\
&= 0x + (x + (-x)) && \text{(A2)}\\
&= (0x + x) + (-x) && \text{(M4)}\\
&= (0x + 1x) + (-x) && \text{(M3)}\\
&= (0+1)x + (-x) && (\text{A1})_F\\
&= (1+0)x + (-x) && (\text{A3})_F\\
&= 1x + (-x) && \text{(M4)}\\
&= x + (-x) && \text{(A4)}\\
&= 0
\end{aligned}
$$

$\blacksquare$

<u>Example</u> (linear space) Set of all vectors $(a_1, a_2, \dots, a_n)$ with $a_i \in F$. Addition, multiplication are defined componentwise. This space is denoted F^n.

let $x, y \in F^n$ be $x = (a_1, a_2, \dots, a_n)$ & $y = (b_1, b_2, \dots, b_n)$

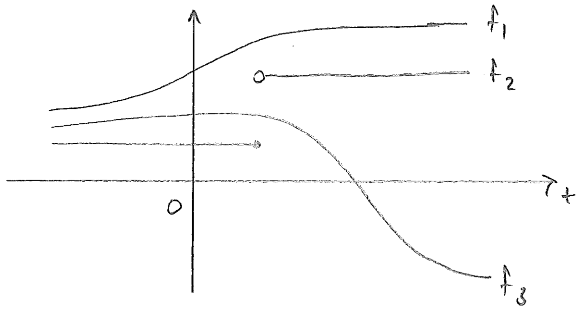
Addition : $x+y = (a_1+b_1, a_2+b_2, \dots, a_n+b_n)$

Scalar mult. $cx = (ca_1, ca_2, \dots, ca_n)$

Most common examples are $\mathbb{R}^n$ & $\mathbb{C}^n$.

2017

<u>Example</u> (linear space) Set of all real-valued functions $t \mapsto f(t)$ defined on the real line with $F = \mathbb{R}$.



Define addition as: $(f_1+f_2)(t) = f_1(t) + f_2(t) \quad \forall t \in \mathbb{R}$

Define multiplic. as: $(af)(t) = af(t) \quad \forall t \in \mathbb{R}$

<u>Example</u> (linear space) set of all polynomials with deg. $\leq n$ with coefficients in F. Note that this linear space is a subset of the previous one for $F=\mathbb{R}$.

— o —

<u>Definition</u> Let V be a linear space defined over field F. A subset $W \subset V$ is called a <u>subspace</u> if the following hold:

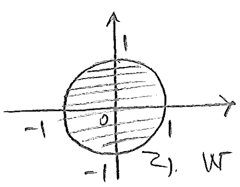
(S1) $w_1 + w_2 \in W \quad \forall\, w_1, w_2 \in W$

(S2) $aw \in W \quad \forall\, w \in W$ & $a \in F$

<u>Example</u> linear space $V = \mathbb{R}^2$

subspace $W = \left\{ \begin{bmatrix} a \\ 0 \end{bmatrix} : a \in \mathbb{R} \right\}$

<u>Example</u> $V = \mathbb{R}^2$



W (unit disc) NOT a subspace

<u>Example</u> linear space V = set of all real valued functions $t \mapsto f(t)$

subspace W_1 = set of all continuous functions $g: \mathbb{R} \to \mathbb{R}$

subspace W_2 = set of all functions $h: \mathbb{R} \to \mathbb{R}$ periodic with period π.

<u>Definition</u> Let X be a linear space & $Y, Z \subset X$ be subsets. We define the <u>sum</u> of Y and Z as $Y+Z := \{y+z : y \in Y \;\&\; z \in Z\}$.

<u>Example</u> Let Y, Z be two subspaces of X. Show that $Y+Z$ is also a subspace.

<u>proof</u> Let $w_1, w_2 \in Y+Z$. Then we can find $y_1, y_2 \in Y$ & $z_1, z_2 \in Z$ satisfying $w_1 = y_1 + z_1$ & $w_2 = y_2 + z_2$. Hence

$$\begin{aligned} w_1 + w_2 &= (y_1+z_1) + (y_2+z_2) \\ &= \underbrace{(y_1+y_2)}_{\in Y} + \underbrace{(z_1+z_2)}_{\in Z} \\ &\in Y+Z \quad \Rightarrow \quad (S1) \text{ is satisfied} \end{aligned}$$

Moreover, for any scalar a,

$$\begin{aligned} a w_1 &= a(y_1+z_1) \\ &= \underbrace{a y_1}_{\in Y} + \underbrace{a z_1}_{\in Z} \\ &\in Y+Z \quad \Rightarrow \quad (S2) \text{ is satisfied, too.} \end{aligned}$$

▨

<u>Example</u> Let Y, Z be subspaces of X. Show that $Y \cap Z$ is also a subspace.

<u>proof</u> Let $w_1, w_2 \in Y \cap Z$

$$\left.\begin{array}{l} w_1, w_2 \in Y \Rightarrow w_1 + w_2 \in Y \\ w_1, w_2 \in Z \Rightarrow w_1 + w_2 \in Z \end{array}\right\} w_1 + w_2 \in Y \cap Z \Rightarrow (S1)$$

Choose any scalar a. Then

$$\left.\begin{array}{l} w_1 \in Y \Rightarrow a w_1 \in Y \\ w_1 \in Z \Rightarrow a w_1 \in Z \end{array}\right\} a w_1 \in Y \cap Z \Rightarrow (S2)$$

▨

<u>Exercise</u> How about $Y \cup Z$?

2017

<u>Definition</u> Let $(\mathcal{V}, F)$ & $(\mathcal{W}, F)$ be two linear spaces defined over the same scalar field F. The <u>product space</u> of $\mathcal{V}$ and $\mathcal{W}$ is defined as

$$\mathcal{V} \times \mathcal{W} = \{ (v, w) : v \in \mathcal{V} , w \in \mathcal{W} \}$$ together with the operations:

$$(v, w) + (\hat{v}, \hat{w}) = (v+\hat{v}, w+\hat{w}) \qquad \text{vector addition}$$

$$a(v, w) = (av, aw) \qquad \text{scalar multiplication}$$

— o —

<u>Definition</u> A <u>linear combination</u> of n vectors $x_1, x_2, \dots, x_n$ of a linear space $\mathcal{X}$ is a vector of the form

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n \qquad (a_i \in F)$$

The <u>span</u> of $\{x_1, x_2, \dots, x_n\}$, denoted $sp\{x_1, x_2, \dots, x_n\}$ is the set of all linear combinations of $x_1, x_2, \dots, x_n$. That is,

$$sp\{x_1, x_2, \dots, x_n\} = \{ \sum_{i=1}^{n} a_i x_i : a_i \in F \}$$

<u>Definition</u> Vectors $x_1, x_2, \dots, x_n \in \mathcal{X}$ are said to be <u>linearly independent</u> if

$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = 0 \implies a_i = 0$ for all i. Otherwise, they are said to be <u>linearly dependent</u>.

<u>Examples</u> $\left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$ lin. ind. set in $\mathbb{R}^2$

$\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} \right\}$ lin. dep. set in $\mathbb{R}^3$

because $$2 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - 4 \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} + 2 \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Example Let V be the linear space of polynomials of deg ≤ 2 with complex coefficients. That is, $p \in V$ means $p(t) = a_0 t^2 + a_1 t + a_2$ with $a_0, a_1, a_2 \in \mathbb{C}$.

Consider the subset $S = \{p_1, p_2, p_3\}$ where $p_1(t) = 1$, $p_2(t) = t$, $p_3(t) = t^2$.

Is S lin. ind ? Apply definition.

$$\underbrace{\alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 = 0}_{\text{zero poly.}} \Rightarrow \alpha_1 p_1(t) + \alpha_2 p_2(t) + \alpha_3 p_3(t) = 0 \quad \forall t$$

$$\Rightarrow \alpha_1 + \alpha_2 t + \alpha_3 t^2 = 0 \quad \forall t$$

$$\Rightarrow \alpha_1 = \alpha_2 = \alpha_3 = 0$$

$$\Rightarrow p_1, p_2, p_3 \text{ are lin. ind.}$$

Example $S = \{\cos t, \sin t, \cos(t - \frac{\pi}{3})\}$

Note that $\cos(t - \frac{\pi}{3}) = \cos\frac{\pi}{3} \cos t + \sin\frac{\pi}{3} \sin t$

$$\Rightarrow -\frac{1}{2}\cos t - \frac{\sqrt{3}}{2}\sin t + 1\cos(t - \frac{\pi}{3}) = 0 \quad \forall t$$

$\Rightarrow$ set S is lin. dep.

— o —

Definition (Basis) Let V be a linear space and (finite) set of vectors $S = \{x_1, x_2, \dots, x_n\}$ be a subset of V. S is said to be a basis for V iff

1) $sp(S) = V$,

2) S is a linearly ind. set.

A linear space has many bases. If one of those bases has $n < \infty$ vectors then every other basis must also have n vectors. The number n is called the dimension of V and V is said to be finite-dimensional.

<u>Example</u> $V = \mathbb{R}^2$

two bases: $S_1 = \left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$; $S_2 = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right\}$

Note that : → each S_i is linearly ind. $i = 1, 2$

→ $sp(S_i) = \mathbb{R}^2$

→ # of elements of $S_i = 2 \Rightarrow \dim V = 2$

<u>Definition</u> An <u>ordered basis</u> $(x_1, x_2, \dots, x_n)$ is such that basis vectors are given a specific ordering.

Let $(x_1, x_2, \dots, x_n)$ be an ordered basis of V. Then for each $y \in V$ there is a unique n-tuple of scalars $(a_1, a_2, \dots, a_n)$ such that

$$y = a_1 x_1 + a_2 x_2 + \dots + a_n x_n$$

Scalars $(a_1, a_2, \dots, a_n)$ are called the coordinates of y with respect to the ordered basis $(x_1, x_2, \dots, x_n)$

<u>Example</u> Let $B_1 = (x_1, x_2)$ be a basis of $\mathbb{R}^2$. Let the coordinates of the vectors $y_1, y_2, y_3 \in \mathbb{R}^2$ w.r.t. B_1 be $[y_1]_{B_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $[y_2]_{B_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $[y_3]_{B_1} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

Find the representation of y_3 w.r.t. $B_2 = (y_1, y_2)$.

<u>Sol'n</u> Let $[y_3]_{B_2} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$. We can write

$y_1 = x_1 + x_2$, $y_2 = x_1$, $y_3 = 2x_1 + 3x_2$

$$\begin{aligned} 2x_1 + 3x_2 &= y_3 \\ &= \alpha y_1 + \beta y_2 \\ &= \alpha(x_1 + x_2) + \beta x_1 \\ &= (\alpha + \beta) x_1 + \alpha x_2 \end{aligned}$$

Hence,

$$\left.\begin{aligned} \alpha + \beta &= 2 \\ \alpha &= 3 \end{aligned}\right\} \Rightarrow \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$\Rightarrow \underbrace{\begin{bmatrix} \alpha \\ \beta \end{bmatrix}}_{\text{rep. w.r.t. } B_2} = \underbrace{\begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}}_{\text{change of coordinates matrix}} \underbrace{\begin{bmatrix} 2 \\ 3 \end{bmatrix}}_{\text{rep. w.r.t. } B_1}$$

$$\Rightarrow [y_3]_{B_2} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

<u>Exercise</u> Show that for a given ordered basis, each vector $x \in V$ has a unique representation.

Linear Transformation (Linear mapping)

Definition Let V and W be linear spaces over the same field F. A linear transformation T is a mapping $T: V \to W$ satisfying

$$T(a_1x_1 + a_2x_2) = a_1T(x_1) + a_2T(x_2) \qquad \text{for all } a_1, a_2 \in F \ \& \ x_1, x_2 \in V$$

Example Let $V = W$ be the set of polynomials with $\deg \leq n$ in s and $T = \frac{d}{ds}$.

T is a linear transformation because:

$$\left(\text{let } x_1 = \sum_{k=0}^{n} c_k s^k \ \& \ x_2 = \sum_{k=0}^{n} d_k s^k\right)$$

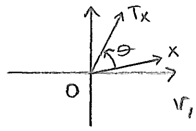
$$T(a_1x_1 + a_2x_2) = \frac{d}{ds}\left\{a_1 \sum c_k s^k + a_2 \sum d_k s^k\right\}$$

$$= a_1 \frac{d}{ds} \sum c_k s^k + a_2 \frac{d}{ds} \sum d_k s^k$$

$$= a_1 T x_1 + a_2 T x_2 \qquad \blacksquare$$

Example Let $V = W = \mathbb{R}^2$ and T be rotation around origin by θ radians.

Then T is a lin. trans.



Example Let $V = \{$ continuous functions of type $f: [0,1] \to \mathbb{R} \}$, $W = \mathbb{R}$ and T be defined as $Tf = \int_0^1 f(s)\,ds$. Then T is a lin. mapping.

— o —

Null Space & Range Space

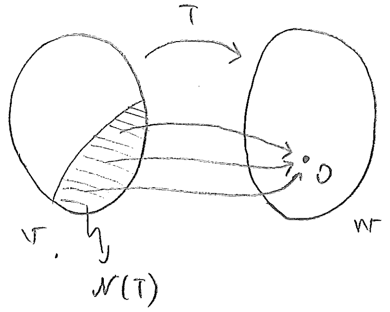
Definition Given a linear transformation $T: V \to W$, the null space of T, denoted by $\mathcal{N}(T)$, is the set of all $x \in V$ satisfying $Tx = 0_W$. That is,

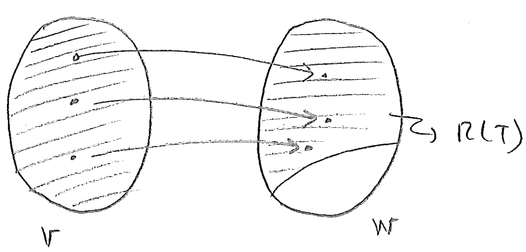
$$\mathcal{N}(T) = \{x \in V : Tx = 0\}.$$

The range space of T, denoted by $\mathcal{R}(T)$ is the set of all $w \in W$ satisfying $Tx = w$ for some $x \in V$. That is,

$$\mathcal{R}(T) = \{w \in W : w = Tx \text{ for some } x \in V\}.$$

That is,





Note that $\mathcal{N}(T) \subset V$ and $R(T) \subset W$.

Claim $\mathcal{N}(T)$ is a subspace of V.

Proof Let $v_1, v_2 \in \mathcal{N}(T)$ and $a \in F$ be arbitrary. We have

$$T(v_1+v_2) = Tv_1 + Tv_2$$
$$= 0 + 0$$
$$= 0 \implies v_1+v_2 \in \mathcal{N}(T) \quad (1)$$

Also, $T(av_1) = aTv_1$
$$= a\cdot 0$$
why?
$$= 0 \implies av_1 \in \mathcal{N}(T) \quad (2)$$

(1) & (2) imply $\mathcal{N}(T)$ is a subspace ▨

Because either $a=0$. Then $a\cdot 0 = 0\cdot 0 = 0$ since $0\cdot x = 0$ for all $x \in V$ (shown earlier).
Or $a \neq 0$. Then a^{-1} exists and

$$a\cdot 0 = a\cdot 0 + 0$$
$$= a\cdot 0 + 1\cdot 0$$
$$= a\cdot 0 + aa^{-1}0$$
$$= a(0 + a^{-1}0)$$
$$= a(a^{-1}0)$$
$$= aa^{-1}0$$
$$= 1\cdot 0$$
$$= 0$$ ▨

Claim $R(T)$ is a subspace of W.

Proof Exercise.

Def A lin. trans. $T: V \to W$ is said to be
1) onto if $R(T) = W$.
2) one-to-one if $v_1 \neq v_2 \implies Tv_1 \neq Tv_2$.

— o —

Theorem Let $T: V \to W$ be a linear transformation. Then the following are equivalent.

1) The mapping T is one-to-one (i.e., $v_1 \neq v_2 \implies Tv_1 \neq Tv_2$)
2) $\mathcal{N}(T) = \{0\}$

Proof $1 \Rightarrow 2$. Suppose not. That is, T is one-to-one but $\mathcal{N}(T) \neq \{0\}$. Then there exists $v \neq 0$ such that $Tv = 0$. We also have $T(0) = 0$. Hence $v \neq 0$ yet $Tv = T0$, which contradicts T is one-to-one.

2017

2 ⇒ 1. Suppose not. That is, $N(T) = \{0\}$ but T is not one-to-one. Then there exist $v_1 \neq v_2$ with $Tv_1 = Tv_2$. Let $v_3 = v_1 - v_2 \neq 0$. Then $Tv_3 = T(v_1 - v_2) = Tv_1 - Tv_2 = 0$. That is $v_3 \in N(T)$. But $v_3 \neq 0$. Hence $N(T) \neq \{0\}$. Contradiction. ∎

— o —

Definition (revisited) Given n-dimensional linear space V over (field) $\mathbb{R}$, let $B = (x_1, x_2, \dots, x_n)$ be an ordered basis for V. Given $v \in V$, let

$$v = \sum_{i=1}^{n} \alpha_i x_i \quad \text{where } \alpha_i \in \mathbb{R} \text{ for } i = 1, 2, \dots, n.$$

Recall that $\alpha_1, \alpha_2, \dots, \alpha_n$ are the coordinates of v w.r.t. the basis B. The column vector $\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{R}^n$ is called the representation of v w.r.t. B and denoted by $[v]_B$. That is, $[v]_B = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$.

Example Let $\mathbb{R}^{2\times 2}$ denote the linear space of real-valued 2×2 matrices where addition & scalar multiplic. are defined as

$$\begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix} + \begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix} = \begin{bmatrix} a_1+a_2 & b_1+b_2 \\ c_1+c_2 & d_1+d_2 \end{bmatrix}, \quad \alpha \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \alpha a & \alpha b \\ \alpha c & \alpha d \end{bmatrix}$$

Let $B = \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right)$ be our (ordered) basis.

Note that $\dim \mathbb{R}^{2\times 2} = 4$. Given $v = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbb{R}^{2\times 2}$ we can write

$$v = a\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + d\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \quad \text{Therefore, } [v]_B = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}.$$

Example Let V be the linear space of polynomials with $\deg \leq 2$.

Let $B = (p_1, p_2, p_3)$ with $p_1(t) = 1$, $p_2(t) = t$, and $p_3(t) = t^2$. Then given $p \in V$ with $p(t) = a + bt + ct^2$ we can write $[v]_B = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$.

Theorem Let V be an n-dimensional linear space over $\mathbb{R}$. Given two bases $B_1 = (v_1, v_2, \dots, v_n)$ and $B_2 = (w_1, w_2, \dots, w_n)$ of V, there exists $P \in \mathbb{R}^{n\times n}$ such that $[x]_{B_1} = P[x]_{B_2}$ $\forall x \in V$. In particular, the j^{th} column $\begin{bmatrix} P_{1j} \\ P_{2j} \\ \vdots \\ P_{nj} \end{bmatrix}$ of P satisfies $w_j = P_{1j}v_1 + P_{2j}v_2 + \dots + P_{nj}v_n$.

Proof Exercise.

Demonstration of the $n=3$ case Let $B_1 = (v_1, v_2, v_3)$ and $B_2 = (w_1, w_2, w_3)$.

We can find scalars $a, b, c, \dots$ such that

$$\left.\begin{aligned} w_1 &= a v_1 + b v_2 + c v_3 \\ w_2 &= d v_1 + e v_2 + f v_3 \\ w_3 &= g v_1 + h v_2 + k v_3 \end{aligned}\right\} (1)$$

Given $x \in V$, let $[x]_{B_1} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$ and $[x]_{B_2} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$. That is,

$$\begin{aligned} x &= \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 && (2) \\ &= \beta_1 w_1 + \beta_2 w_2 + \beta_3 w_3 && (3) \end{aligned}$$

By (1) & (3) we can write

$$\begin{aligned} x &= \beta_1(a v_1 + b v_2 + c v_3) + \beta_2(d v_1 + e v_2 + f v_3) + \beta_3(g v_1 + h v_2 + k v_3) \\ &= (\beta_1 a + \beta_2 d + \beta_3 g) v_1 + (\beta_1 b + \beta_2 e + \beta_3 h) v_2 + (\beta_1 c + \beta_2 f + \beta_3 k) v_3 \\ &= \underbrace{[a \ d \ g]\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}}_{\alpha_1} \cdot v_1 + \underbrace{[b \ e \ h]\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}}_{\alpha_2} \cdot v_2 + \underbrace{[c \ f \ k]\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}}_{\alpha_3} \cdot v_3 \end{aligned}$$

by (2)

Hence,

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \underbrace{\begin{bmatrix} a & d & g \\ b & e & h \\ c & f & k \end{bmatrix}}_{P} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \implies [x]_{B_1} = P[x]_{B_2} \quad \blacksquare$$

2017

<u>Claim</u> Matrix P is invertible. $([x]_{B_1} = P[x]_{B_2})$

proof Let $Q \in \mathbb{R}^{n\times n}$ satisfy $[x]_{B_2} = Q[x]_{B_1}$ for all x. Then

$$\begin{aligned}[x]_{B_2} &= Q[x]_{B_1} \\ &= Q\{P[x]_{B_2}\} \\ &= [QP][x]_{B_2} \qquad (1)\end{aligned}$$

Since (1) holds for all $[x]_{B_2}$ we have to have $QP = I$ (identity matrix).

Hence $P^{-1} = Q$ exists.

Matrix Representation of a Linear Transformation

<u>Example</u> Let $T: \underset{\text{dim}=2}{V} \to \underset{\text{dim}=4}{W}$ be a linear transformation

and $B = (v_1, v_2)$ & $C = (w_1, w_2, w_3, w_4)$

B: basis for $(V, \mathbb{R})$ C: basis for $(W, \mathbb{R})$ ($\mathbb{R}$: field)

We can find scalars $a, b, c, \ldots$ such that

$$\in W \quad \begin{aligned} Tv_1 &= aw_1 + bw_2 + cw_3 + dw_4 \\ Tv_2 &= ew_1 + fw_2 + gw_3 + hw_4 \end{aligned}$$

Now, given $v \in V$ and $w \in W$ with $w = Tv$ let

$$[v]_B = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \quad \text{and} \quad [w]_C = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$$

We can write

$$\beta_1 w_1 + \beta_2 w_2 + \beta_3 w_3 + \beta_4 w_4 = w$$
$$= Tv$$
$$= T(\alpha_1 v_1 + \alpha_2 v_2)$$
$$= \alpha_1 Tv_1 + \alpha_2 Tv_2$$
$$= \alpha_1 (a w_1 + b w_2 + \cdots) + \alpha_2 (e w_1 + f w_2 + \cdots)$$
$$= (\alpha_1 a + \alpha_2 e) w_1 + (\alpha_1 b + \alpha_2 f) w_2 + \cdots$$
$$= \underbrace{\begin{bmatrix} a & e \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}}_{\beta_1} \cdot w_1 + \underbrace{\begin{bmatrix} b & f \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}}_{\beta_2} \cdot w_2 + \cdots \quad \cdots$$

$$\Rightarrow \quad \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}}_{[Tv]_C} = \underbrace{\begin{bmatrix} a & e \\ b & f \\ c & g \\ d & h \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}}_{[v]_B} \quad \Rightarrow \quad [Tv]_C = A[v]_B$$

Our example can be generalized as:

<u>Theorem</u>. Given lin. trans. $T: V \to W$ and bases $B = (v_1, v_2, \ldots, v_n)$ for V & $C = (w_1, w_2, \ldots, w_m)$ for W, there exists $A \in \mathbb{R}^{m \times n}$ such that

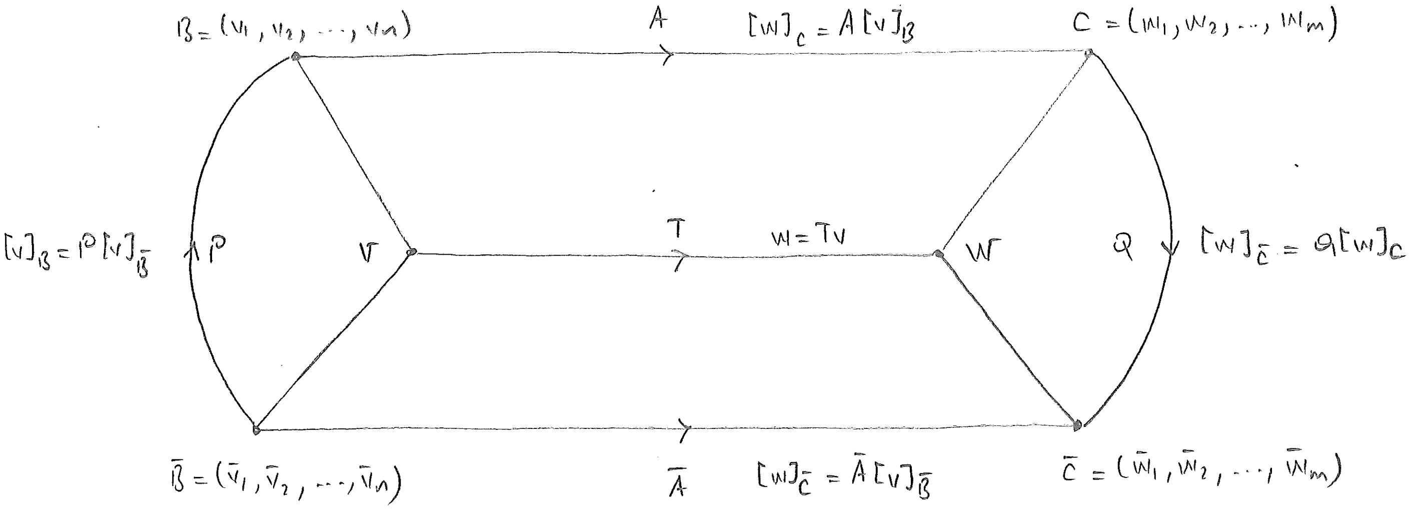
$$[Tx]_C = A[x]_B \quad \forall x \in V$$

where the j^{th} column of A : $\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}$ satisfy

$$Tv_j = a_{1j} w_1 + a_{2j} w_2 + \cdots + a_{mj} w_m$$

The matrix A is called the <u>matrix representation</u> of T w.r.t. bases B & C.

General Picture



$A \sim \bar{A}$ relation?

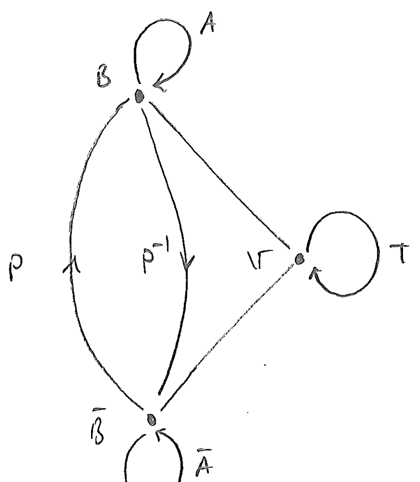
- $$\begin{aligned} [w]_{\bar{C}} &= Q[w]_C \\ &= QA[v]_B \\ &= \underbrace{QAP}_{\bar{A}}[v]_{\bar{B}} \end{aligned}$$

Hence,

$$\boxed{\underset{m\times n}{\bar{A}} = \underset{m\times m}{Q}\,\underset{m\times n}{A}\,\underset{n\times n}{P}}$$

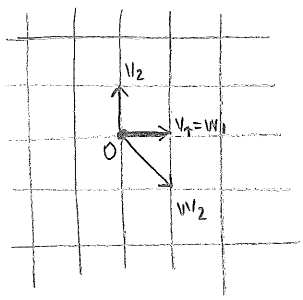
— o —

Special case If $W = V$, that is T is a map from V to itself, $T: V \to V$, then bases B & C ($\bar{B}$ & $\bar{C}$) can be taken to be the same and we have



$$\Rightarrow \boxed{\bar{A} = P^{-1}AP}$$

<u>Example</u> Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be $90°$ ccw rotation about the origin. (Note that T is linear.) Let $B_1 = (v_1, v_2)$ and $B_2 = (w_1, w_2)$ be two bases for $\mathbb{R}^2$ where v_1, v_2, w_1, w_2 are shown below.



a) Find the matrix representations $A_{11}, A_{12}, A_{21}, A_{22}$ such that $[Tx]_{B_i} = A_{ij} [x]_{B_j}$ for $(i,j) = (1,1), (1,2), (2,1), (2,2)$

b) Find coordinate change matrices P_{12}, P_{21} such that $[x]_{B_1} = P_{12} [x]_{B_2}$ and $[x]_{B_2} = P_{21} [x]_{B_1}$

c) Verify $A_{ij} = P_{ik} A_{kl} P_{lj}$ for all $i, j, k, l \in \{1, 2\}$ with $P_{ii} = I$.

<u>Sol'n</u> a) <u>$A_{11} = ?$</u>

Given $x \in \mathbb{R}^2$, let $[x]_{B_1} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$, i.e., $x = \alpha_1 v_1 + \alpha_2 v_2$.

$[Tx]_{B_1} = ?$

$$Tx = T(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 (Tv_1) + \alpha_2 (Tv_2) \quad (1)$$

$Tv_1, Tv_2 = ?$ (in terms of v_1, v_2)

$$\left.\begin{aligned} Tv_1 &= v_2 \\ Tv_2 &= -v_1 \end{aligned}\right\} \quad (2)$$

$$(1) \,\&\, (2) \Rightarrow Tx = \alpha_1 v_2 - \alpha_2 v_1 \Rightarrow [Tx]_{B_1} = \begin{bmatrix} -\alpha_2 \\ \alpha_1 \end{bmatrix}$$

$$[Tx]_{B_1} = A_{11} [x]_{B_1} \Rightarrow \begin{bmatrix} -\alpha_2 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} & A_{11} & \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$$

We can write

$$\begin{bmatrix} -\alpha_2 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \alpha_1 + \begin{bmatrix} -1 \\ 0 \end{bmatrix} \alpha_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$$

$$\boxed{\Rightarrow A_{11} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}}$$

Or, we could have directly applied the formula

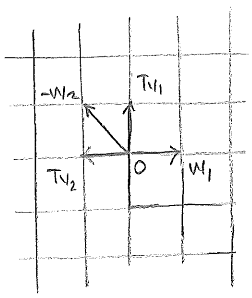
$Tv_1 = v_2 = 0 \cdot v_1 + 1 \cdot v_2 \Rightarrow$ first column of A_{11} is $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$Tv_2 = -v_1 = -1 \cdot v_1 + 0 \cdot v_2 \Rightarrow$ second column of A_{11} is $\begin{bmatrix} -1 \\ 0 \end{bmatrix}$

$\Rightarrow A_{11} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ as expected.

$A_{21} = ?$ $\qquad [Tx]_{B_2} = A_{21} [x]_{B_1}$

$Tv_1, Tv_2 = ?$ (in terms of w_1, w_2)



$Tv_1 = w_1 - w_2$

$Tv_2 = -w_1$

$\Rightarrow Tx = T(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 (Tv_1) + \alpha_2 (Tv_2) = \alpha_1 (w_1 - w_2) + \alpha_2 (-w_1)$

$= (\alpha_1 - \alpha_2) w_1 - \alpha_1 w_2$

$\Rightarrow [Tx]_{B_2} = \begin{bmatrix} \alpha_1 - \alpha_2 \\ -\alpha_1 \end{bmatrix}$

$[Tx]_{B_2} = A_{21} [x]_{B_1} \Rightarrow \begin{bmatrix} \alpha_1 - \alpha_2 \\ -\alpha_1 \end{bmatrix} = \begin{bmatrix} & A_{21} & \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$

$$\boxed{\Rightarrow A_{21} = \begin{bmatrix} 1 & -1 \\ -1 & 0 \end{bmatrix}}$$

Or, $\left. \begin{array}{l} Tv_1 = 1 \cdot w_1 - 1 \cdot w_2 \\ Tv_2 = -1 \cdot w_1 + 0 \cdot w_2 \end{array} \right\} \Rightarrow A_{21} = \begin{bmatrix} 1 & -1 \\ -1 & 0 \end{bmatrix}$

Exercise Find A_{12}, A_{22}.

b) $\underline{P_{21} = ?}$ Let $[x]_{\mathcal{B}_1} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$. $[x]_{\mathcal{B}_2} = ?$

$v_1, v_2 = ?$ (in terms of w_1, w_2)

$$\left.\begin{aligned} v_1 &= w_1 \\ v_2 &= w_1 - w_2 \end{aligned}\right\}$$

$\Rightarrow \quad x = \alpha_1 v_1 + \alpha_2 v_2 = \alpha_1 w_1 + \alpha_2 (w_1 - w_2) = (\alpha_1 + \alpha_2) w_1 - \alpha_2 w_2$

$\Rightarrow \quad [x]_{\mathcal{B}_2} = \begin{bmatrix} \alpha_1 + \alpha_2 \\ -\alpha_2 \end{bmatrix}$

$[x]_{\mathcal{B}_2} = P_{21} [x]_{\mathcal{B}_1} \quad \Rightarrow \quad \begin{bmatrix} \alpha_1 + \alpha_2 \\ -\alpha_2 \end{bmatrix} = \begin{bmatrix} & P_{21} & \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \quad \Rightarrow \quad \boxed{P_{21} = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}}$

Or, $v_1 = 1 \cdot w_1 + 0 \cdot w_2 \Rightarrow$ first column of P_{21} is $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$v_2 = 1 \cdot w_1 - 1 \cdot w_2 \Rightarrow$ second column of P_{21} is $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$

$\Rightarrow P_{21} = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}$ as expected.

$\underline{P_{12} = ?}$ $\quad P_{12} = P_{21}^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}$

c) Let's verify $A_{11} = P_{12} A_{21} P_{11} = P_{12} A_{21}$

$P_{12} A_{21} = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = A_{11}$ ✓

The rest of this part is left as an exercise.

$\underline{\text{Example}}$ Let $T: \mathbb{R}^{2\times 2} \to \mathbb{R}^{2\times 2}$ be such that $Tx = Sx + xS^T$ with $S = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

For the basis $\mathcal{B} = \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right)$ find the matrix representation A satisfying $[Tx]_{\mathcal{B}} = A[x]_{\mathcal{B}}$.

Sol'n Let $x = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be given. Then $[x]_B = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$.

Now, $Tx = Sx + xS^T = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} b+c & d-a \\ d-a & -b-c \end{bmatrix}$

$$= (b+c)\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + (d-a)\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + (d-a)\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + (-b-c)\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Therefore, $[Tx]_B = \begin{bmatrix} b+c \\ d-a \\ d-a \\ -b-c \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \end{bmatrix} a + \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} b + \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} c + \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} d$

$$= \underbrace{\begin{bmatrix} 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & -1 & 0 \end{bmatrix}}_{A}\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = A[x]_B$$

Verify that $A = \begin{bmatrix} [Tv_1]_B & [Tv_2]_B & [Tv_3]_B & [Tv_4]_B \end{bmatrix}$

~ first column ~ second col. ...

Example Let V be the linear space of poly. with real coefficients and with deg ≤ 3. Consider $T: V \to V$ with $Tf = f'$ (the derivative of f). For the basis $B = (1, s, s^2, s^3)$ find the matrix representation A of T satisfying $[Tf]_B = A[f]_B$.

Sol'n Let $f \in V$ with $f(s) = a + bs + cs^2 + ds^3$ be given. Then $[f]_B = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$.

Now, $Tf = \frac{d}{ds} f(s) = b + 2cs + 3ds^2$. Hence $[Tf]_B = \begin{bmatrix} b \\ 2c \\ 3d \\ 0 \end{bmatrix}$

We can write

$$[Tf]_B = \begin{bmatrix} b \\ 2c \\ 3d \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} a + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} b + \begin{bmatrix} 0 \\ 2 \\ 0 \\ 0 \end{bmatrix} c + \begin{bmatrix} 0 \\ 0 \\ 3 \\ 0 \end{bmatrix} d$$

$$= \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{A}\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = A[f]_B$$

__Theorem__ Let $T: V \to W$ be a linear transformation. Then

$$\dim R(T) + \dim N(T) = \dim V$$

__Proof__ Let $n = \dim V$ and the set $\{v_1, v_2, \dots, v_k\}$ $(k \le n)$ be a basis for $N(T)$. Now, complete this basis to a larger basis that spans the entire V.

$$sp\{\underbrace{v_1, v_2, \dots, v_k}_{\text{span } N(T)}, v_{k+1}, \dots, v_n\} = V$$

Let $y \in R(T)$ be an arbitrary vector. By definition there exists $x \in V$ such that $y = Tx$. Now, we can find scalars $\alpha_1, \alpha_2, \dots, \alpha_n$ such that $x = \sum_{i=1}^{n} \alpha_i v_i$.

Then,
$$y = Tx = T\left(\sum_{i=1}^{n} \alpha_i v_i\right) = T\left(\sum_{i=1}^{k} \alpha_i v_i + \sum_{i=k+1}^{n} \alpha_i v_i\right)$$

$$= \underbrace{\sum_{i=1}^{k} \alpha_i (Tv_i)}_{\substack{=0 \text{ since } v_i \in N(T) \\ \text{for } i=1,2,\dots,k}} + \sum_{i=k+1}^{n} \alpha_i (Tv_i) = \sum_{i=k+1}^{n} \alpha_i Tv_i$$

We obtained $y = \sum_{i=k+1}^{n} \alpha_i Tv_i$. Since $y \in R(T)$ was arbitrary, this means that we can express any vector in $R(T)$ as a lin. combination of $Tv_{k+1}, Tv_{k+2}, \dots, Tv_n$. Hence, $sp\{Tv_{k+1}, Tv_{k+2}, \dots, Tv_n\} = R(T)$.

— o —

__Claim__ The set $\{Tv_{k+1}, Tv_{k+2}, \dots, Tv_n\}$ is lin. independent. (Therefore it makes a basis for $R(T)$.)

__proof of claim__ suppose not. Then we can find scalars $\beta_{k+1}, \beta_{k+2}, \dots, \beta_n$ not all zero such that

$$\sum_{i=k+1}^{n} \beta_i Tv_i = 0 \Rightarrow T\left(\sum_{i=k+1}^{n} \beta_i v_i\right) = 0 \Rightarrow \sum_{i=k+1}^{n} \beta_i v_i \in N(T) \qquad (1)$$

(1) $\Rightarrow \quad -\sum_{i=k+1}^{n} \beta_i v_i \in N(T)$

$\Rightarrow$ we can find scalars $\beta_1, \beta_2, \dots, \beta_k$ such that

$$\sum_{i=1}^{k} \beta_i v_i = -\sum_{i=k+1}^{n} \beta_i v_i \qquad \text{since} \quad \operatorname{sp}\{v_1, v_2, \dots, v_k\} = N(T)$$

$$\Rightarrow \quad \sum_{i=1}^{k} \beta_i v_i + \sum_{i=k+1}^{n} \beta_i v_i = 0$$

$$\Rightarrow \quad \sum_{i=1}^{n} \beta_i v_i = 0 \quad \left(\text{and not all } \beta_i \text{ are zero}\right) \qquad (2)$$

(2) contradicts that $\{v_1, v_2, \dots, v_n\}$ is a basis. (end of proof of claim)

— o —

<u>proof of thm.</u> (cont'd) Now, we obtained

1) $\{Tv_{k+1}, Tv_{k+2}, \dots, Tv_n\}$ spans $R(T)$
2) $\{Tv_{k+1}, Tv_{k+2}, \dots, Tv_n\}$ is lin. ind.

Therefore $\{Tv_{k+1}, Tv_{k+2}, \dots, Tv_n\}$ is a basis for $R(T)$.

Finally, $\left.\begin{array}{l} \dim N(T) = k \\ \dim R(T) = n-k \end{array}\right\} \quad \dim N(T) + \dim R(T) = n.$ ▨

<u>Definition</u> (rank) Given a linear transformation $T: V \to W$, let A be the matrix representation of T w.r.t. some bases. Then the <u>rank</u> of the matrix A is defined as

$$\boxed{\operatorname{rank}(A) = \dim R(T) = \dim V - \dim N(T)}$$

<u>Fact</u> Let $A \in \mathbb{R}^{m \times n}$. Then

$\operatorname{rank}(A)$ = max. number of lin. ind. column vectors of A.
= max. number of lin. ind. row vectors of A.

Normed Linear Spaces

Consider a linear space $\mathcal{X}$ over the field $\mathbb{F}$, where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$. Let there be a function $x \mapsto \|x\|$ that assigns to each $x \in \mathcal{X}$, a nonnegative real number $\|x\| \geq 0$. Such function is called a norm if it satisfies the following properties.

(P1) $\|x+y\| \leq \|x\| + \|y\| \quad \forall x,y \in \mathcal{X}$ (triangle property)

(P2) $\|\alpha x\| = |\alpha| . \|x\| \quad \forall x \in \mathcal{X}$ & $\forall \alpha \in \mathbb{F}$

(P3) $\|x\| = 0 \iff x = 0$

The expression "$\|x\|$" is read "the norm of x" and the function $\|\cdot\| : \mathcal{X} \to \mathbb{R}_{\geq 0}$ is said to be a norm on $\mathcal{X}$. The triplet $(\mathcal{X}, \mathbb{F}, \|\cdot\|)$ is called normed space.

— o —

Remark An important use of the norm: to quantify "closeness". The "distance" between two vectors $x,y \in \mathcal{X}$ can be taken as $\|x-y\|$. Also, since $\|x\| = \|x-0\|$, norm $\|x\|$ can be interpreted as the distance of x to the origin 0.

— o —

Example Let $\mathcal{X} = \mathbb{R}^2$, $\mathbb{F} = \mathbb{R}$

$$x = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \in \mathcal{X}$$

i) $\|x\|_1 := |\alpha_1| + |\alpha_2|$. Is this a norm? (Check if it satisfies norm conditions.)

P1) $\|x+y\| = \left\| \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \right\| \qquad \left(y = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \right)$

$$= \left\| \begin{bmatrix} \alpha_1+\beta_1 \\ \alpha_2+\beta_2 \end{bmatrix} \right\|$$

$$= |\alpha_1+\beta_1| + |\alpha_2+\beta_2|$$

$$\leq |\alpha_1| + |\beta_1| + |\alpha_2| + |\beta_2|$$

$$= (|\alpha_1| + |\alpha_2|) + (|\beta_1| + |\beta_2|)$$

$$= \|x\| + \|y\| \quad \checkmark$$

P2) $\|\gamma x\| = \left\| \begin{bmatrix} \gamma\alpha_1 \\ \gamma\alpha_2 \end{bmatrix} \right\| = |\gamma\alpha_1| + |\gamma\alpha_2| = |\gamma| \left(|\alpha_1| + |\alpha_2| \right)$

$= |\gamma| \cdot \|x\|$ ✓

P3) $\|x\| = 0 \iff |\alpha_1| + |\alpha_2| = 0$

$\iff \alpha_1, \alpha_2 = 0$

$\iff x = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ ✓

ii) $\|x\| = \|x\|_2 := (\alpha_1^2 + \alpha_2^2)^{1/2}$ "2-norm" or "ℓ_2-norm" or "Euclidean norm", the norm we are most familiar with. (Exercise: show that it is indeed a norm.)

iii) $\|x\|_\infty := \max \{ |\alpha_1|, |\alpha_2| \}$ "ℓ_∞ norm". (Exercise: show that it is a norm.)

— o —

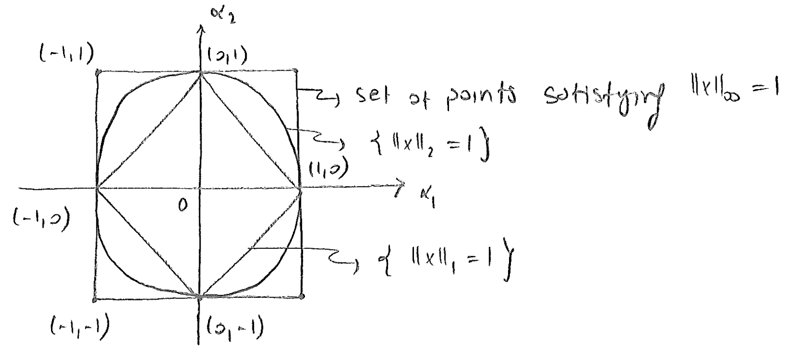
All these three norms can be generalized into what we call a "p-norm"

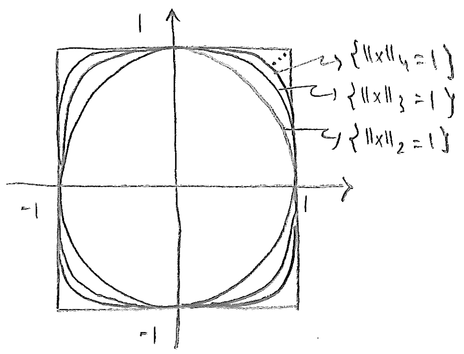
$$\|x\|_p := \left(|\alpha_1|^p + |\alpha_2|^p \right)^{1/p} \quad \text{for } 1 \leq p < \infty$$

Note that $\lim_{p \to \infty} \|x\|_p = \|x\|_\infty$

— o —

Visualization of these norms:





Example : Previous example can be generalized to $\mathbb{R}^n$.

$$X = \mathbb{R}^n, \quad F = \mathbb{R}$$

$$\|x\|_p = \left(|\alpha_1|^p + |\alpha_2|^p + \cdots + |\alpha_n|^p \right)^{1/p}$$

Definition The set $\{x \in X : \|x\| \leq 1\}$ is called a unit ball in $\mathbb{R}^n$.

Example (Norms on function spaces)

Let $X = \{ f \mid f : [0,1] \to \mathbb{R}, \ f \text{ continuous} \}$

We define the p-norm as $\|f\|_p := \left(\int_0^1 |f(t)|^p dt \right)^{1/p}$

Special cases $\|f\|_1 = \int_0^1 |f(t)|\, dt$

$\|f\|_2 = \left(\int_0^1 f^2(t) dt \right)^{1/2}$ $\Rightarrow$ Most important & popular norm "Root Mean Square" (RMS)

$\|f\|_\infty = \max\limits_{0 \leq t \leq 1} |f(t)|$

Matlab Exercise Let $A = \begin{bmatrix} -1 & -6 \\ 1 & 4 \end{bmatrix}$. Choose a nonzero arbitrary initial vector $x_0 \in \mathbb{R}^2$.

Perform the recursive computation $x_{k+1} = \dfrac{A x_k}{\|A x_k\|}$ for $k = 0, 1, 2, \ldots$

Observe the following :

1) $x_k \to \bar{x}$, where $\bar{x} \in \mathbb{R}^2$ is some fixed vector. (x_k converges to $\bar{x}$)

2) There exists $\lambda \in \mathbb{R}$ such that $A\bar{x} = \lambda \bar{x}$.

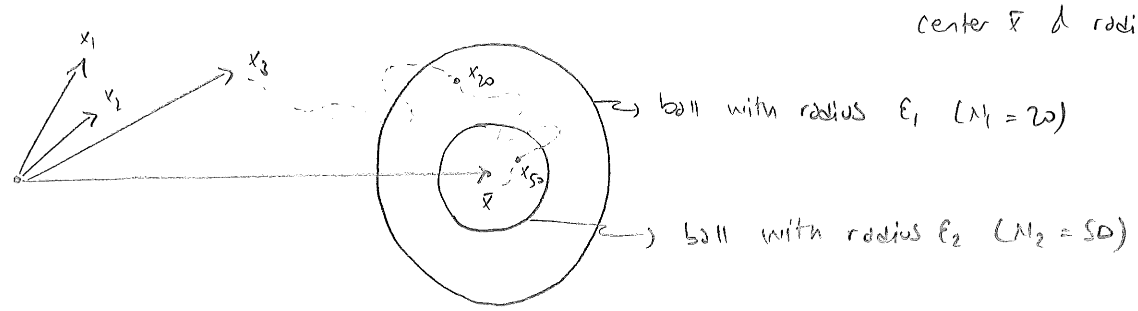
Repeat the experiment with a new initial vector x_0.

Convergence

Let $(X, F, \|\cdot\|)$ be a normed space. Let $(x_k)_{k=1}^{\infty}$ be a sequence of vectors in X, i.e., $x_k \in X$ $k = 1, 2, \dots$ The sequence is said to be convergent to the limit $\bar{x} \in X$ if $\|x_k - \bar{x}\| \to 0$ as $k \to \infty$.

Equivalently : $(x_k)_{k=1}^{\infty}$ is convergent to $\bar{x}$ if for any given $\varepsilon > 0$, there exists N (depending on ε) such that $k \geq N$ implies $\underbrace{\|x_k - \bar{x}\| \leq \varepsilon}$

"x_k within ball with center $\bar{x}$ & radius ε"

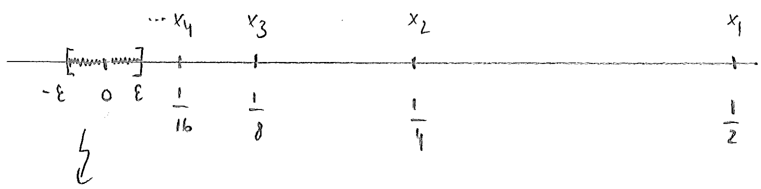


Usually one expects $\varepsilon \searrow$, $N \nearrow$.

A sequence that is not convergent is called divergent.

Example $X = \mathbb{R}$, $\|x\| = |x|$ (norm of $x \in \mathbb{R}$ is the absolute value)

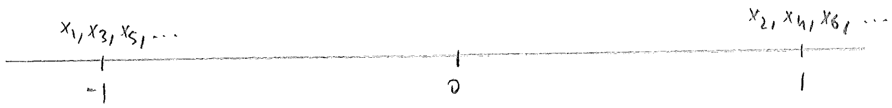
i) Consider the sequence $\left(\frac{1}{2^k}\right)_{k=1}^{\infty}$. The sequence is convergent to $\bar{x} = 0$.



ε-ball in $\mathbb{R}$
$[-\varepsilon, \varepsilon]$

Say $\varepsilon = \frac{1}{20}$. Then for $k \geq 5$ $\|x_k - 0\| \leq \varepsilon$.

ii) This time consider the sequence $\left([-1]^k\right)_{k=1}^{\infty}$



Let $\varepsilon = \frac{1}{2}$. Then we can never find N large enough so that $\|x_k - 0\| \le \varepsilon$. Therefore the sequence is not convergent to 0. (What if it is converging to some other point? See the note below.)

<u>Note</u> In most engineering applications we are interested in the convergence of an iterative algorithm. But the problem is that we usually do not know where. (otherwise, why the algorithm?)

One shortcoming of our convergence definition

$$\{\forall \varepsilon > 0 \quad \exists N \quad \text{such that} \quad \|x_k - \bar{x}\| \le \varepsilon \quad \text{for all} \quad k \ge N\}$$

is that it needs the limiting element $\bar{x} \in X$ to verify convergence. An element we may not know beforehand. An attempt to get rid of $\bar{x}$-dependence has resulted in the following concept.

<u>Definition</u> Let $(X, F, \|\cdot\|)$ be a normed space. A sequence $(x_k)_{k=1}^{\infty}$ is said to be a <u>Cauchy sequence</u> if $\forall \varepsilon > 0$, $\exists N$ such that $\|x_n - x_m\| \le \varepsilon$ for all $n, m \ge N$.

<u>Remark</u> : Every convergent sequence is Cauchy. The converse need not be true.

<u>Example</u> Let $(X, F, \|\cdot\|) = (\mathbb{Q}, \mathbb{Q}, |\cdot|)$ Consider the

$\mathbb{Q}$: rational numbers; $|\cdot|$: absolute value.

sequence $\left(1 + \sum_{n=1}^{k} \frac{1}{n!}\right)_{k=1}^{\infty}$. Then we have $x_1 = 2$, $x_2 = \frac{5}{2}$, $x_3 = \frac{16}{6}$, $x_4 = \frac{65}{24}, \ldots$

This sequence is Cauchy but in the limit $x_k \to e = 2.71828\ldots \notin \mathbb{Q}$

Hence, it is not convergent.

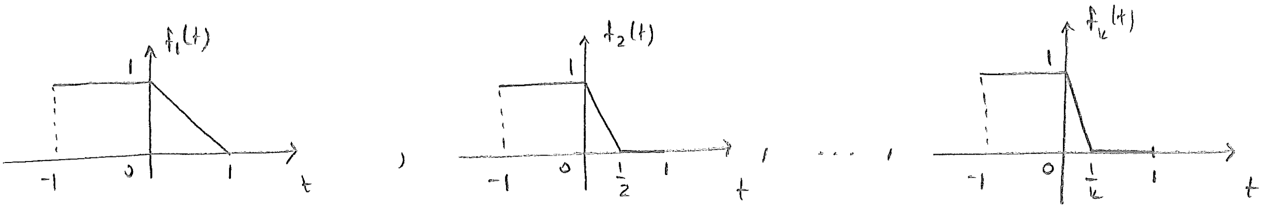
<u>Definition</u> A normed space is said to be <u>complete</u> if every Cauchy sequence is convergent. A complete normed space is called a <u>Banach space</u>.

<u>Example</u> [A normed space that is not complete.]

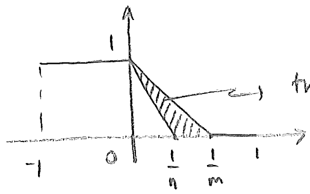
Let $X = \{ f \mid f : [-1, 1] \to \mathbb{R} , f \text{ continuous} \}$

Define $\|f\| := \int_{-1}^{1} |f(t)|\,dt$ as the norm.

Now, consider the sequence $(f_k)_{k=1}^{\infty}$ as follows



Then,



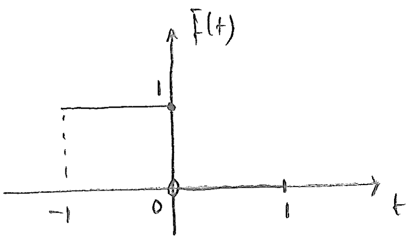
the area of the shaded triangle equals $\|f_n - f_m\|$

Hence $\|f_n - f_m\| = \frac{1}{2} \left| \frac{1}{n} - \frac{1}{m} \right|$. Note that $\|f_n - f_m\| \to 0$ as $N \to \infty$ & $n, m \geq N$.

Therefore the sequence is Cauchy.

<u>Question</u> : Is this sequence convergent ?

<u>Answer</u> : NO. Because the sequence "converges" to $\bar{f} \notin X$ below.



$\bar{f}$ discontinuous (at $t=0$)

$\to \bar{f} \notin X$.

2017

Matrix Norms

Example $\mathcal{X} = \mathbb{R}^{m \times n}$. Let $A = [a_{ij}]$. Then $\|A\| = \max_{i,j} |a_{ij}|$ is a norm.

P1)
$$\begin{aligned} \|A+B\| &= \max_{i,j} |a_{ij} + b_{ij}| \\ &\le \max_{i,j} (|a_{ij}| + |b_{ij}|) \\ &\le \max_{i,j} |a_{ij}| + \max_{i,j} |b_{ij}| \\ &= \|A\| + \|B\| \end{aligned}$$

P2)
$$\begin{aligned} \|\alpha A\| &= \max_{i,j} |\alpha a_{ij}| \\ &= |\alpha| \max_{i,j} |a_{ij}| \\ &= |\alpha| \cdot \|A\| \end{aligned}$$

P3)
$$\begin{aligned} \|A\| = 0 &\iff \max_{i,j} |a_{ij}| = 0 \\ &\iff a_{ij} = 0 \quad \forall i,j \\ &\iff A = 0 \end{aligned}$$

Example $\mathcal{X} = \mathbb{R}^{m \times n}$. Let $A = [a_{ij}]$.

Define $\|A\| = \max_{i} \sum_{j=1}^{n} |a_{ij}| = \max_{i} \{\text{absolute sum of } i^{th} \text{ row}\}$

e.g. $A = \begin{bmatrix} 2 & 1 & -1 \\ -2 & 3 & -1 \end{bmatrix} \begin{array}{l} \rightarrow |2| + |1| + |-1| = 4 \\ \rightarrow |-2| + |3| + |-1| = 6 \end{array} \Big\} \max\{4, 6\} = 6$

$\Rightarrow \|A\| = 6$

Exercise Show that this a norm.

2017

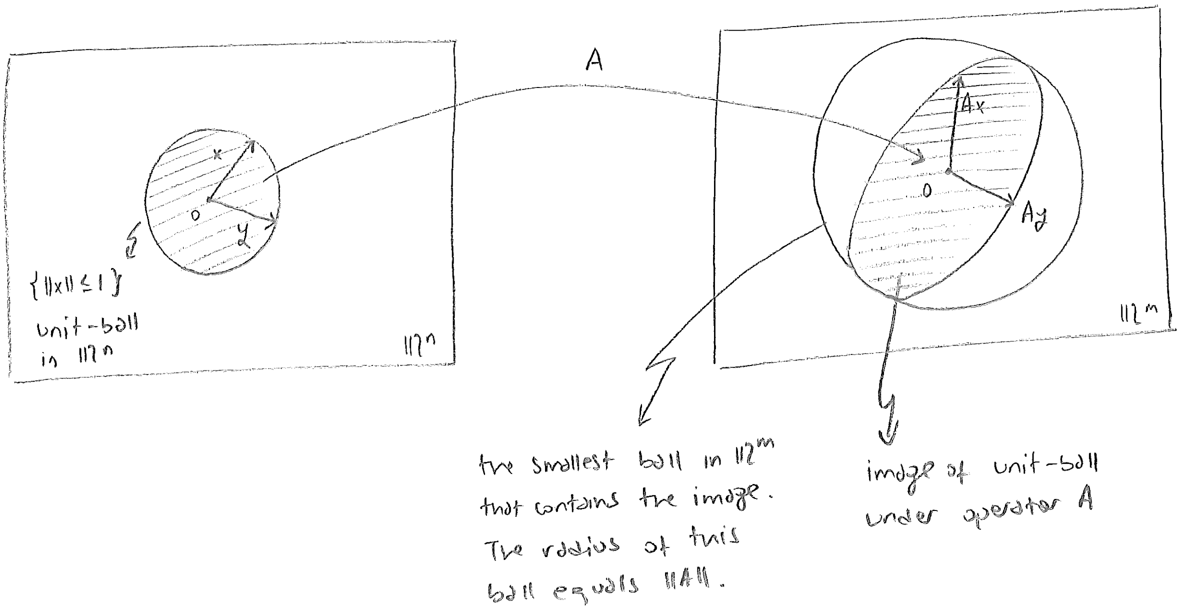
The induced matrix norm (matrix norm defined in terms of vector norms)

Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be an $m \times n$ matrix. Let $\|\cdot\|_{\mathbb{R}^n}$ & $\|\cdot\|_{\mathbb{R}^m}$ denote norms (vector norms) in $\mathbb{R}^n$ & $\mathbb{R}^m$. The induced norm is defined as

$$\boxed{\|A\| := \max_{0 \neq x \in \mathbb{R}^n} \frac{\|Ax\|_{\mathbb{R}^m}}{\|x\|_{\mathbb{R}^n}}}$$

Remark : An equivalent definition is $\boxed{\|A\| = \max_{\|x\|_{\mathbb{R}^n} = 1} \|Ax\|_{\mathbb{R}^m}}$ (why?)

Visualization



Remark
$$\|Ax\| = \frac{\|Ax\|}{\|x\|} \cdot \|x\|$$
$$\leq \left\{ \max_y \frac{\|Ay\|}{\|y\|} \right\} \cdot \|x\|$$
$$= \|A\| \cdot \|x\|$$

Therefore $\boxed{\|Ax\| \leq \|A\| \cdot \|x\|}$ for induced norms.

Remark There always exists x^* such that $\|Ax^*\| = \|A\| \cdot \|x^*\|$.

Example : Choose $\|\cdot\|_\infty$ as the norm in $\mathbb{R}^n$ and $\mathbb{R}^m$. Find a closed form expression of the induced norm $\|A\| = \max\limits_{\|x\|_\infty = 1} \|Ax\|_\infty$ for $A \in \mathbb{R}^{m\times n}$.

Sol'n $\quad \|A\| = \max\limits_{i=1,2,\dots,m} \sum\limits_{j=1}^{n} |a_{ij}|$ (max absolute row sum)

proof For $A=0$, the result is obvious. Suppose $A \neq 0$.

Let $A = \begin{bmatrix} - & r_1 & - \\ - & r_2 & - \\ & \vdots & \\ - & r_m & - \end{bmatrix}$ with $r_i = [a_{i1} \; a_{i2} \; \cdots \; a_{in}]_{1\times n}$ the i^{th} row of A

Note that $\|A\| = \max\limits_{\|x\|_\infty = 1} \|Ax\|_\infty = \max\limits_{\|x\|_\infty=1} \left\| \begin{bmatrix} r_1 x \\ r_2 x \\ \vdots \\ r_m x \end{bmatrix} \right\|_\infty = \max\limits_{\substack{\|x\|_\infty = 1 \\ i=1,2,\dots,m}} |r_i x|$

Let $z \in \mathbb{R}^n$ (with $\|z\|_\infty = 1$) and $k \in \{1,2,\dots,m\}$ be such that

$|r_k z| \geq |r_i x|$ for all $\|x\|_\infty = 1$ and $i \in \{1,2,\dots,m\}$.

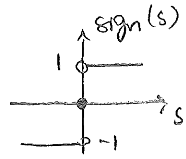
Hence, $\|A\| = |r_k z|$ and we can write

$$\begin{aligned}
\|A\| &= |a_{k1} z_1 + a_{k2} z_2 + \cdots + a_{kn} z_n| \\
&\leq |a_{k1}|\cdot|z_1| + |a_{k2}|\cdot|z_2| + \cdots + |a_{kn}|\cdot|z_n| \\
&\leq |a_{k1}| + |a_{k2}| + \cdots + |a_{kn}| \\
&\leq \max_{i=1,2,\dots,m} \sum_{j=1}^{n} |a_{ij}| \qquad (1)
\end{aligned}$$

because $|z_i| \leq 1$ (because $\|z\|_\infty = 1$)

Now, let $\ell \in \{1,2,\dots,m\}$ be such that $\max\limits_{i=1,2,\dots,m} \sum\limits_{j=1}^{n} |a_{ij}| = \sum\limits_{j=1}^{n} |a_{\ell j}|$.

Also, let $y \in \mathbb{R}^n$ be $y = \begin{bmatrix} \text{sign}(a_{\ell 1}) \\ \text{sign}(a_{\ell 2}) \\ \vdots \\ \text{sign}(a_{\ell n}) \end{bmatrix}$



Note that $\|y\|_\infty = 1$ since $A \neq 0$.

This time we write

$$\|A\| = \max_{\|x\|_\infty = 1} \|Ax\|_\infty$$

since $\|y\|_\infty = 1$

$$\geq \|Ay\|_\infty$$

$$Ay = \begin{bmatrix} r_1 y \\ r_2 y \\ \vdots \\ r_m y \end{bmatrix}$$

$$= \max_{i=1,2,\dots,m} |r_i y|$$

$$\geq |r_\ell y|$$

$$r_\ell = [a_{\ell 1} \cdots a_{\ell n}], \quad y = \begin{bmatrix} \text{sign}(a_{\ell 1}) \\ \vdots \\ \text{sign}(a_{\ell n}) \end{bmatrix}$$

$$= \left| \sum_{j=1}^{n} a_{\ell j} \cdot \text{sign}(a_{\ell j}) \right|$$

$$s \cdot \text{sign}(s) = |s|$$

$$= \left| \sum_{j=1}^{n} |a_{\ell j}| \right|$$

$$= \sum_{j=1}^{n} |a_{\ell j}|$$

$$= \max_{i=1,2,\dots,m} \sum_{j=1}^{n} |a_{ij}| \qquad (2)$$

Eqn. (1) & (2) imply $\|A\| = \max_{i=1,2,\dots,m} \sum_{j=1}^{n} |a_{ij}|$. ∎

— o —

<u>Example</u> Choose 2-norm in $\mathbb{R}^n$ & $\mathbb{R}^m$

$$\|A\| = \max_{\|x\|=1} \|Ax\|$$

$$= \max_{\|x\|=1} \sqrt{(Ax)^T Ax}$$

$$= \max_{\|x\|=1} \sqrt{x^T A^T A x}$$

Recall that

$$\|x\|_2 = \left\| \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \right\|_2 = \left(\sum_{i=1}^{n} \alpha_i^2 \right)^{1/2} = \sqrt{x^T x}$$

2017

Inner Product Space

An inner product space is a linear space with an additional structure called inner product.

Definition Let X be a linear space over the field F. (F is either $\mathbb{R}$ or $\mathbb{C}$)

An inner product is a map of the form

$$\langle \cdot , \cdot \rangle : X \times X \to F$$

that satisfies the following three conditions for all vectors $x, y, z \in X$ and all scalars $\alpha \in F$.

1) $\langle x, y \rangle = \overline{\langle y, x \rangle}$ (conjugate symmetry)

2a) $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$

2b) $\langle x+y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$

$\Big\}$ (linearity in the first argument)

3) $\langle x, x \rangle \geq 0$ with equality only for $x = 0$ (positive-definiteness)

— o —

Question How about $\langle x, y+z \rangle = ?$ & $\langle x, \alpha y \rangle = ?$

Answer

$$\begin{aligned} \langle x, y+z \rangle &= \overline{\langle y+z, x \rangle} \\ &= \overline{\langle y, x \rangle + \langle z, x \rangle} \\ &= \overline{\langle y, x \rangle} + \overline{\langle z, x \rangle} \\ &= \langle x, y \rangle + \langle x, z \rangle \end{aligned}$$

&

$$\begin{aligned} \langle x, \alpha y \rangle &= \overline{\langle \alpha y, x \rangle} \\ &= \overline{\alpha \langle y, x \rangle} \\ &= \bar{\alpha} \cdot \overline{\langle y, x \rangle} \\ &= \bar{\alpha} \cdot \langle x, y \rangle \end{aligned}$$

Example $\mathcal{X} = \mathbb{C}^n$ with $x = [x_1 \ x_2 \cdots x_n]^T \in \mathcal{X}$

Then $\langle x, y\rangle := \sum_{i=1}^{n} x_i \bar{y}_i$ is an inner product.

Let us check whether the conditions (1, 2a, 2b, 3) are satisfied:

1) $\langle y, x\rangle = \sum_{i=1}^{n} y_i \bar{x}_i = \sum \overline{x_i \bar{y}_i} = \overline{\sum x_i \bar{y}_i} = \overline{\langle x, y\rangle}$ ✓

2a) $\langle \alpha x, y\rangle = \sum_{i=1}^{n} \alpha x_i \bar{y}_i = \alpha \sum x_i \bar{y}_i = \alpha \langle x, y\rangle$ ✓

2b) $\langle x+y, z\rangle = \sum (x_i + y_i)\bar{z}_i = \sum (x_i \bar{z}_i + y_i \bar{z}_i) = \sum x_i \bar{z}_i + \sum y_i \bar{z}_i = \langle x, z\rangle + \langle y, z\rangle$ ✓

3) $\langle x, x\rangle = \sum x_i \bar{x}_i = \sum |x_i|^2 \geq 0$ & $\sum |x_i|^2 = 0 \Leftrightarrow x = 0$ ✓

Example Let $\mathcal{X} = \{ f \mid f : [0,1] \to \mathbb{C} ,\ f \text{ continuous} \}$ over the field $\mathbb{C}$.

A standard inner product in $\mathcal{X}$ is:

$$\langle f, g\rangle := \int_0^1 f(t)\, \overline{g(t)}\, dt$$

Theorem The Cauchy-Schwarz inequality

$$\boxed{|\langle x, y\rangle|^2 \leq \langle x, x\rangle \langle y, y\rangle} \qquad \forall x, y$$

Proof If $y = 0$ then both $\langle x, y\rangle$ & $\langle y, y\rangle = 0$ and the inequality holds.

Now, suppose $y \neq 0$. Let $\lambda = \dfrac{\langle x, y\rangle}{\langle y, y\rangle}$. $\left(\text{Note that } \bar{\lambda} = \dfrac{\langle y, x\rangle}{\langle y, y\rangle}\right)$

We can write:

$$\begin{aligned} 0 &\leq \langle x - \lambda y, x - \lambda y\rangle \\ &= \langle x, x\rangle + \langle -\lambda y, x\rangle + \langle x, -\lambda y\rangle + \langle -\lambda y, -\lambda y\rangle \\ &= \langle x, x\rangle - \lambda\langle y, x\rangle - \bar{\lambda}\langle x, y\rangle + \lambda\bar{\lambda}\langle y, y\rangle \end{aligned}$$

 2017

We then proceed as

$$0 \le \langle x,x\rangle - \frac{\langle x,y\rangle}{\langle y,y\rangle}\langle y,x\rangle - \frac{\langle y,x\rangle}{\langle y,y\rangle}\langle x,y\rangle + \frac{\langle x,y\rangle\langle y,x\rangle}{\langle y,y\rangle^2}\langle y,y\rangle$$

$$= \langle x,x\rangle - \frac{|\langle x,y\rangle|^2}{\langle y,y\rangle} - \frac{|\langle x,y\rangle|^2}{\langle y,y\rangle} + \frac{|\langle x,y\rangle|^2}{\langle y,y\rangle}$$

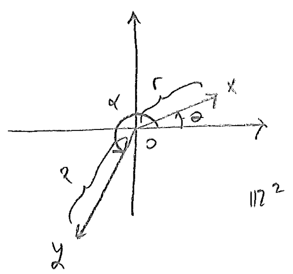
$$= \langle x,x\rangle - \frac{|\langle x,y\rangle|^2}{\langle y,y\rangle} \qquad (1)$$

(1) $\Rightarrow$ $0 \le \langle x,x\rangle\langle y,y\rangle - |\langle x,y\rangle|^2$ ▨

Example (Geometric demonstration of Cauchy-Schwarz ineq. in $\mathbb{R}^2$)

Let $X = \mathbb{R}^2$ with $\langle x,y\rangle = y^T x$

Given $x,y \in \mathbb{R}^2$ we can write $x = r\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix}$ & $y = \rho\begin{bmatrix}\cos\alpha\\ \sin\alpha\end{bmatrix}$



Then $\langle x,y\rangle^2 = (y^T x)^2 = \left(r\rho\,[\cos\alpha \;\; \sin\alpha]\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix}\right)^2$

$$= r^2\rho^2(\cos\alpha\cos\theta + \sin\alpha\sin\theta)^2$$

$$= r^2\rho^2\cos^2(\alpha-\theta)$$

$$\le r^2\rho^2$$

$$= r^2(\cos^2\theta + \sin^2\theta)\,\rho^2(\cos^2\alpha+\sin^2\alpha)$$

$$= \langle x,x\rangle \cdot \langle y,y\rangle$$

Remark Observe that for $X = \mathbb{R}^n$ with $\langle x,y \rangle = y^T x$ we can write

$$\langle x,x \rangle = [x_1\ x_2 \cdots x_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1^2 + x_2^2 + \cdots + x_n^2 = \|x\|_2^2$$

That is, $\sqrt{\langle x,x \rangle}$ is a norm. This is true for any inner product:

Theorem $\sqrt{\langle x,x \rangle} = \|x\|$ is a norm.

Proof We need to show that the norm properties are satisfied.

P1) $\|x+y\| \leq \|x\| + \|y\|$?

$$\begin{aligned}
\|x+y\|^2 &= \langle x+y, x+y \rangle \\
&= \langle x,x \rangle + \langle x,y \rangle + \langle y,x \rangle + \langle y,y \rangle \\
&= \langle x,x \rangle + \langle y,y \rangle + \langle x,y \rangle + \overline{\langle x,y \rangle} \\
&= \langle x,x \rangle + \langle y,y \rangle + 2\,\mathrm{Re}\{\langle x,y \rangle\} \\
&\leq \langle x,x \rangle + \langle y,y \rangle + 2\,|\langle x,y \rangle| \\
&\leq \langle x,x \rangle + \langle y,y \rangle + 2\,\langle x,x \rangle^{1/2} \langle y,y \rangle^{1/2} \quad \text{Cauchy-Schwarz ineq.} \\
&= \left(\langle x,x \rangle^{1/2} + \langle y,y \rangle^{1/2} \right)^2 \\
&= (\|x\| + \|y\|)^2 \qquad (1)
\end{aligned}$$

$(1) \Rightarrow \|x+y\| \leq \|x\| + \|y\|$ ✓

P2) $\|\alpha x\| = |\alpha| \cdot \|x\|$?

$$\begin{aligned}
\|\alpha x\| &= \langle \alpha x, \alpha x \rangle^{1/2} \\
&= (\alpha \bar{\alpha} \langle x,x \rangle)^{1/2} \\
&= (|\alpha|^2 \langle x,x \rangle)^{1/2} \\
&= |\alpha| \cdot \langle x,x \rangle^{1/2} \\
&= |\alpha| \cdot \|x\| \quad ✓
\end{aligned}$$

P3) $\|x\| = 0 \Leftrightarrow x = 0$?

This directly comes from the definition of inner product (the pos.-def. property) ✓

2 Kasım 2017

Remark Every inner product space is a normed space. Converse is not true.

Counterexample The one norm $\|\cdot\|_1$ in $\mathbb{R}^2$ cannot be generated by an inner product.

Because Suppose not. Then $\left\langle \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\rangle = (|x_1| + |x_2|)^2$. We can write

$$\langle x-y, x-y \rangle = \langle x,x \rangle + \langle x,-y \rangle + \langle -y,x \rangle + \langle -y,-y \rangle$$

$$= \langle x,x \rangle + \langle y,y \rangle - 2\langle x,y \rangle$$

& $\langle x+y, x+y \rangle = \langle x,x \rangle + \langle y,y \rangle + 2\langle x,y \rangle$.

Then

$$\langle x-y, x-y \rangle + \langle x+y, x+y \rangle = 2(\langle x,x \rangle + \langle y,y \rangle)$$

$$\Rightarrow \|x-y\|_1^2 + \|x+y\|_1^2 = 2(\|x\|_1^2 + \|y\|_1^2)$$

Choose $x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ & $y = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Then

$$\underbrace{\|x-y\|_1^2}_{4} + \underbrace{\|x+y\|_1^2}_{4} = 2(\underbrace{\|x\|_1^2}_{1} + \underbrace{\|y\|_1^2}_{1}) \Rightarrow 8 = 4 \Rightarrow \text{contradiction.} \quad \boxtimes$$

— o —

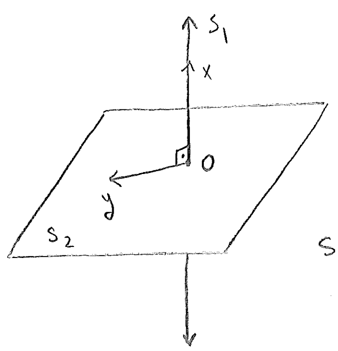
Definition An inner product space that is complete with respect to the norm induced by the inner product is called a <u>Hilbert Space</u>.

— o —

Orthogonality

Definition Two vectors $x, y \in X$ are said to be orthogonal if $\langle x,y \rangle = 0$. More generally, two subsets $S_1, S_2 \subset X$ are said to be orthogonal if $\langle x,y \rangle = 0$ $\forall x \in S_1$ & $\forall y \in S_2$.

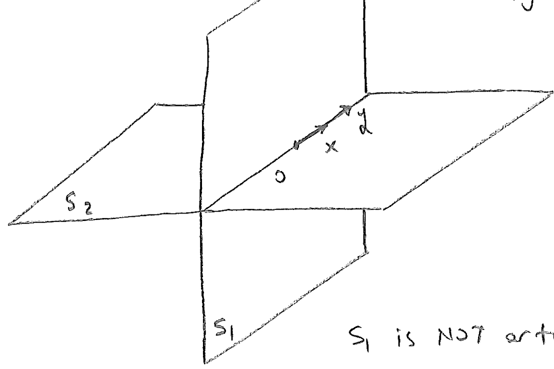
Example Let $X = \mathbb{R}^3$ & $\langle x,y \rangle = x^T y$



$\langle x,y \rangle = 0$

S_1 is orthogonal to S_2

$\langle x,y \rangle \neq 0$



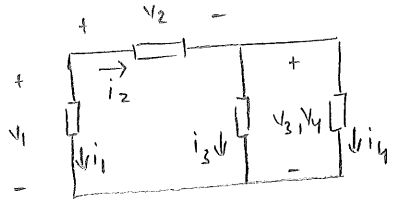
S_1 is NOT orthogonal to S_2!

Example Let $A \in \mathbb{R}^{m\times n}$. Then $N(A)$ and $R(A^T)$ are orthogonal.

Because Let $x \in N(A)$ & $y \in R(A^T)$. $y \in R(A^T) \Rightarrow \exists z \in \mathbb{R}^m$ s.t. $y = A^T z$

Then $\langle x,y\rangle = x^T y = x^T(A^T z) = (Ax)^T z = 0$ (since $Ax=0$)

Example Consider the circuit



$v = \begin{bmatrix} v_1\\ v_2\\ v_3\\ v_4\end{bmatrix} \in \mathbb{R}^4$ the voltage vector, $i = \begin{bmatrix} i_1\\ i_2\\ i_3\\ i_4\end{bmatrix} \in \mathbb{R}^4$ the current vector

$$\text{KCL} \Rightarrow \left.\begin{array}{l} i_1+i_2=0\\ \& \ -i_2+i_3+i_4=0\end{array}\right\} \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0\\ 0 & -1 & 1 & 1\end{bmatrix}}_{A:\ \text{(reduced) incidence matrix}} \begin{bmatrix} i_1\\ i_2\\ i_3\\ i_4\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix} \Rightarrow Ai=0 \Rightarrow i \in N(A)$$

$$\text{KVL} \Rightarrow \underbrace{\begin{bmatrix} 1 & 0\\ 1 & -1\\ 0 & 1\\ 0 & 1\end{bmatrix}}_{A^T} \underbrace{\begin{bmatrix} v_1\\ v_3\end{bmatrix}}_{e} = \begin{bmatrix} v_1\\ v_2\\ v_3\\ v_4\end{bmatrix} \Rightarrow A^T e = v \Rightarrow v \in R(A^T)$$

Hence $\langle i,v\rangle = 0$. This is true for any circuit: "The voltage vector is always orthogonal to the current vector." (Tellegen's Thm.)

Example The total energy $E = \sum\limits_{k \text{ over capacitors}} \frac{1}{2} C_k v_k^2 + \sum\limits_{k \text{ over inductors}} \frac{1}{2} L_k i_k^2$ of a passive ($R_k, L_k, C_k > 0 \ \forall k$) RLC network never increases.

$$\dot{E} = \sum_{k\in \text{Cap.}} C_k \dot{v}_k v_k + \sum_{k \in \text{Ind.}} L_k \dot{i}_k i_k \qquad \Big)\ C_k\dot{v}_k = i_k,\ L_k \dot{i}_k = v_k$$

$$= \sum_{k\in\text{Cap.}} i_k v_k + \sum_{k\in\text{Ind.}} v_k i_k \qquad \Big)\ \sum_{k\in R,L,C} i_k v_k = 0 \ \text{(Tellegen's Thm.)}$$

$$= -\sum_{k \in \text{Res.}} i_k v_k \qquad \Big)\ v_k = R_k i_k$$

$$= -\sum_{k\in\text{Res.}} R_k i_k^2 \qquad \Rightarrow \dot{E} \le 0. \ \blacksquare$$

Example $\mathcal{X} = \mathbb{R}^n$, $\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i = x^T y$ where $x = [x_1 \; x_2 \cdots x_n]^T$

Consider the canonical basis set $\mathcal{B} = \{e_1, e_2, \ldots, e_n\}$

where $e_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow$ ith entry. Then we have $\langle e_i, e_j \rangle = \underbrace{\delta_{ij}}_{\text{"kronecker delta"}} := \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}$

Therefore any two (distinct) vectors from the set $\mathcal{B}$ are orthogonal.

Example $\mathcal{X} = \{ f \mid f : [-\pi, \pi] \to \mathbb{C}, \; f \text{ continuous} \}$

Let $\langle f, g \rangle = \int_{-\pi}^{\pi} f(t)\overline{g(t)}\,dt$ be the inner product.

Now, consider the set of vectors $\{e^{jnt}\}_{n=-\infty}^{\infty} \subset \mathcal{X}$

$$\Rightarrow \langle e^{jnt}, e^{jmt} \rangle = \int_{-\pi}^{\pi} e^{jnt}\overline{e^{jmt}}\,dt = \int_{-\pi}^{\pi} e^{jnt} e^{-jmt}\,dt = \int_{-\pi}^{\pi} \underbrace{e^{j(n-m)t}}_{\text{sinusoidal with period } 2\pi \text{ for all } n \neq m} dt$$

Therefore $\langle e^{jnt}, e^{jmt} \rangle = 2\pi \delta_{nm}$ and any two vectors from the set $\{e^{jnt}\}_{n=-\infty}^{\infty}$ are orthogonal.

Likewise, consider the set $\{\sin(nt)\}_{n=1}^{\infty} \subset \mathcal{X}$

Then we can write

$$\langle \sin(nt), \sin(mt) \rangle = \int_{-\pi}^{\pi} \sin(nt)\sin(mt)\,dt$$

$\sin\alpha \sin\beta = \frac{1}{2}\{\cos(\alpha-\beta) - \cos(\alpha+\beta)\}$

$$= \int_{-\pi}^{\pi} \left\{ \tfrac{1}{2}\cos(n-m)t - \tfrac{1}{2}\cos(n+m)t \right\} dt$$

$$= \begin{cases} \pi & \text{if } n = m \\ 0 & \text{if } n \neq m \end{cases}$$

Hence, $\langle \sin(nt), \sin(mt) \rangle = \pi \delta_{nm}$.

<u>Question</u> Let V be an inner product space and $S_1 = \{v_1, v_2, \dots, v_n\}$ with $v_i \in V$ be a linearly independent set. Can we obtain another set $S_2 = \{w_1, w_2, \dots, w_n\}$ such that

1) span S_2 = span S_1 and
2) the vectors in S_2 are pairwise orthogonal. That is, $\langle w_i, w_j \rangle = 0 \quad \forall\, i \neq j$?

The answer is:

<u>Gram-Schmidt Orthogonalization</u>

<u>The G.-S. Algorithm</u>: Given a lin. ind. set $\{v_1, v_2, \dots, v_n\}$ obtain (construct) $\{w_1, w_2, \dots, w_n\}$ as follows:

$$w_1 := v_1$$

$$w_2 := v_2 - \frac{\langle v_2, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1$$

$$w_3 := v_3 - \frac{\langle v_3, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 - \frac{\langle v_3, w_2 \rangle}{\langle w_2, w_2 \rangle} w_2$$

$$\vdots$$

$$w_n := v_n - \sum_{i=1}^{n-1} \frac{\langle v_n, w_i \rangle}{\langle w_i, w_i \rangle} w_i$$

— o —

<u>Claim 1</u> $\langle w_i, w_j \rangle = 0 \quad \forall\, i \neq j$. That is, $\{w_1, w_2, \dots, w_n\}$ is a pairwise orthogonal set.

<u>Proof</u> (By induction) Suppose for some $k < n$, the set $\{w_1, w_2, \dots, w_k\}$ is pairwise orthogonal. Then for all $j \in \{1, 2, \dots, k\}$ we can write

2017

$$\langle w_{k+1}, w_j \rangle = \langle v_{k+1} - \sum_{i=1}^{k} \frac{\langle v_{k+1}, w_i \rangle}{\langle w_i, w_i \rangle} w_i , w_j \rangle$$

$$= \langle v_{k+1}, w_j \rangle - \left\langle \sum_{i=1}^{k} \frac{\langle v_{k+1}, w_i \rangle}{\langle w_i, w_i \rangle} w_i , w_j \right\rangle$$

$$= \langle v_{k+1}, w_j \rangle - \sum_{i=1}^{k} \frac{\langle v_{k+1}, w_i \rangle}{\langle w_i, w_i \rangle} \underbrace{\langle w_i, w_j \rangle}_{=0 \ \forall i \neq j}$$

$$= \langle v_{k+1}, w_j \rangle - \frac{\langle v_{k+1}, w_j \rangle}{\langle w_j, w_j \rangle} \langle w_j, w_j \rangle$$

$$= \langle v_{k+1}, w_j \rangle - \langle v_{k+1}, w_j \rangle$$

$$= 0$$

Therefore $\{w_1, w_2, \dots, w_{k+1}\}$ is pairwise orthogonal. Since $\{w_1\}$ is pairwise orthogonal (?), the result follows by induction. ☒

<u>Claim 2</u> $\quad \text{span}\{w_1, w_2, \dots, w_n\} = \text{span}\{v_1, v_2, \dots, v_n\}$

<u>Proof</u> (by induction) Suppose $\text{span}\{w_1, w_2, \dots, w_k\} = \text{span}\{v_1, v_2, \dots, v_k\}$ for some $k < n$.

Let $x \in \text{span}\{v_1, v_2, \dots, v_{k+1}\}$. Then we can find $\alpha_1, \alpha_2, \dots, \alpha_{k+1}$ such that

$$x = \sum_{i=1}^{k+1} \alpha_i v_i = \sum_{i=1}^{k} \alpha_i v_i + \alpha_{k+1} \left\{ w_{k+1} + \sum_{i=1}^{k} \frac{\langle v_{k+1}, w_i \rangle}{\langle w_i, w_i \rangle} w_i \right\}$$

$$= \underbrace{\underbrace{\sum_{i=1}^{k} \alpha_i v_i}_{\in \text{sp}\{w_1, \dots, w_k\}} + \alpha_{k+1} w_{k+1} + \alpha_{k+1} \sum_{i=1}^{k} \frac{\langle v_{k+1}, w_i \rangle}{\langle w_i, w_i \rangle} w_i}_{\in \text{sp}\{w_1, \dots, w_{k+1}\}}$$

Hence $x \in \text{sp}\{w_1, \dots, w_{k+1}\}$. Therefore

$$\text{span}\{v_1, v_2, \dots, v_{k+1}\} \subset \text{span}\{w_1, w_2, \dots, w_{k+1}\} \qquad (1)$$

Let now $y \in \operatorname{span}\{w_1, w_2, \ldots, w_{k+1}\}$. Then we can find $\beta_1, \beta_2, \ldots, \beta_{k+1}$ such that

$$y = \sum_{i=1}^{k+1} \beta_i w_i = \sum_{i=1}^{k} \beta_i w_i + \beta_{k+1} \left\{ v_{k+1} - \sum_{i=1}^{k} \frac{\langle v_{k+1}, w_i \rangle}{\langle w_i, w_i \rangle} w_i \right\}$$

$$= \underbrace{\underbrace{\sum_{i=1}^{k} \beta_i w_i - \beta_{k+1} \sum_{i=1}^{k} \frac{\langle v_{k+1}, w_i \rangle}{\langle w_i, w_i \rangle} w_i}_{\in \operatorname{sp}\{v_1, \ldots, v_k\}} + \beta_{k+1} v_{k+1}}_{\in \operatorname{sp}\{v_1, \ldots, v_{k+1}\}}$$

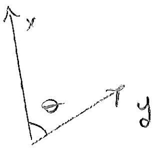
Hence $y \in \operatorname{sp}\{v_1, \ldots, v_{k+1}\}$. Therefore

$$\operatorname{span}\{w_1, w_2, \ldots, w_{k+1}\} \subset \operatorname{span}\{v_1, v_2, \ldots, v_{k+1}\} \qquad (2)$$

Now, (1) & (2) imply $\operatorname{span}\{w_1, \ldots, w_{k+1}\} = \operatorname{span}\{v_1, \ldots, v_{k+1}\}$. The result follows by induction since $\operatorname{span}\{w_1\} = \operatorname{span}\{v_1\}$. ▨

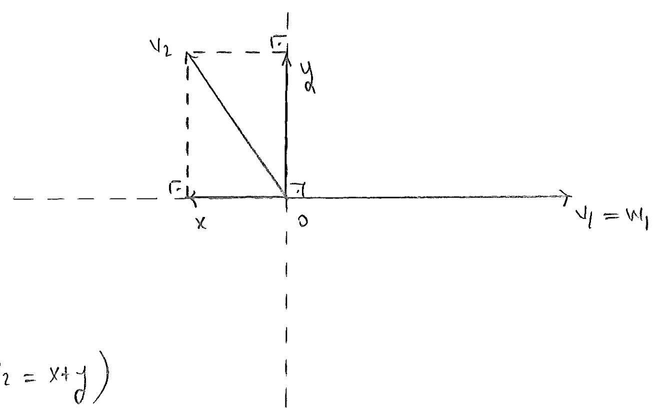
Example (Visualization of Gram-Schmidt process in $\mathbb{R}^2$.

Recall: for $x, y \in \mathbb{R}^2$



$\langle x, y \rangle = x^T y = \|x\| \cdot \|y\| \cdot \cos\theta$.)

Given $\{v_1, v_2\}$



$(v_2 = x + y)$

Note that y is orthogonal to w_1.

Hence we can use y as our w_2.

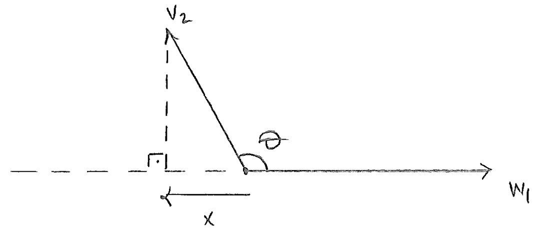
Question: How to express y in terms of w_1 & v_2?

Answer: $y = v_2 - x$

2017

$y = v_2 - x$

$= v_2 - \underbrace{\|x\|}_{\text{length of } x} \cdot \underbrace{\left(\frac{x}{\|x\|}\right)}_{\text{unit vector along } x}$ (1)



Note that

$\|x\| = \|v_2\| \cdot \cos(180° - \theta)$ (2)

$\frac{x}{\|x\|} = -\frac{w_1}{\|w_1\|}$ (3)

Combining (1), (2), & (3) yields

$y = v_2 - \|v_2\| \cdot \cos(180° - \theta) \cdot \left(-\frac{w_1}{\|w_1\|}\right)$

$\cos(180° - \theta) = -\cos\theta$

$= v_2 - \frac{\|v_2\| \cdot \|w_1\| \cdot \cos\theta}{\|w_1\|^2} \cdot w_1$

$= v_2 - \frac{\langle v_2, w_1 \rangle}{\langle w_1, w_1 \rangle} \cdot w_1$

$= w_2$

<u>Example</u> Let $V = \{ f \mid f : [-1,1] \to \mathbb{R},\ f \text{ continuous} \}$

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

Let $S_1 = \{ f_1, f_2, f_3 \}$ where $f_1(t) = 1$, $f_2(t) = t$, $f_3(t) = t^2$ $\forall t \in [-1,1]$

Obtain $S_2 = \{ g_1, g_2, g_3 \}$ (by Gram-Schmidt Alg.) such that

→ $\operatorname{Span} S_2 = \operatorname{Span} S_1$ &

→ S_2 is a pairwise orthogonal set.

Sol'n $\boxed{g_1 = f_1}$

$$g_2 = f_2 - \frac{\langle f_2, g_1\rangle}{\langle g_1, g_1\rangle} g_1$$

$$g_3 = f_3 - \frac{\langle f_3, g_1\rangle}{\langle g_1, g_1\rangle} g_1 - \frac{\langle f_3, g_2\rangle}{\langle g_2, g_2\rangle} g_2$$

compute $\langle f_2, g_1\rangle$, $\langle g_1, g_1\rangle$, $\langle f_3, g_1\rangle$, $\langle f_3, g_2\rangle$, $\langle g_2, g_2\rangle$

$$\langle f_2, g_1\rangle = \int_{-1}^{1} \underbrace{f_2(t)}_{t}\underbrace{g_1(t)}_{1}dt = \int_{-1}^{1} t\,dt = 0 \implies \boxed{g_2 = f_2}$$

$$\langle f_3, g_1\rangle = \int_{-1}^{1} t^2 dt = \frac{t^3}{3}\Big|_{-1}^{1} = \frac{2}{3}$$

$$\langle g_1, g_1\rangle = \int_{-1}^{1} dt = 2$$

$$\langle f_3, g_2\rangle = \int_{-1}^{1} t^3 dt = 0$$

$$\implies g_3 = f_3 - \frac{\langle f_3, g_1\rangle}{\langle g_1, g_1\rangle} g_1 = f_3 - \frac{2/3}{2} g_1 = \boxed{f_3 - \frac{1}{3} g_1}$$

Hence, $g_1(t) = 1$, $g_2(t) = t$, $g_3(t) = t^2 - \frac{1}{3}$.

Exercise Verify pairwise orthogonality: $\langle g_1, g_2\rangle = \langle g_2, g_3\rangle = \langle g_3, g_1\rangle = 0$.

Gram-Schmidt Orthonormalization

Same as G.-S. orthogonalization with the additional requirement that the orthogonal vectors $\{w_1, w_2, \dots, w_n\}$ are also of unit length. That is, $\langle w_i, w_j\rangle = 0 \quad \forall i \neq j$ & $\langle w_i, w_i\rangle = 1$. (In short, $\langle w_i, w_j\rangle = \delta_{ij}$.)

Algorithm: Given lin. ind. $\{v_1, v_2, \dots, v_n\}$

Step 1: $w_1 = \frac{v_1}{\|v_1\|}$

Step 2: $w_2 = \frac{y_2}{\|y_2\|}$ where $y_2 = v_2 - \langle v_2, w_1\rangle w_1$

Step 3: $w_3 = \frac{y_3}{\|y_3\|}$ where $y_3 = v_3 - \langle v_3, w_1\rangle w_1 - \langle v_3, w_2\rangle w_2$

$\vdots$

Step n: $w_n = \frac{y_n}{\|y_n\|}$ where $y_n = v_n - \sum_{i=1}^{n-1} \langle v_n, w_i\rangle w_i$

Direct Sum Decomposition

Definition Let X be a vector space and let $M_1, M_2, \ldots, M_k$ be subspaces of X. The sum of these subspaces is defined as

$$M_1 + M_2 + \cdots + M_k = \{ m \in X : m = m_1 + m_2 + \cdots + m_k \text{ where } m_i \in M_i \}$$

Theorem The sum of subspaces is also a subspace of X.

Proof Exercise.

Definition Let $M_1, M_2, \ldots, M_k$ be subspaces of a vector space X. These subspaces are said to be linearly independent if

$$m_1 + m_2 + \cdots + m_k = 0, \; m_i \in M_i \quad \text{implies} \quad m_1 = m_2 = \cdots = m_k = 0$$

Definition Given $M_1, M_2, \ldots, M_k$ subspaces of a vector space X let

i) $M = M_1 + M_2 + \cdots + M_k$ &

ii) $M_1, M_2, \ldots, M_k$ are linearly ind.

Then M is said to be the direct sum of subspaces $M_1, M_2, \ldots, M_k$ and denoted by $M = M_1 \oplus M_2 \oplus \cdots \oplus M_k$.

(direct sum symbol)

Furthermore, if $M = X$ (the linear space itself) then

$$X = M_1 \oplus M_2 \oplus \cdots \oplus M_k$$

is called a direct sum decomposition of X.

2017

<u>Example</u> Let $\mathcal{X} = \mathbb{R}^4$, $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \in \mathbb{R}^4$.

Let $M_1 = \{x \in \mathbb{R}^4 : x_3, x_4 = 0\} \Rightarrow M_1 = \text{span}\left\{ \begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix} \right\}$

$M_2 = \{x \in \mathbb{R}^4 : x_1, x_2 = 0\} \Rightarrow M_2 = \text{span}\left\{ \begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix} \right\}$

$M_3 = \{x \in \mathbb{R}^4 : x_1 = 0\} \Rightarrow M_3 = \text{span}\left\{ \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix} \right\}$

observe:

→ M_1, M_2, M_3 are subspaces

→ $M_1 + M_2 = \mathbb{R}^4$, $M_1 + M_3 = \mathbb{R}^4$, $M_2 + M_3 \neq \mathbb{R}^4$

→ M_1 & M_2 are lin. ind. $\Rightarrow$ $\mathbb{R}^4 = M_1 \oplus M_2$

→ M_1 & M_3 are not lin. ind.

because: $\underbrace{\begin{bmatrix}0\\8\\0\\0\end{bmatrix}}_{\in M_1} + \underbrace{\begin{bmatrix}0\\-8\\0\\0\end{bmatrix}}_{\in M_3} = \begin{bmatrix}0\\0\\0\\0\end{bmatrix}$

Therefore we cannot write $\cancel{\mathbb{R}^4 = M_1 \oplus M_3}$

<u>Definition</u> Let $\mathcal{X}$ be an inner product space. Two subspaces M_1 & M_2 are said to be orthogonal if

$\langle m_1, m_2 \rangle = 0$ for all $m_1 \in M_1$ & $m_2 \in M_2$

To denote M_1 & M_2 are orthogonal we write $M_1 \perp M_2$ (or $M_2 \perp M_1$)

<u>Definition</u> Let $M = M_1 \oplus M_2 \oplus \cdots \oplus M_k$ and let $M_i \perp M_j$ for all $i \neq j$. Then M is said to be the <u>orthogonal direct sum</u> of subspaces $M_1, M_2, \dots, M_k$.
(Notation: $M = M_1 \overset{\perp}{\oplus} M_2 \overset{\perp}{\oplus} \cdots \overset{\perp}{\oplus} M_k$.)

<u>Example</u> Let $\{v_1, v_2, \dots, v_n\}$ be a basis for a subspace M and $\{w_1, w_2, \dots, w_n\}$ have been obtained from $\{v_1, v_2, \dots, v_n\}$ by Gram-Schmidt orthogonalization process. Then we can write

$$M = sp\{v_1\} \oplus sp\{v_2\} \oplus \cdots \oplus sp\{v_n\} \quad \text{and}$$

$$M = sp\{w_1\} \overset{\perp}{\oplus} sp\{w_2\} \overset{\perp}{\oplus} \cdots \overset{\perp}{\oplus} sp\{w_n\}.$$

<u>Definition</u> Let M be a subspace of an inner product space $\mathcal{X}$. The <u>orthogonal complement</u> $M^\perp$ of the subspace M is defined as

$$M^\perp := \{ x \in \mathcal{X} : \langle x, m \rangle = 0 \quad \forall m \in M \}$$

<u>Theorem</u> $M^\perp$ is itself a subspace.

<u>Proof</u> Given $x, y \in M^\perp$ and scalars α, β we can write for $m \in M$

$$\langle \alpha x + \beta y, m \rangle = \alpha \langle x, m \rangle + \beta \langle y, m \rangle = 0 \implies \alpha x + \beta y \in M^\perp \quad \blacksquare$$

<u>Example</u> $\mathcal{X} = \mathbb{R}^3$, $\langle x, y \rangle = y^T x$, $M = \text{span}\left\{ \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$. $M^\perp = ?$

$$m \in M \implies m = \alpha \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \text{ for some } \alpha, \beta \in \mathbb{R} \implies m = \begin{bmatrix} 0 & -1 \\ -1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

Let $x = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \in M^\perp$. Then $m^T x = 0 \quad \forall m \in M$

$$\implies \underbrace{[\alpha \ \ \beta] \begin{bmatrix} 0 & -1 & 1 \\ -1 & 0 & 1 \end{bmatrix}}_{m^T} \underbrace{\begin{bmatrix} a \\ b \\ c \end{bmatrix}}_{x} = 0 \quad \forall \alpha, \beta$$

$$\implies \begin{bmatrix} 0 & -1 & 1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\implies \begin{cases} -b + c = 0 \\ -a + c = 0 \end{cases} \implies a = b = c \implies M^\perp = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

Theorem Let X be an inner product space and M is a subspace of X. We can always write $X = M \oplus M^{\perp}$. That is, X can always be written as the direct sum of a subspace and its orthogonal complement.

proof We need to show two things:

1) M and $M^{\perp}$ are linearly independent
2) Any $x \in X$ can be expressed as $x = x_1 + x_2$ where $x_1 \in M$ & $x_2 \in M^{\perp}$.

part I Let $x_1 + x_2 = 0$ with $x_1 \in M$ & $x_2 \in M^{\perp}$. (Linear independence requires $x_1 = 0$ and $x_2 = 0$.) We can write

$$
\begin{aligned}
0 &= \langle 0, x_1 \rangle \\
&= \langle x_1 + x_2, x_1 \rangle \\
&= \langle x_1, x_2 \rangle + \cancel{\langle x_2, x_1 \rangle}^{\,0} \quad \text{since } x_1 \perp x_2 \\
&= \|x_1\|^2
\end{aligned}
$$

$\Rightarrow \quad x_1 = 0$

Likewise,

$\langle x_1 + x_2, x_2 \rangle = 0 \Rightarrow x_2 = 0$

Therefore, M & $M^{\perp}$ are lin. ind.

part II Let $k = \dim M$ & $l = \dim M^{\perp}$.

Let $M = \operatorname{span}\{v_1, v_2, \dots, v_k\}$
& $M^{\perp} = \operatorname{span}\{w_1, w_2, \dots, w_l\}$

apply Gram-Schmidt orthonormalization and obtain $\{e_1, e_2, \dots, e_k\}$ & $\{f_1, f_2, \dots, f_l\}$

such that $M = \operatorname{span}\{e_1, e_2, \dots, e_k\}$ with $\langle e_i, e_j \rangle = \delta_{ij}$

& $M^{\perp} = \operatorname{span}\{f_1, f_2, \dots, f_l\}$ with $\langle f_i, f_j \rangle = \delta_{ij}$

Note that $\langle e_i, f_j \rangle = 0 \quad \forall i, j$.

Let now $x \in X$ be given. Define

$$
\begin{aligned}
x_1 &:= \langle x, e_1 \rangle e_1 + \langle x, e_2 \rangle e_2 + \dots + \langle x, e_k \rangle e_k \\
x_2 &:= \langle x, f_1 \rangle f_1 + \langle x, f_2 \rangle f_2 + \dots + \langle x, f_l \rangle f_l
\end{aligned}
$$

observe that $x_1 \in M$ & $x_2 \in M^{\perp}$

Claim $x = x_1 + x_2$

Suppose not. Then define $v := x - (x_1+x_2) \neq 0$. Observe that for $j = 1,2,\dots,k$

$$\langle v, e_j \rangle = \langle x - \sum_{i=1}^{k} \langle x, e_i\rangle e_i - \sum_{i=1}^{\ell} \langle x, f_i \rangle f_i \, , \, e_j \rangle$$

$$= \langle x, e_j \rangle - \underbrace{\sum_i \langle x, e_i \rangle \underbrace{\langle e_i, e_j \rangle}_{\delta_{ij}}}_{\langle x, e_j \rangle \underbrace{\langle e_j, e_j \rangle}_{1}} - \sum_i \langle x, f_i \rangle \underbrace{\langle f_i, e_j \rangle}_{0}$$

$$= \langle x, e_j \rangle - \langle x, e_j \rangle = 0$$

Therefore $v \in M^{\perp}$ (1).

Likewise, $\langle v, f_j \rangle = 0$ for all $j = 1,2,\dots,\ell$. Hence $v \in M$ (2).

(1) & (2) $\Rightarrow$ $\langle v, v \rangle = 0$ $\Rightarrow$ $v = 0$ (contradiction)

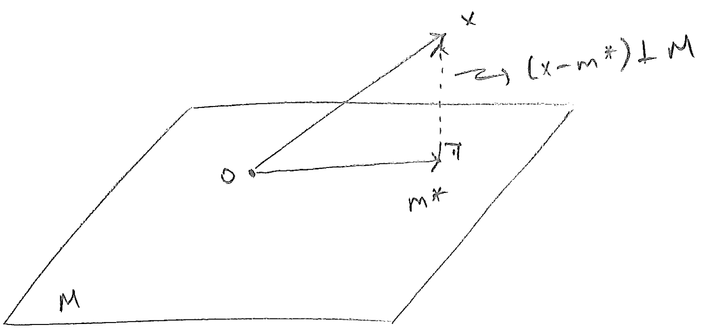
Hence the claim holds. ▨

<u>Theorem</u> (Projection Thm.) Let H be a Hilbert Space (inner product space, complete wrt the norm induced by the inner product) and let M be a finite dimensional subspace of H. For each $x \in H$, the following minimization problem has a sol'n.

$$\boxed{\min_{m \in M} \|x - m\|}$$

In words: "We can find a closest vector to x lying in the subspace M."

If $m^* \in M$ is a solution then $x - m^*$ is orthogonal to M. Furthermore, the solution m^* is <u>unique</u> (and is called the orthogonal projection of x onto M.)



proof We can write $H = M \oplus M^\perp$. Then for each given $x \in H$, there exist a unique (why unique ?) pair (x_1, x_2) satisfying $x_1 \in M$, $x_2 \in M^\perp$, and $x = x_1 + x_2$.

Let $\sigma_m = \|x-m\|^2$ for $m \in M$.

$$\Rightarrow \sigma_m = \|x_1+x_2-m\|^2 = \|\underbrace{(x_1-m)}_{\in M} + \underbrace{x_2}_{\in M^\perp}\|^2$$

$$\Rightarrow \sigma_m = \langle (x_1-m)+x_2 , (x_1-m)+x_2 \rangle$$

$$= \langle x_1-m, x_1-m\rangle + \cancelto{0}{\langle \underbrace{x_1-m}_{\in M}, \underbrace{x_2}_{\in M^\perp}\rangle} + \cancelto{0}{\langle x_2, x_1-m\rangle} + \langle x_2, x_2\rangle$$

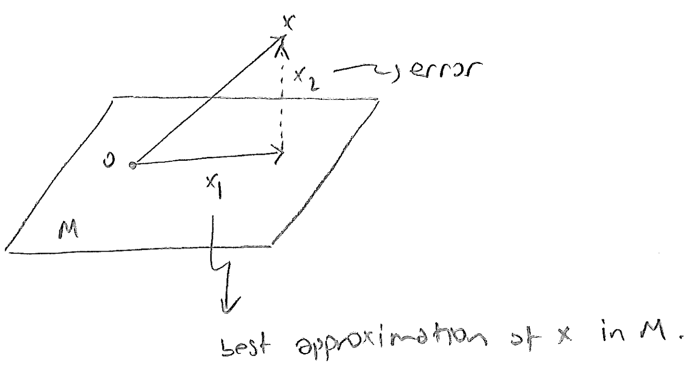
$$= \|x_1-m\|^2 + \|x_2\|^2 \qquad (1)$$

It is clear from (1) that $\min\limits_{m\in M} \sigma_m$ is attained by letting $m = x_1$.

Hence $\boxed{m^* = x_1}$ Since x_1 is unique, so is m^*.

Furthermore, $x - m^* = x - x_1 = x_2$ is orthogonal to M. ▨

<u>Remark</u> $m^* = x_1$ can be interpreted as the <u>best approximation</u> of x chosen from the vectors in M. The vector x_2 can be interpreted as the <u>error in approximation</u> (Error must be orthogonal to the subspace M.)



2017

Computational aspects of the projection thm

Problem : Given $x \in H$, subspace $M \subset H$, and a basis $\{v_1, v_2, \dots, v_k\}$ for M find $x_1 \in M$ where $x = x_1 + x_2$ with $x_2 \in M^\perp$.

Sol'n : There exist scalars $\alpha_1, \alpha_2, \dots, \alpha_k$ such that $x_1 = \sum_{i=1}^{k} \alpha_i v_i$ (since $x_1 \in M$)

Note that once we obtain $\alpha_1, \alpha_2, \dots, \alpha_k$; the vector x_1 is no longer unknown.

Question: How to compute $\alpha_1, \alpha_2, \dots, \alpha_k$?

Answer : $x_2 \perp M \Rightarrow \langle x_2, v_j \rangle = 0$ for $j = 1, 2, \dots, k$

$$\Rightarrow \langle x - \sum_{i=1}^{k} \alpha_i v_i , v_j \rangle = 0 \qquad \forall j$$

$$\Rightarrow \sum_{i=1}^{k} \alpha_i \langle v_i, v_j \rangle = \langle x, v_j \rangle \qquad \forall j$$

$$\Rightarrow \begin{cases} \alpha_1 \langle v_1, v_1 \rangle + \alpha_2 \langle v_2, v_1 \rangle + \cdots + \alpha_k \langle v_k, v_1 \rangle = \langle x, v_1 \rangle & (j=1) \\ \alpha_1 \langle v_1, v_2 \rangle + \alpha_2 \langle v_2, v_2 \rangle + \cdots + \alpha_k \langle v_k, v_2 \rangle = \langle x, v_2 \rangle & (j=2) \\ \vdots & \vdots \\ \alpha_1 \langle v_1, v_k \rangle + \alpha_2 \langle v_2, v_k \rangle + \cdots + \alpha_k \langle v_k, v_k \rangle = \langle x, v_k \rangle & (j=k) \end{cases}$$

$$\Rightarrow \underbrace{\begin{bmatrix} \langle v_1, v_1 \rangle & \cdots & \langle v_k, v_1 \rangle \\ \vdots & \ddots & \vdots \\ \langle v_1, v_k \rangle & \cdots & \langle v_k, v_k \rangle \end{bmatrix}}_{G \text{ (known)}} \underbrace{\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix}}_{\alpha \text{ (unknown)}} = \underbrace{\begin{bmatrix} \langle x, v_1 \rangle \\ \vdots \\ \langle x, v_k \rangle \end{bmatrix}}_{b \text{ (known)}}$$

$$\Rightarrow G\alpha = b \Rightarrow \boxed{\alpha = G^{-1} b}$$

Exercise Why G^{-1} exists ?

Remark If $\{v_1, v_2, \dots, v_k\}$ is an orthonormal basis, then we have $G = I$.

Then $\alpha = b$. That is, $\boxed{\alpha_i = \langle x, v_i \rangle}$ $i = 1, 2, \dots, k$, Example Let H be the space of square integrable functions with domain $[-\pi, \pi]$ with inner product $\langle f_1, f_2 \rangle = \int_{-\pi}^{\pi} f_1(t)\overline{f_2(t)}\,dt$. Let M be the subspace spanned as

$$M = \text{Span}\left\{ \frac{e^{jkt}}{\sqrt{2\pi}} \right\}_{k=-N}^{N}$$

$\Rightarrow$ normalization term : $\left\langle \frac{e^{jkt}}{\sqrt{2\pi}}, \frac{e^{jkt}}{\sqrt{2\pi}} \right\rangle = 1$

Note that $\dim M = 2N+1$ and the basis set $\left\{ \frac{e^{jkt}}{\sqrt{2\pi}} \right\}_{k=-N}^{N}$ is orthonormal.

Because $\left\langle \frac{e^{jnt}}{\sqrt{2\pi}}, \frac{e^{jmt}}{\sqrt{2\pi}} \right\rangle = \int_{-\pi}^{\pi} \frac{1}{2\pi} e^{j(n-m)t}\,dt = \delta_{nm}$.

Now, let $g \in H$ be an arbitrary function. Then we can uniquely decompose g as $g = g_1 + g_2$ with $g_1 \in M$ & $g_2 \in M^{\perp}$.

$$g_1 \in M \Rightarrow g_1 = \sum_{k=-N}^{N} \alpha_k \left(\frac{e^{jkt}}{\sqrt{2\pi}} \right).$$

$\Rightarrow$ basis vectors

Question $\alpha_k = ?$

Answer According to the projection thm we have $\alpha_k = \left\langle g, \frac{e^{jkt}}{\sqrt{2\pi}} \right\rangle$.

Therefore $\alpha_k = \int_{-\pi}^{\pi} g(t) \frac{e^{-jkt}}{\sqrt{2\pi}}\,dt$.

$$\Rightarrow g_1(t) = \sum_{k=-N}^{N} \underbrace{\left\{ \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} g(t) e^{-jkt}\,dt \right\}}_{\alpha_k} \frac{e^{jkt}}{\sqrt{2\pi}}$$

$g_1(t)$ is the <u>best approximation</u> to $g(t)$ within the subspace M. The function $g_1(t)$ turns out to be the finite Fourier series representation of $g(t)$. As $N \to \infty$ we obtain the <u>Fourier series</u> representation.

Application of the projection thm. in $\mathbb{C}^n$

Let $\{m_1, m_2, \ldots, m_k\}$ $(k \le n)$ be a basis for a subspace M of $\mathbb{C}^n$.

Let $\langle x, y \rangle = y^* x$ be the inner product, where "$*$" indicates conjugate transpose.

eg $\begin{bmatrix} 3+j4 \\ 1-j2 \end{bmatrix}^* = \begin{bmatrix} 3-j4 & 1+j2 \end{bmatrix}$

Given an arbitrary vector $x \in \mathbb{C}^n$ we can write

$$x = x_1 + x_2 \quad \text{with} \quad x_1 \in M \ \& \ x_2 \in M^\perp \quad (\text{Recall: } x_1, x_2 \text{ unique!})$$

Problem Compute x_1 (and x_2).

Sol'n We can write

$$x_1 = \sum_{i=1}^{k} \alpha_i m_i = \underbrace{\begin{bmatrix} | & | & & | \\ m_1 & m_2 & \cdots & m_k \\ | & | & & | \end{bmatrix}}_{B\ (n \times k)} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix}}_{\alpha} \Rightarrow x_1 = B\alpha$$

(α: unknown)

Observe: $x_2 \perp M \Rightarrow (x - x_1) \perp M \Rightarrow \langle x - B\alpha, m_i \rangle = 0 \quad i = 1, 2, \ldots, k$

$$\Rightarrow m_i^* (x - B\alpha) = 0 \quad \forall i$$

$$\Rightarrow B^* (x - B\alpha) = 0 \qquad B^* = \begin{bmatrix} - m_1^* - \\ \vdots \\ - m_k^* - \end{bmatrix}_{k \times n}$$

Then: $B^* B \alpha = B^* x \Rightarrow \alpha = [B^* B]^{-1} B^* x$

$$\Rightarrow x_1 = B [B^* B]^{-1} B^* x \qquad x_1 = B\alpha$$

Hence we've obtained $\boxed{x_1 = B[B^*B]^{-1}B^* x}$

→ x_1 is the orthogonal projection of x onto M.

→ the matrix $[B[B^*B]^{-1}B^*]_{n \times n}$ is the projection (onto M) matrix.

As for x_2: $x_2 = x - x_1 = x - B[B^*B]^{-1}B^* x = \underbrace{[I - B[B^*B]^{-1}B^*]}_{\text{the projection (onto } M^\perp) \text{ matrix}} x$

Remark: The proj. matrix is basis (B) independent. (why?)

2017

Remark An orthogonal projection matrix $P \in \mathbb{C}^{n\times n}$ satisfies

1) $P^* = P$ (for $\mathbb{R}^{n\times n}$ we can write $P^T = P$)

2) $P^2 = P$

Exercise Let P satisfy 1 & 2. Then $\hat{P} = I - P$ also satisfies 1 & 2.

Example [orthogonal projection] Find the orthogonal projection of the vector

$x = \begin{bmatrix}1\\0\\0\end{bmatrix} \in \mathbb{R}^3$ onto the subspace $M = \operatorname{span}\left\{ \underbrace{\begin{bmatrix}1\\1\\1\end{bmatrix}}_{m_1}, \underbrace{\begin{bmatrix}1\\-1\\1\end{bmatrix}}_{m_2} \right\}$

Sol'n $B = \begin{bmatrix}1 & 1\\ 1 & -1\\ 1 & 1\end{bmatrix} \Rightarrow B^T B = \begin{bmatrix}1 & 1 & 1\\ 1 & -1 & 1\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\\ 1 & 1\end{bmatrix} = \begin{bmatrix}3 & 1\\ 1 & 3\end{bmatrix}$

$$\Rightarrow [B^T B]^{-1} = \frac{1}{8}\begin{bmatrix}3 & -1\\ -1 & 3\end{bmatrix}$$

$$\Rightarrow B[B^T B]^{-1}B^T = \begin{bmatrix}1 & 1\\ 1 & -1\\ 1 & 1\end{bmatrix}\left\{\frac{1}{8}\begin{bmatrix}3 & -1\\ -1 & 3\end{bmatrix}\right\}\begin{bmatrix}1 & 1 & 1\\ 1 & -1 & 1\end{bmatrix} = \begin{bmatrix}\frac{1}{2} & 0 & \frac{1}{2}\\ 0 & 1 & 0\\ \frac{1}{2} & 0 & \frac{1}{2}\end{bmatrix} =: P$$

$\Rightarrow$ P is the projection matrix for $M = \operatorname{span}\left\{\begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\\1\end{bmatrix}\right\}$

(observe: $P^2 = P$ & $P^T = P$)

$\Rightarrow P \underbrace{\begin{bmatrix}1\\0\\0\end{bmatrix}}_{x} = \underbrace{\begin{bmatrix}1/2\\0\\1/2\end{bmatrix}}_{x_1}$ $\qquad x_1 = \begin{bmatrix}1/2\\0\\1/2\end{bmatrix} \in M$ is the projection of x onto M.

Exercise verify that $\underbrace{x_2 = x - x_1}_{\in M^\perp}$ is orthogonal to both m_1 and m_2.

2017

Application of the projection theorem to the solution of $Ax=b$

Consider the linear equation expressed as

$Ax=b$ where $A \in \mathbb{C}^{m\times n}$ & $b \in \mathbb{C}^{m\times 1}$ are known

$x \in \mathbb{C}^{n\times 1}$ unknown

q1) When is there a solution x ?

q2) If a solution exists, when is it unique ?

A1) When (and only when) $b \in \mathcal{R}(A)$.

A2) When (and only when) $\mathcal{N}(A) = \{0\}$.

Example Let $A = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$ and $b = \begin{bmatrix}2.1\\1.9\\2.0\\2.2\end{bmatrix}$.

$$Ax=b \Rightarrow \begin{bmatrix}1\\1\\1\\1\end{bmatrix}\underbrace{x}_{\text{1-by-1 vector (i.e. scalar)}} = \begin{bmatrix}2.1\\1.9\\2.0\\2.2\end{bmatrix}$$

Now, $\mathcal{R}(A) = \text{span}\left\{\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\right\}$ and clearly $b \notin \mathcal{R}(A)$. Therefore there is no $x \in \mathbb{R}$ satisfying $Ax=b$. No solution exists. Now what ?

— o —

When there is no exact solution usually we go for the "best approximation." That is, we look for an $\hat{x}$ that would solve the following minimization problem

$$\boxed{\min_{x\in\mathbb{C}^n} \|Ax-b\|^2}$$

Remark If a solution to $Ax=b$ exists, then $\min_x \|Ax-b\|^2 = 0$.

Otherwise $\min_x \|Ax-b\|^2 > 0$.

<u>Solving $\min_{x\in\mathbb{C}^n} \|Ax-b\|^2$</u> $\quad (A\in\mathbb{C}^{m\times n},\ b\in\mathbb{C}^{m\times 1})$

Since $R(A)$ is a subspace of $\mathbb{C}^m$ we can write

$$\mathbb{C}^m = R(A)\oplus R(A)^\perp \quad \text{and} \quad b = \underbrace{b_1}_{\in R(A)} + \underbrace{b_2}_{\in R(A)^\perp}$$

$$\Rightarrow \|Ax-b\|^2 = \langle Ax-b, Ax-b\rangle$$

$$= \langle (Ax-b_1)-b_2,\ (Ax-b_1)-b_2\rangle$$

$$= \langle Ax-b_1, Ax-b_1\rangle + \langle b_2,b_2\rangle + \underbrace{\langle Ax-b_1, -b_2\rangle}_{=0} + \underbrace{\langle -b_2, Ax-b_1\rangle}_{=0\ (\text{why?})}$$

$$= \|Ax-b_1\|^2 + \|b_2\|^2$$

Therefore $\|Ax-b\|^2 = \|Ax-b_1\|^2 + \|b_2\|^2$. This implies the solution:

<u>Sol'n</u> To minimize $\|Ax-b\|^2$ choose some $\hat{x}$ satisfying $A\hat{x}=b_1$ [and we can always find such $\hat{x}$ because $b_1\in R(A)$]. Then

$$\min_x \|Ax-b\|^2 = \|b_2\|^2$$

<u>Example</u> The length x of a metal rod is inaccurately measured four times and four different values l_1, l_2, l_3, l_4 are obtained. What is the best approximation to x?

$$\underbrace{\begin{bmatrix}1\\1\\1\\1\end{bmatrix}}_{A} x = \underbrace{\begin{bmatrix}l_1\\l_2\\l_3\\l_4\end{bmatrix}}_{b}$$

Best approximation is the solution to $Ax=b_1$ where $b = \underbrace{b_1}_{\in R(A)} + \underbrace{b_2}_{\in R(A)^\perp}$

Note that b_1 is the projection of b onto $R(A) = \text{span}\left\{\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\right\}$.

2017

Projection matrix P onto $\mathcal{R}(A)$? (Once we have P we can write $b_1 = Pb$)

$P = B(B^TB)^{-1}B^T$ where B is any matrix whose columns make a basis for $\mathcal{R}(A)$.

Clearly, we can take $B = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$. Then

$$P = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \left([1\ 1\ 1\ 1] \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right)^{-1} [1\ 1\ 1\ 1] = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}$$

$$\Rightarrow b_1 = P \begin{bmatrix} l_1 \\ \vdots \\ l_4 \end{bmatrix} = \begin{bmatrix} \frac{1}{4}(l_1+l_2+l_3+l_4) \\ \vdots \\ \frac{1}{4}(l_1+l_2+l_3+l_4) \end{bmatrix}$$

$\Rightarrow$ The solution of $Ax = b_1$ then is $\boxed{\hat{x} = \frac{1}{4}(l_1+l_2+l_3+l_4)}$

Example Now consider the following scenario.

$$\underbrace{\begin{bmatrix} 2 & 1 \\ 2 & 1 \\ 2 & 1 \\ 2 & 1 \end{bmatrix}}_{A} x = \underbrace{\begin{bmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \end{bmatrix}}_{b}$$

Clearly, $\mathcal{R}(A) = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}$, same as before.

Hence, the projection of b onto $\mathcal{R}(A)$ is also same as before.

$$b_1 = \begin{bmatrix} \bar{l} \\ \bar{l} \\ \bar{l} \\ \bar{l} \end{bmatrix} \quad \text{where } \bar{l} = \frac{1}{4}(l_1+l_2+l_3+l_4)$$

Now, $Ax = b_1$ reads $\begin{bmatrix} 2 & 1 \\ 2 & 1 \\ 2 & 1 \\ 2 & 1 \end{bmatrix} x = \begin{bmatrix} \bar{l} \\ \bar{l} \\ \bar{l} \\ \bar{l} \end{bmatrix}$. That is, the solution is no longer unique

e.g., $x = \begin{bmatrix} 0 \\ \bar{l} \end{bmatrix}$, $x = \begin{bmatrix} \bar{l}/2 \\ 0 \end{bmatrix}$, $x = \begin{bmatrix} 500\,\bar{l} \\ -999\,\bar{l} \end{bmatrix}$, ...

Question: How to choose among all x satisfying $Ax = b_1$?

A meaningful answer: Choose the one with minimum norm $\|x\|$. (why?)

Minimum norm $\|x\|$ solution to $Ax = b_1$

Let $\hat{x}$ & $\tilde{x}$ be two solutions to $Ax = b_1$. That is, we have both $A\hat{x} = b_1$ and $A\tilde{x} = b_1$. We can decompose uniquely:

$$\hat{x} = \underbrace{\hat{x}_1}_{\in N(A)^{\perp}} + \underbrace{\hat{x}_2}_{\in N(A)} \qquad \text{and} \qquad \tilde{x} = \underbrace{\tilde{x}_1}_{\in N(A)^{\perp}} + \underbrace{\tilde{x}_2}_{\in N(A)}$$

Claim $\tilde{x}_1 = \hat{x}_1$

Proof $A(\hat{x}_1 - \tilde{x}_1) = A(\hat{x}_1 - \tilde{x}_1) + \underbrace{A(\hat{x}_2 - \tilde{x}_2)}_{=0 \text{ (why?)}} = A(\underbrace{\hat{x}_1 + \hat{x}_2}_{\hat{x}}) - A(\underbrace{\tilde{x}_1 + \tilde{x}_2}_{\tilde{x}}) = b_1 - b_1 = 0$

Hence; $\hat{x}_1 - \tilde{x}_1 \in N(A)$ (1).

Also, $\hat{x}_1 - \tilde{x}_1 \in N(A)^{\perp}$ (2) (why?)

(1) & (2) $\Rightarrow \langle \hat{x}_1 - \tilde{x}_1, \hat{x}_1 - \tilde{x}_1 \rangle = 0 \Rightarrow \hat{x}_1 - \tilde{x}_1 = 0.$ ∎

Therefore for all x satisfying $Ax = b_1$ $\text{proj}_{N(A)^{\perp}} x = x_1$ is the same.

Moreover, $x_1 + z$ is a solution for all $z \in N(A)$ because

$$A(x_1 + z) = \underbrace{Ax_1}_{b_1} + \underbrace{Az}_{0} = b_1$$

Question How to choose $z \in N(A)$ so that $\|x_1 + z\|^2$ is minimum?

Answer $\|x_1 + z\|^2 = \langle x_1 + z, x_1 + z \rangle = \langle x_1, x_1 \rangle + \langle z, z \rangle + \cancelto{0}{\langle x_1, z \rangle} + \cancelto{0}{\langle z, x_1 \rangle}$

$$= \|x_1\|^2 + \|z\|^2$$

Clearly, we choose $z = 0$. Therefore the minimum norm solution to $Ax = b_1$ is x_1, which we find by:

1) First, find any solution x to $Ax = b_1$.
2) Then $x_1 = \left[\text{proj}_{N(A)^{\perp}}\right] x$.

2017

To construct the projection matrix $\left[\operatorname{proj}_{N(A)^\perp}\right]$ the following thm may be useful.

<u>Theorem</u> $N(A)^\perp = R(A^*)$ (therefore $\operatorname{proj}_{N(A)^\perp} = \operatorname{proj}_{R(A^*)}$)

<u>Proof</u>
$$\begin{aligned} x \in N(A) &\iff Ax = 0 \\ &\iff \langle y, Ax\rangle = 0 \quad \forall y \\ &\iff \langle A^*y, x\rangle = 0 \quad \forall y \\ &\iff x \perp R(A^*) \\ &\iff x \in R(A^*)^\perp \end{aligned}$$

$\langle \eta, \xi\rangle = \xi^* \eta$

Hence $N(A) = R(A^*)^\perp \Rightarrow N(A)^\perp = \left(R(A^*)^\perp\right)^\perp = R(A^*)$. ☑

— o —

<u>Example</u> (Revisited)

$$\underbrace{\begin{bmatrix} 2 & 1 \\ 2 & 1 \\ 2 & 1 \\ 2 & 1 \end{bmatrix}}_{A} x = \underbrace{\begin{bmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \end{bmatrix}}_{b} \Rightarrow \begin{bmatrix} 2 & 1 \\ 2 & 1 \\ 2 & 1 \\ 2 & 1 \end{bmatrix} x = \underbrace{\begin{bmatrix} \bar{l} \\ \bar{l} \\ \bar{l} \\ \bar{l} \end{bmatrix}}_{b_1 = \operatorname{proj}_{R(A)} b} \qquad \bar{l} = \frac{1}{4}(l_1 + l_2 + l_3 + l_4)$$

$\Rightarrow$ A solution : $x = \begin{bmatrix} 0 \\ \bar{l} \end{bmatrix}$ to $Ax = b_1$

Another solution $x = \begin{bmatrix} 500\bar{l} \\ -999\bar{l} \end{bmatrix}$.

$\Rightarrow$ min. norm solution $x_1 = \underbrace{\operatorname{proj}_{N(A)^\perp}}_{?} x$

$N(A)^\perp = R(A^*)$, $A^* = \begin{bmatrix} 2 & 2 & 2 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix} \Rightarrow R(A^*) = \operatorname{span}\left\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right\}$

$\Rightarrow \operatorname{proj}_{N(A)^\perp} = \operatorname{proj}_{R(A^*)} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}\left(\begin{bmatrix} 2 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right)^{-1}\begin{bmatrix} 2 & 1 \end{bmatrix} = \begin{bmatrix} 4/5 & 2/5 \\ 2/5 & 1/5 \end{bmatrix}$

$\Rightarrow x_1 = \begin{bmatrix} 4/5 & 2/5 \\ 2/5 & 1/5 \end{bmatrix}\begin{bmatrix} 0 \\ \bar{l} \end{bmatrix} = \begin{bmatrix} 2\bar{l}/5 \\ \bar{l}/5 \end{bmatrix}$ or $x_1 = \begin{bmatrix} 4/5 & 2/5 \\ 2/5 & 1/5 \end{bmatrix}\begin{bmatrix} 500\bar{l} \\ -999\bar{l} \end{bmatrix} = \begin{bmatrix} 2\bar{l}/5 \\ \bar{l}/5 \end{bmatrix}$.

SUMMARY

Start with $\boxed{Ax = b}$ Solution x may or may not exist. A solution exists if and only if $b \in R(A)$

$A \in \mathbb{C}^{m\times n}$, $b \in \mathbb{C}^{m\times 1}$ } known
$x \in \mathbb{C}^{n\times 1}$: unknown

$\downarrow$

$\boxed{Ax = \text{proj}_{R(A)} b}$ Now a solution must exist. (Note that if $b \in R(A)$ then $\text{proj}_{R(A)} b = b$.)

$\downarrow$

$\boxed{x_{min} = \text{proj}_{R(A^*)} x}$ We obtain x_{min} (solution with minimum norm) from an arbitrary solution x ($Ax = \text{proj}_{R(A)} b$) by projecting x onto $N(A)^\perp$ ($x_{min} = \text{proj}_{N(A)^\perp} x$)

Note that x_{min} is unique and satisfies

1) $\|Ax_{min} - b\| \leq \|Ax - b\|$ for all x

2) $\|x_{min}\| \leq \|x\|$ for all x satisfying $\|Ax-b\| = \|Ax_{min} - b\|$

Recall Given $A \in \mathbb{C}^{m\times n}$ let $B \in \mathbb{C}^{m\times k}$ be a full column rank matrix satisfying $R(B) = R(A)$. Then

$$\text{proj}_{R(A)} = B(B^*B)^{-1}B^*$$

Recall $N(A)^\perp = R(A^*)$. Therefore $\text{proj}_{N(A)^\perp} = \text{proj}_{R(A^*)}$

Example Consider the discrete-time LTI system

$$x_{k+1} = Ax_k + Bu_k \quad , \quad A \in \mathbb{R}^{n\times n} \quad , \quad B \in \mathbb{R}^{n\times 1}$$

Suppose the system is controllable, i.e., range $[B \ AB \cdots A^{n-1}B] = \mathbb{R}^n$

Given $x_0 \in \mathbb{R}^n$ and $N \geq n$ find the min-norm input sequence $(u_0, u_1, \dots, u_{N-1})$ satisfying $x_N = 0$.

$$x_1 = Ax_0 + Bu_0$$

$$x_2 = Ax_1 + Bu_1 = A(Ax_0 + Bu_0) + Bu_1 = A^2x_0 + ABu_0 + Bu_1$$

$$\vdots$$

$$\underbrace{x_N}_{0} = A^N x_0 + \underbrace{[A^{N-1}B \quad A^{N-2}B \quad \cdots \quad AB \quad B]}_{W} \underbrace{\begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_{N-1} \end{bmatrix}}_{u} \qquad (1)$$

Since the system is controllable & $N \geq n$ $\exists$ a solution $u \in \mathbb{R}^N$ exists for (1).

How about the min. norm solution?

Let $\hat{u} \in \mathbb{R}^N$ solves $W\hat{u} + A^N x_0 = 0$

min norm solution $u_{min} = \text{proj}_{\mathcal{N}(W)^\perp} \hat{u}$

$$\text{proj}_{\mathcal{N}(W)^\perp} = \text{proj}_{\mathcal{R}(W^T)} = W^T(WW^T)^{-1}W \qquad (\text{why } [WW^T]^{-1} \text{ exists?})$$

$$\Rightarrow u_{min} = W^T(WW^T)^{-1}W\hat{u} \qquad \& \quad W\hat{u} = -A^N x_0$$

$$= -W^T(WW^T)^{-1}A^N x_0$$

u_{min} is called the "minimum energy" control since it solves the optim. problem:

$$J(x_0) = \min \sum_{k=0}^{N-1} u_k^2 \quad \text{subject to} \quad \begin{cases} x_{k+1} = Ax_k + Bu_k \\ x_N = 0 \end{cases}$$

2017

Some special cases of the problem $\min_x \|Ax-b\|$

$\left(\begin{array}{l} A \in \mathbb{C}^{m\times n} \\ b \in \mathbb{C}^{m\times 1} \\ x \in \mathbb{C}^{n\times 1} \end{array} \right.$

1) Columns of A lin. ind. (i.e. A full column rank, $m \geq n$)

We can write $\text{proj}_{\mathcal{R}(A)} = A[A^*A]^{-1}A^*$.

Then $Ax = \text{proj}_{\mathcal{R}(A)} b \Rightarrow Ax = A[A^*A]^{-1}A^*b$

$\Rightarrow x = [A^*A]^{-1}A^*b$ minimizes $\|Ax-b\|$

Note that $N(A) = \{0\}$ because the columns are lin. ind. Hence $N(A)^{\perp} = \mathbb{C}^n$.

Then $\text{proj}_{N(A)^\perp} x = x$. That is, $\boxed{x = [A^*A]^{-1}A^*b}$ is the unique solution that minimizes $\|Ax-b\|$. (Remark: $[A^*A]^{-1}A^*$ is called the left inverse of A.)

2) Rows of A lin. ind. (i.e. A full row rank, $m \leq n$)

We have $N(A^*) = \{0\} \Rightarrow \mathcal{R}(A)^\perp = \{0\} \Rightarrow \mathcal{R}(A) = \mathbb{C}^m \Rightarrow b \in \mathcal{R}(A)$

Hence there exists a solution satisfying $\|Ax-b\| = 0$. Then the min. norm solution is $\text{proj}_{N(A)^\perp} x = \text{proj}_{\mathcal{R}(A^*)} x$.

We can write $\text{proj}_{\mathcal{R}(A^*)} = A^*[AA^*]^{-1}A$. Hence $\text{proj}_{\mathcal{R}(A^*)} x = A^*[AA^*]^{-1}\underbrace{Ax}_{b}$

That is, $\boxed{x_{min} = A^*[AA^*]^{-1}b}$ is the min. norm solution satisfying $\|Ax-b\| = 0$.

(Remark: $A^*[AA^*]^{-1}$ is called the right inverse of A.)

3) Both rows & columns of A lin. ind. (i.e. A invertible, $m=n$)

In this case the solution to $Ax=b$ exists for all b and it equals $x = A^{-1}b$, where A^{-1} is the inverse of A ($A^{-1}A = AA^{-1} = I$). Note that for such A the right and left inverses both equal A^{-1}.

Exercise obtain x_{min} for

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}}_{b}$$

Answer: $x_{min} = \left[-\frac{23}{56} \quad \frac{1}{9} \quad \frac{31}{36} \right]^T$

$$\text{proj}_{\mathcal{R}(A)} = \begin{bmatrix} 1/9 & 2/9 & 2/9 \\ 2/9 & 17/18 & -1/18 \\ 2/9 & -1/18 & 17/18 \end{bmatrix}, \quad \text{proj}_{\mathcal{R}(A^T)} = \begin{bmatrix} 5/6 & 1/3 & -1/6 \\ 1/3 & 1/3 & 1/3 \\ -1/6 & 1/3 & 5/6 \end{bmatrix}$$

$b_1 = \left[\frac{1}{3} \quad \frac{13}{6} \quad -\frac{5}{6} \right]^T$

Spectral Analysis of Linear Operators

Definition [Invariant Subspaces] Let $A : V \to V$ be a linear transformation defined over the vector space V. A subspace M of V is said to be invariant under A if $A(x) \in M$ for all $x \in M$.

Examples 1) $R(A)$ is invariant under A.

Let $x \in R(A) \Rightarrow Ax \in R(A)$ by definition

2) $N(A)$ is invariant under A, too.

Let $x \in N(A) \Rightarrow Ax = 0 \in N(A)$.

Definition Powers of a linear operator are defined as

$$A^k(x) := \underbrace{A(A(\cdots A(x)\cdots))}_{A \text{ applied } k \text{ times}}$$

This allows us to construct new linear transformations written as polynomials of A:

Given $p(s) = \alpha_0 s^n + \alpha_1 s^{n-1} + \cdots + \alpha_{n-1} s + \alpha_n$

let $p(A) := \alpha_0 A^n + \alpha_1 A^{n-1} + \cdots + \alpha_{n-1} A + \alpha_n I$

where I is the identity operator, i.e., $I(x) = x \ \forall x$. Note that $p(A) : V \to V$.

Exercise Show that $R(p(A))$ and $N(p(A))$ are invariant under A.

Definition Let $A \in \mathbb{C}^{n \times n}$ be the matrix representation of a linear operator from V to V, with $\dim V = n$. The eigenvalues of A, denoted by $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{C}$ are defined as the n roots of the equation $\det(sI - A) = 0$. The (nth order) polynomial $d(s) := \det(sI - A)$ is known as the characteristic polynomial of A.

<u>Definition</u> The vector(s) $e_i \in V$ satisfying $e_i \neq 0$ and $Ae_i = \lambda_i e_i$ is called the eigenvector(s) of A corresponding to the eigenvalue λ_i.

<u>Example</u> Let $A \in \mathbb{C}^{n\times n}$ and λ_i be an eigenvalue of A. Then $N(A-\lambda_i I)$ is invariant under A.

<u>Proof</u> Let $x \in N(A-\lambda_i I)$. Define $y = Ax$. Then

$$(A-\lambda_i I)y = (A-\lambda_i I)Ax = A^2x - \lambda_i Ax = A(Ax-\lambda_i x) = A\underbrace{(A-\lambda_i I)x}_{0} = 0$$

Hence $y \in N(A-\lambda_i I)$. ∎

Note that $x \in N(A-\lambda_i I)$ implies x (if nonzero) is an eigenvector of A with corresponding eigenvalue λ_i.

<u>Theorem</u> Let $A \in \mathbb{C}^{n\times n}$ be the matrix representation of a linear operator $T: \mathbb{C}^n \to \mathbb{C}^n$ with respect to (this is not essential!) canonical basis

$$\left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right).$$

Suppose

1) $\mathbb{C}^n = M_1 \oplus M_2 \oplus \cdots \oplus M_k$

2) Each subspace M_i is invariant under T.

Let $\dim(M_i) = n_i$ and the columns of the $n \times n_i$ matrix $B_i = [b_1^i \; b_2^i \cdots b_{n_i}^i]$ make a basis for M_i. Then with respect to the basis (for $\mathbb{C}^n$)

$(b_1^1, \dots, b_{n_1}^1; \; \dots \; ; b_1^k, \dots, b_{n_k}^k)$ the transformation T has a block diagonal matrix representation

$$\bar{A} = \begin{bmatrix} \bar{A}_1 & 0 & \cdots & 0 \\ 0 & \bar{A}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{A}_k \end{bmatrix}$$

where $\bar{A}_i \in \mathbb{C}^{n_i \times n_i}$. In particular, $\bar{A} = B^{-1}AB$ where $B \in \mathbb{C}^{n\times n}$ is given by $B = [B_1 \,\vdots\, B_2 \,\vdots\, \cdots \,\vdots\, B_k]$.

<u>Proof</u> Exercise. Hint: see the next example.

Example

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & -5 & -4 \end{bmatrix} \quad \text{Let} \quad M_1 = \text{span}\left\{ \underbrace{\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}}_{b_1^1}, \underbrace{\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}}_{b_2^1} \right\} \;\&\; M_2 = \text{span}\left\{ \underbrace{\begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix}}_{b_1^2} \right\}$$

We have $\mathbb{R}^3 = M_1 \oplus M_2$ (M_1 and M_2 are lin. ind. subspaces)

1) Is M_1 invariant under A ?

$$A b_1^1 = A \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = 0 \cdot b_1^1 - 1 \cdot b_2^1 \in M_1$$

$$A b_2^1 = A \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} = 1 \cdot b_1^1 - 2 \cdot b_2^1 \in M_1$$

$$A \underbrace{\begin{bmatrix} -2 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}}_{B_1} = \begin{bmatrix} -2 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} 0 & 1 \\ -1 & -2 \end{bmatrix}}_{\bar{A}_1}$$

2) Is M_2 invariant under A ?

$$A b_1^2 = A \begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \\ -8 \end{bmatrix} = -2 b_1^2 \in M_2 \implies A \underbrace{\begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix}}_{B_2} = \begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix} \underbrace{\begin{bmatrix} -2 \end{bmatrix}}_{\bar{A}_2}$$

Therefore $$A \underbrace{[B_1, \; B_2]}_{B} = [B_1, \; B_2] \underbrace{\begin{bmatrix} \bar{A}_1 & 0_{2\times1} \\ 0_{1\times2} & \bar{A}_2 \end{bmatrix}}_{\bar{A}}$$

$$\implies \bar{A} = \left[\begin{array}{cc|c} 0 & 1 & 0 \\ -1 & -2 & 0 \\ \hline 0 & 0 & -2 \end{array}\right] = B^{-1} A B = \underbrace{\begin{bmatrix} 2 & 5 & 2 \\ -4 & -8 & -3 \\ 1 & 2 & 1 \end{bmatrix}}_{B^{-1}} \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & -5 & -4 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} -2 & -1 & 1 \\ 1 & 0 & -2 \\ 0 & 1 & 4 \end{bmatrix}}_{B}$$

Let A be an $n \times n$ matrix with n <u>distinct</u> eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Also let $e_1, e_2, \ldots, e_n$ be the eigenvectors corresponding to these eigenvalues, i.e., $Ae_i = \lambda_i e_i$ for $i = 1, 2, \ldots, n$.

<u>Claim</u> The set of eigenvectors $\{e_1, e_2, \ldots, e_n\}$ form a linearly independent set. Moreover, $N(A - \lambda_i I) = \text{span}\{e_i\}$ for $i = 1, 2, \ldots, n$.

<u>Proof</u> [Part I] Consider the equation

$$\sum_{i=1}^{n} \alpha_i e_i = 0 \qquad [\text{lin. independence requires all } \alpha_i = 0]$$

Define the matrix $K_j = [A - \lambda_1 I][A - \lambda_2 I] \cdots [A - \lambda_{j-1} I][A - \lambda_{j+1} I] \cdots [A - \lambda_n I]$

j^{th} term missing!

Note that the matrices $[A - \lambda_k I]$ and $[A - \lambda_m I]$ commute, i.e.,

$$[A - \lambda_k I][A - \lambda_m I] = A^2 - \lambda_m A - \lambda_k A + \lambda_k \lambda_m I = [A - \lambda_m I][A - \lambda_k I]$$

Then

$$\begin{aligned}
0 &= K_j(0) \\
&= K_j\left(\sum_{i=1}^{n} \alpha_i e_i\right) \\
&= \sum_{i=1}^{n} \alpha_i K_j(e_i) \qquad K_j(e_i) = 0 \text{ for all } i \neq j \\
&= \alpha_j K_j(e_j) \\
&= \alpha_j \{ [A - \lambda_1 I] \cdots [A - \lambda_{j-1} I][A - \lambda_{j+1} I] \cdots [A - \lambda_n I] e_j \} \\
&= \alpha_j \{ \underbrace{(\lambda_j - \lambda_1) \cdots (\lambda_j - \lambda_{j-1})(\lambda_j - \lambda_{j+1}) \cdots (\lambda_j - \lambda_n)}_{\neq 0 \text{ since eigenvalues distinct}} e_j \}
\end{aligned}$$

Therefore $\alpha_j = 0$. Since the index j was arbitrary we have to have $\alpha_j = 0 \ \forall j$. Hence the lin. independence of the eigenvectors.

2017

[part II] Clearly $N(A-\lambda_i I) \supset \text{span}\{e_i\}$. How about $N(A-\lambda_i I) \subset \text{span}\{e_i\}$?

Suppose not. Then there exists $f_i \in N(A-\lambda_i I)$ but $f_i \notin \text{span}\{e_i\}$. Then

$$f_i = \sum_{j=1}^{n} \beta_j e_j \qquad \text{for some scalars } \beta_1, \beta_2, \ldots, \beta_n \qquad (\text{why?})$$

$$\Rightarrow f_i - \beta_i e_i = \sum_{j \neq i} \beta_j e_j \qquad (1)$$

$$\Rightarrow \underbrace{[A-\lambda_i I](f_i - \beta_i e_i)}_{=0 \ (\text{why?})} = [A-\lambda_i I] \sum_{j \neq i} \beta_j e_j$$

$$= \sum_{j \neq i} \beta_j [A-\lambda_i I] e_j$$

$$= \sum_{j \neq i} \beta_j \underbrace{(\lambda_j - \lambda_i)}_{\neq 0} e_j$$

$\Rightarrow \beta_j = 0 \quad \forall j \neq i$ because e_j are lin. ind.

Then (1) $\Rightarrow f_i - \beta_i e_i = 0$

$\Rightarrow f_i \in \text{span}\{e_i\}$ which contradicts $f_i \notin \text{span}\{e_i\}$. ☑

— o —

As a result of this claim we can write

$$\mathbb{C}^n = N(A-\lambda_1 I) \oplus N(A-\lambda_2 I) \oplus \cdots \oplus N(A-\lambda_n I)$$

provided that the below condition holds

(C1) $\lambda_i \neq \lambda_j \quad \forall i \neq j$. (n distinct eigenvalues)

Suppose condition (C1) holds.

Choose $V = \begin{bmatrix} | & | & & | \\ e_1 & e_2 & \cdots & e_n \\ | & | & & | \end{bmatrix}_{n\times n}$ i.e. columns of V are eigenvectors (V^{-1} exists)

Then $AV = [Ae_1 \ Ae_2 \ \cdots \ Ae_n]$

$$= [\lambda_1 e_1 \ \lambda_2 e_2 \cdots \lambda_n e_n] = \underbrace{[e_1 \ e_2 \cdots e_n]}_{V} \underbrace{\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}}_{\Lambda}$$

$\Rightarrow \ \Lambda = V^{-1} A V$

Therefore Λ is the block diagonal representation with 1×1 blocks.

In this case we call Λ the <u>diagonal representation</u> of A.

<u>Remark</u> Not all A have diagonal representations because not all $A \in \mathbb{C}^{n\times n}$ have n lin. independent eigenvectors. E.g. $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$.

— o —

Recall that the characteristic polynomial of $A \in \mathbb{C}^{n\times n}$ is

$$d(s) = \det(sI - A) = s^n + \alpha_1 s^{n-1} + \cdots + \alpha_{n-1} s + \alpha_n \quad (n\text{th degree poly.})$$

The roots of $d(s)$ = eigenvalues of A. In general $d(s)$ is of form

$$d(s) = (s-\lambda_1)^{r_1} (s-\lambda_2)^{r_2} \cdots (s-\lambda_\sigma)^{r_\sigma} \quad \text{where}$$

σ : the number of distinct eigenvalues

r_i : the multiplicity of the eigenvalue λ_i

Note that $r_1 + r_2 + \cdots + r_\sigma = n$.

2017

<u>Theorem</u> [Cayley-Hamilton Thm.] Every matrix $A \in \mathbb{C}^{n\times n}$ satisfies its own characteristic equation : $d(A) = 0_{n\times n}$ where $d(s)$ is the characteristic poly. of A.

<u>Example</u> Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$

$$\Rightarrow d(s) = \det(sI - A) = \begin{vmatrix} s-1 & -2 \\ -3 & s-4 \end{vmatrix} = (s-1)(s-4) - 6 = s^2 - 5s - 2$$

$$\Rightarrow d(A) = A^2 - 5A - 2I = \underbrace{\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}}_{\begin{bmatrix} 7 & 10 \\ 15 & 22 \end{bmatrix}} - 5\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} - 2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 7 & 10 \\ 15 & 22 \end{bmatrix} - \begin{bmatrix} 5 & 10 \\ 15 & 20 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

<u>$A^4 = ?$</u>

$$A^2 = 5A + 2I$$

$$A^3 = A \cdot A^2 = 5A^2 + 2A = 5(5A + 2I) + 2A = 27A + 10I$$

$$A^4 = A \cdot A^3 = 27A^2 + 10A = 27(5A + 2I) + 10A = 145A + 54I$$

Hence in general for each $N \geq n$ we can find scalars $\alpha_0, \alpha_1, \dots, \alpha_{n-1}$ such that

$$A^N = \alpha_0 I + \alpha_1 A + \dots + \alpha_{n-1} A^{n-1}$$

<u>$A^{-1} = ?$</u>

$$\left.\begin{aligned} 0 &= A^{-1} \cdot 0 \\ &= A^{-1}(A^2 - 5A - 2I) \\ &= A - 5I - 2A^{-1} \end{aligned}\right\} \begin{aligned} &2A^{-1} = A - 5I \Rightarrow A^{-1} = \tfrac{1}{2}A - \tfrac{5}{2}I \\ &\Rightarrow A^{-1} = \tfrac{1}{2}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} - \tfrac{5}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ \tfrac{3}{2} & -\tfrac{1}{2} \end{bmatrix} \end{aligned}$$

<u>Exercise</u> Recall for the system $\dot{x} = \underset{n\times n}{A}x + \underset{n\times 1}{B}u$ the controllability matrix is defined as $C = [B \;\; AB \; \cdots \; A^{n-1}B]$. Show that

$$\text{range } C = \text{range } [B \;\; AB \; \cdots \; A^{N-1}B] \quad \text{for all } N > n.$$

Proof of Cayley-Hamilton Thm. Initially suppose the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ are distinct. Then we can write

$$A = V \Lambda V^{-1} = V \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ & & \ddots \\ 0 & & & \lambda_n \end{bmatrix} V^{-1}$$ where the columns of V are eigenvectors

Note that $A^k = V \Lambda^k V^{-1}$

$$= V \begin{bmatrix} \lambda_1^k & & 0 \\ & \lambda_2^k & \\ & & \ddots \\ 0 & & & \lambda_n^k \end{bmatrix} V^{-1}$$

ex $A^3 = (V\Lambda V^{-1})(V\Lambda V^{-1})(V\Lambda V^{-1})$

$= V \Lambda I \Lambda I \Lambda V^{-1}$

$= V \Lambda^3 V^{-1}$

This implies that given any poly. $p(s)$ we can write $p(A) = V p(\Lambda) V^{-1}$. Hence

$$d(A) = V d(\Lambda) V^{-1} = V \begin{bmatrix} d(\lambda_1) & & \\ & d(\lambda_2) & \\ & & \ddots \\ & & & d(\lambda_n) \end{bmatrix} V^{-1}$$

ex $p(s) = s^2 + 2s + 3$

$p(A) = A^2 + 2A + 3I$

$= V\Lambda^2 V^{-1} + 2V\Lambda V^{-1} + 3VIV^{-1}$

$= V(\Lambda^2 + 2\Lambda + 3I)V^{-1}$

$= V p(\Lambda) V^{-1}$

Since λ_i are the roots of $d(s)$, $d(\lambda_i) = 0$ for all i.

$$\Rightarrow d(A) = V \begin{bmatrix} 0 \end{bmatrix} V^{-1} = 0_{n \times n}.$$

Consider now the general case. That is, the eigenvalues of A need not be distinct. We use the following fact.

Fact Let $M \in \mathbb{C}^{n \times n}$. For each $\delta > 0$ there exists a matrix $\tilde{M} \in \mathbb{C}^{n \times n}$ with distinct eigenvalues satisfying $\|M - \tilde{M}\| \le \delta$.

E.g. $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow d(s) = (s-a)(s-d) - bc = s^2 - (a+d)s + ad - bc$

$= (s-1)^2 = s^2 - 2s + 1$

perturb M: $\tilde{M} = \begin{bmatrix} a+\varepsilon_1 & b+\varepsilon_2 \\ c+\varepsilon_3 & d+\varepsilon_4 \end{bmatrix}$, $|\varepsilon_i| \ll 1$

$\Rightarrow \tilde{d}(s) = s^2 - (\tilde{a} + \tilde{d})s + \tilde{a}\tilde{d} - \tilde{b}\tilde{c}$

$= s^2 - \tilde{2}s + \tilde{1} = (s - 0.997)(s - 1.00082)$

Let us choose a sequence $\{A_k\}_{k=1}^{\infty}$ satisfying $\|A - A_k\| \leq \frac{1}{k}$ and A_k has distinct eigenvalues. (Such a sequence exists by the above fact.) Let us define $d_k(s) := \det(sI - A_k)$. Now since $\det(\cdot)$ is a continuous function we can write $d(A) = \lim_{k\to\infty} d_k(A_k)$. Note that $d_k(A_k) = 0 \quad \forall k$.

Hence $d(A) = 0$. ▨

— o —

<u>Minimal Polynomial</u>

<u>Definition</u> For an $n\times n$ matrix A, the <u>minimal polynomial</u> $m(s)$ is the monic polynomial (i.e. the highest order term's coefficient is unity, $m(s) = \underline{1}\cdot s^k + d_1 \cdot s^{k-1} + \dots$) with smallest degree such that $m(A) = 0_{n\times n}$.

<u>Ex</u> Let $P \in \mathbb{R}^{n\times n}$ be an orthogonal projection matrix and $P \neq I$, $P \neq 0$.

Then $m(s) = s(s-1)$. Why?

Because $P^2 = P \Rightarrow 0 = P^2 - P = P(P-I) = m(P)$

How about $P = I$ & $P = 0$?

<u>Theorem</u> Given $A \in \mathbb{C}^{n\times n}$, let $m(s)$ be its minimal polynomial. Then

1) $m(s)$ is unique.
2) $m(s)$ divides $d(s)$, i.e., there exists a poly. $q(s)$ such that $d(s) = q(s)\, m(s)$.
3) Every root of $d(s)$ is also a root of $m(s)$.

<u>Proof</u> [1] Suppose not. Let $m_1(s)$ & $m_2(s)$ be two distinct minimal polynomials. Then we can construct $m_3(s) = m_1(s) - m_2(s)$. Note that

$$\deg(m_3) < \deg(m_1) = \deg(m_2) \qquad (1)$$

Also,

$$m_3(A) = \underbrace{m_1(A)}_{0} - \underbrace{m_2(A)}_{0} = 0 \qquad (2)$$

(1) & (2) contradict that m_1 & m_2 are minimal.

 2017

[2] Suppose not. Then $d(s) = q(s)m(s) + r(s)$ with $r(s) \neq 0$ and

$$\deg(r) < \deg(m) \qquad (3)$$

Also,

$$\begin{aligned} 0 &= d(A) \\ &= q(A)\underbrace{m(A)}_{0} + r(A) \\ &= r(A) \qquad (4) \end{aligned}$$

(3) & (4) contradict that $m(s)$ is minimal.

[3] Let $d(s) = (s-\lambda_1)^{r_1}(s-\lambda_2)^{r_2}\cdots(s-\lambda_\sigma)^{r_\sigma}$ with λ_i being the distinct eigenvalues of A. [σ = # of distinct eigenvalues ; $r_1+r_2+\cdots+r_\sigma = n$] Let e_i be an eigenvector associated to λ_i, i.e., $Ae_i = \lambda_i e_i$. Then

$$0_{n\times 1} = 0_{n\times n} e_i = m(A)e_i = m(\lambda_i)\underbrace{e_i}_{\neq 0}$$

$\Rightarrow m(\lambda_i) = 0 \Rightarrow \lambda_i$ is also a root of $m(s)$. ▨

Remark For $d(s) = (s-\lambda_1)^{r_1}(s-\lambda_2)^{r_2}\cdots(s-\lambda_\sigma)^{r_\sigma}$, by the previous thm we have

$$m(s) = (s-\lambda_1)^{m_1}(s-\lambda_2)^{m_2}\cdots(s-\lambda_\sigma)^{m_\sigma} \text{ with } 1 \le m_i \le r_i \text{ for } i = 1,2,\ldots,\sigma.$$

Theorem Let $N_i := N\left([A-\lambda_i I]^{m_i}\right)$. Then $\mathbb{C}^n = N_1 \oplus N_2 \oplus \cdots \oplus N_\sigma$.

Proof Let $m(s)$ be the minimal poly of A. By partial fraction expansion:

$$\frac{1}{m(s)} = \frac{n_1(s)}{(s-\lambda_1)^{m_1}} + \frac{n_2(s)}{(s-\lambda_2)^{m_2}} + \cdots + \frac{n_\sigma(s)}{(s-\lambda_\sigma)^{m_\sigma}}$$

$$\Rightarrow 1 = n_1(s)P_1(s) + n_2(s)P_2(s) + \cdots + n_\sigma(s)P_\sigma(s) \qquad (1)$$

where

$$P_i(s) = (s-\lambda_1)^{m_1}\cdots\underbrace{(s-\lambda_{i-1})^{m_{i-1}}(s-\lambda_{i+1})^{m_{i+1}}}_{i\text{th term missing}}\cdots(s-\lambda_\sigma)^{m_\sigma}$$

Par. frac. exp.

$$\frac{1}{(s-1)(s+1)^2} = \frac{a}{s-1} + \frac{bs+c}{(s+1)^2}$$

$$a = \frac{1}{4}, \quad b = -\frac{1}{4}, \quad c = -\frac{3}{4}$$

(1) $\Rightarrow$ $I = n_1(A)p_1(A) + \cdots + n_\sigma(A)p_\sigma(A)$ (2)

Let $x \in \mathbb{C}^n$ be an arbitrary vector. Then

ex $1 = (s+1)^2 - (s^2+2s)$

$\Rightarrow$ $I = (A+I)^2 - (A^2+2A)$

(2) $\Rightarrow$ $x = Ix$

$$= \underbrace{n_1(A)p_1(A)x}_{x_1} + \cdots + \underbrace{n_\sigma(A)p_\sigma(A)x}_{x_\sigma}$$

Claim $n_i(A)p_i(A)x \in N_i$

Proof $[A-\lambda_i I]^{m_i} n_i(A)p_i(A)x = n_i(A)\underbrace{[A-\lambda_i I]^{m_i}p_i(A)}_{m(A)=0}x = 0$

polynomials of a matrix commute:

$p(A)q(A) = q(A)p(A)$

Hence $x = x_1 + x_2 + \cdots + x_\sigma$ with $x_i = n_i(A)p_i(A)x \in N_i$. Since x was arbitrary we have

$$\mathbb{C}^n = N_1 + N_2 + \cdots + N_\sigma$$

This however is just regular sum. For direct sum we must establish also the lin. independence of $N_1, N_2, \ldots, N_\sigma$. To this end, let

$$y_1 + y_2 + \cdots + y_\sigma = 0 \quad \text{where} \quad y_i \in N_i \quad \text{for} \quad i = 1, 2, \ldots, \sigma$$

$\Rightarrow$ $p_i(A)y_1 + p_i(A)y_2 + \cdots + p_i(A)y_\sigma = 0$

since $p_i(A)y_j = 0$ for $j \neq i$ because y_j is in the null space of one of the non-missing terms.

$\Rightarrow$ $p_i(A)y_i = 0$ (3)

Recall that $m(s) = p_i(s)(s-\lambda_i)^{m_i}$.

$$\Rightarrow \frac{1}{m(s)} = \frac{n(s)}{(s-\lambda_i)^{m_i}} + \frac{q(s)}{p_i(s)} \qquad \text{(partial fraction expansion)}$$

$\times m(s)$

$$\Rightarrow 1 = n(s)p_i(s) + q(s)(s-\lambda_i)^{m_i}$$

$$\Rightarrow I = n(A)p_i(A) + q(A)(A-\lambda_i I)^{m_i}$$

$\times y_i$

$$\Rightarrow y_i = \underbrace{n(A)p_i(A)y_i}_{=0 \text{ by } (3)} + \underbrace{q(A)(A-\lambda_i I)^{m_i} y_i}_{=0 \text{ by } y_i \in N_i}$$

$$\Rightarrow y_i = 0$$

Since index i was arbitrary we have $y_i = 0 \ \forall i$. Hence N_i are lin. ind. ☑

Summary $A \in \mathbb{C}^{n \times n}$

→ $d(s) = \det(sI - A) = (s-\lambda_1)^{r_1}(s-\lambda_2)^{r_2} \cdots (s-\lambda_\sigma)^{r_\sigma}$ char. poly

λ_i = distinct eigenvalues of A

& $r_1 + r_2 + \cdots + r_\sigma = n$

→ Cayley-Hamilton Thm : $d(A) = 0$.

→ Minimal poly. $m(s) = (s-\lambda_1)^{m_1}(s-\lambda_2)^{m_2} \cdots (s-\lambda_\sigma)^{m_\sigma}$ with $1 \le m_i \le r_i$

$m(A) = 0$

→ $\mathbb{C}^n = N_1 \oplus N_2 \oplus \cdots \oplus N_\sigma$ where $N_i = N\left([A - \lambda_i I]^{m_i}\right)$

Theorem $N(A-\lambda_i I) \underset{\neq}{\subset} N([A-\lambda_i I]^2) \underset{\neq}{\subset} \cdots \underset{\neq}{\subset} N([A-\lambda_i I]^{k_i}) = N([A-\lambda_i I]^{k_i+1}) = \cdots$

for some $k_i \geq 1$.

proof Let $x \in N([A-\lambda_i I]^k)$.

$$\Rightarrow [A-\lambda_i I]^k x = 0$$

$$\Rightarrow [A-\lambda_i I]^{k+1} x = [A-\lambda_i I]\underbrace{[A-\lambda_i I]^k x}_{0} = 0$$

$$\Rightarrow x \in N([A-\lambda_i I]^{k+1})$$

Hence $N([A-\lambda_i I]^k) \subset N([A-\lambda_i I]^{k+1})$.

Now, if $N([A-\lambda_i I]^k) \neq N([A-\lambda_i I]^{k+1})$ then $\dim N([A-\lambda_i I]^k) < \dim N([A-\lambda_i I]^{k+1})$

Also $N([A-\lambda_i I]^k) \subset \mathbb{C}^n \Rightarrow \dim N([A-\lambda_i I]^k) \leq n$ for all k.

Therefore for some k $(k = k_i)$ we have to have

$$N([A-\lambda_i I]^k) = N([A-\lambda_i I]^{k+1}) \qquad (1)$$

Once (1) is satisfied the dimension cannot increase further because

$$x \in N([A-\lambda_i I]^{k+2}) \Leftrightarrow [A-\lambda_i I]^{k+2} x = 0$$

$$\Leftrightarrow [A-\lambda_i I]^{k+1} [A-\lambda_i I] x = 0$$

$$\Leftrightarrow [A-\lambda_i I]^{k} [A-\lambda_i I] x = 0 \qquad (1)$$

$$\Leftrightarrow [A-\lambda_i I]^{k+1} x = 0$$

$$\Leftrightarrow x \in N([A-\lambda_i I]^{k+1})$$

That is, $(1) \Rightarrow N([A-\lambda_i I]^{k+1}) = N([A-\lambda_i I]^{k+2})$.

The result follows by induction. ∎

Theorem $k_i = m_i$.

Proof Let $M_i := [A - \lambda_i I]$. We must show two things:

1) $N(M_i^{m_i - 1}) \neq N(M_i^{m_i})$

2) $N(M_i^{m_i}) = N(M_i^{m_i+1})$

part I Recall that minimal poly. $m(s) = \prod_{i=1}^{\sigma} (s-\lambda_i)^{m_i}$

$$\Rightarrow m(A) = M_1^{m_1} M_2^{m_2} \cdots M_\sigma^{m_\sigma} = 0_{n \times n}$$

$$\& \quad M_1^{m_1} \cdots M_{i-1}^{m(i-1)} \underbrace{[M_i^{m_i - 1}]}_{\text{instead of } M_i^{m_i}} M_{i+1}^{m(i+1)} \cdots M_\sigma^{m_\sigma} \neq 0 \quad \text{(because } m(s) \text{ is minimal)}$$

Then there exists $y \in \mathbb{C}^n$ such that

$$M_1^{m_1} \cdots M_{i-1}^{m(i-1)} [M_i^{m_i - 1}] M_{i+1}^{m(i+1)} \cdots M_\sigma^{m_\sigma} y \neq 0 \quad \text{[Note that the terms commute]}$$

Let now $x := M_1^{m_1} \cdots M_{i-1}^{m(i-1)} \underbrace{\quad}_{i^{th} \text{ term missing}} M_{i+1}^{m(i+1)} \cdots M_\sigma^{m_\sigma} y$

Note that $M_i^{m_i - 1} x \neq 0$ and $M_i^{m_i} x = 0$. Hence part I is proven.

part II Partial fraction expansion:

$$\frac{1}{m(s)} = \frac{n(s)}{(s-\lambda_i)^{m_i}} + \frac{q(s)}{p_i(s)} \quad \text{where } p_i(s) = \prod_{j \neq i} (s - \lambda_j)^{m_j}$$

$\times\ m(s)(s-\lambda_i)^{m_i}$

$$\Rightarrow (s-\lambda_i)^{m_i} = n(s) m(s) + q(s) \underbrace{(s-\lambda_i)^{2m_i}}_{= (s-\lambda_i)^{m_i - 1} (s-\lambda_i)^{m_i+1}}$$

$$\Rightarrow [A - \lambda_i I]^{m_i} = n(A) \underbrace{m(A)}_{0} + q(A) [A - \lambda_i I]^{m_i - 1} [A - \lambda_i I]^{m_i + 1}$$

$$\Rightarrow [A - \lambda_i I]^{m_i} = q(A) [A - \lambda_i I]^{m_i - 1} [A - \lambda_i I]^{m_i+1} \quad (1)$$

Let $x \in N(M_i^{m_i+1})$. Then (1) allows us to write

$$M_i^{m_i} x = q(A)\, M_i^{m_i-1} \underbrace{M_i^{m_i+1} x}_{0} = 0 \implies x \in N(M_i^{m_i})$$

Therefore $N(M_i^{m_i}) \supset N(M_i^{m_i+1})$

we also have $N(M_i^{m_i}) \subset N(M_i^{m_i+1})$

$\Bigg\} \quad N(M_i^{m_i}) = N(M_i^{m_i+1})$ ☒

<u>Theorem</u> $\dim N([A-\lambda_i I]^{m_i}) = r_i$.

<u>Proof</u> Recall $\mathbb{C}^n = N_1 \oplus N_2 \oplus \cdots \oplus N_\sigma$ with $N_i = N([A-\lambda_i I]^{m_i})$. Since each subspace N_i is invariant under A (why?) we can find $P \in \mathbb{C}^{n\times n}$ such that $P = [P_1 \; P_2 \cdots P_\sigma]$ and the linearly independent columns of $P_i \in \mathbb{C}^{n\times k_i}$ span N_i and $AP_i = P_i \bar{A}_i$ with $\bar{A}_i \in \mathbb{C}^{k_i \times k_i}$. Note that $k_i = \dim N_i$. Therefore we want to show $k_i = r_i$. We can write

$$P^{-1}AP = \begin{bmatrix} \bar{A}_1 & 0 & \cdots & 0 \\ 0 & \bar{A}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{A}_\sigma \end{bmatrix} =: \bar{A}$$

$\rightarrow$ block diagonal matrix

Note that
$$\bar{d}(s) = \det(sI - \bar{A}) = \det(sI - P^{-1}AP)$$
$$= \det(P^{-1}[sI - A]P)$$
$$= \underbrace{\det(P^{-1})\det(P)}_{1}\det(sI-A) = d(s)$$

Hence $\bar{d}(s) = d(s)$ ($\bar{A}$ & A have the same char. poly.)

2017

Since $\bar{A}$ is block diagonal we have $\bar{d}(s) = \det(sI - \bar{A}_1)\det(sI - \bar{A}_2) \cdots \det(sI - \bar{A}_\sigma)$

Let μ be an eigenvalue of $\bar{A}_i$ with eigenvector $x \in \mathbb{C}^{k_i}$, i.e., $\bar{A}_i x = \mu x$. Now, consider the vector $v = P_i x$. Note that $v \neq 0$ (why?) and $v \in N_i$ (why?) Then

$$\begin{aligned} Av &= AP_i x \\ &= P_i \bar{A}_i x \\ &= \mu P_i x \\ &= \mu v \end{aligned}$$

Moreover,
$$\begin{aligned} 0 &= [A - \lambda_i I]^{m_i} v \\ &= [A - \lambda_i I]^{m_i - 1}[A - \lambda_i I] v \\ &= (\mu - \lambda_i)[A - \lambda_i I]^{m_i - 1} v \\ &\vdots \\ &= (\mu - \lambda_i)^{m_i} v \qquad (1) \end{aligned}$$

Since $v \neq 0$ eq. (1) implies $\mu = \lambda_i$, i.e., λ_i is the only distinct eigenvalue of $\bar{A}_i$.

Therefore $\det(sI - \bar{A}_i) = (s - \lambda_i)^{k_i}$. We can write

$$\bar{d}(s) = (s-\lambda_1)^{k_1}(s-\lambda_2)^{k_2} \cdots (s-\lambda_\sigma)^{k_\sigma} \quad \&$$

$$d(s) = (s-\lambda_1)^{r_1}(s-\lambda_2)^{r_2} \cdots (s-\lambda_\sigma)^{r_\sigma}$$

Since $\lambda_i \neq \lambda_j$ & $d(s) = \bar{d}(s)$ we have to have $k_i = r_i$. ▨

Summary of Spectral Theory

Given $A \in \mathbb{C}^{n \times n}$

→ char. poly. $d(s) = (s-\lambda_1)^{r_1}(s-\lambda_2)^{r_2}\cdots(s-\lambda_\sigma)^{r_\sigma}$, $\lambda_i \neq \lambda_j$, $\sum r_i = n$, $d(A) = 0$

→ min poly. $m(s) = (s-\lambda_1)^{m_1}(s-\lambda_2)^{m_2}\cdots(s-\lambda_\sigma)^{m_\sigma}$, $1 \leq m_i \leq r_i$, $m(A) = 0$

→ $$\mathbb{C}^n = \underbrace{N([A-\lambda_1 I]^{m_1})}_{\dim = r_1} \oplus \underbrace{N([A-\lambda_2 I]^{m_2})}_{\dim = r_2} \oplus \cdots \oplus \underbrace{N([A-\lambda_\sigma I]^{m_\sigma})}_{\dim = r_\sigma}$$

→ $$N([A-\lambda_i I]) \underset{\neq}{\subset} N([A-\lambda_i I]^2) \underset{\neq}{\subset} \cdots \underset{\neq}{\subset} N([A-\lambda_i I]^{m_i}) = N([A-\lambda_i I]^{m_i+1}) = \cdots$$

2017

Computational Aspects

Recall: Given $A \in \mathbb{C}^{n\times n}$

→ char. poly. $d(s) = (s-\lambda_1)^{r_1}(s-\lambda_2)^{r_2}\cdots(s-\lambda_\sigma)^{r_\sigma}$, $\lambda_i \neq \lambda_j$

→ min. poly. $m(s) = (s-\lambda_1)^{m_1}(s-\lambda_2)^{m_2}\cdots(s-\lambda_\sigma)^{m_\sigma}$, $1 \le m_i \le r_i$, $m(A)=0$

→ $\mathbb{C}^n = N_1 \oplus N_2 \oplus \cdots \oplus N_\sigma$ where $N_i = N([A-\lambda_i I]^{m_i})$ and $\dim N_i = r_i$

→ $N([A-\lambda_i I]) \underset{\neq}{\subset} N([A-\lambda_i I]^2) \underset{\neq}{\subset} \cdots \underset{\neq}{\subset} N([A-\lambda_i I]^{m_i}) = N([A-\lambda_i I]^{m_i+1}) = \cdots$

Problem: Find a coordinate change matrix $P \in \mathbb{C}^{n\times n}$ such that $P = [P_1 \,\vdots\, P_2 \,\vdots\, \cdots \,\vdots\, P_\sigma]$ where r_i columns of P_i span N_i. Therefore

$$P^{-1}AP = \begin{bmatrix} \bar{A}_1 & 0 & \cdots & 0 \\ 0 & \bar{A}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{A}_\sigma \end{bmatrix} .$$

In particular $\underbrace{AP_i}_{n\times r_i} = P_i \underbrace{\bar{A}_i}_{r_i\times r_i}$

Moreover we want the blocks $\bar{A}_i$ be as close to diagonal as possible.

Soll'n: Constructing P_i via Jordan chains

Example: Given $A \in \mathbb{C}^{n\times n}$ let λ_i be the ith eigenvalue. Let $r_i = 12$ and $m_i = 4$.

Furthermore, let $\dim N([A-\lambda_i I]) = 4 \Rightarrow$ number of chains

$\dim N([A-\lambda_i I]^2) = 7$

$\dim N([A-\lambda_i I]^3) = 10$

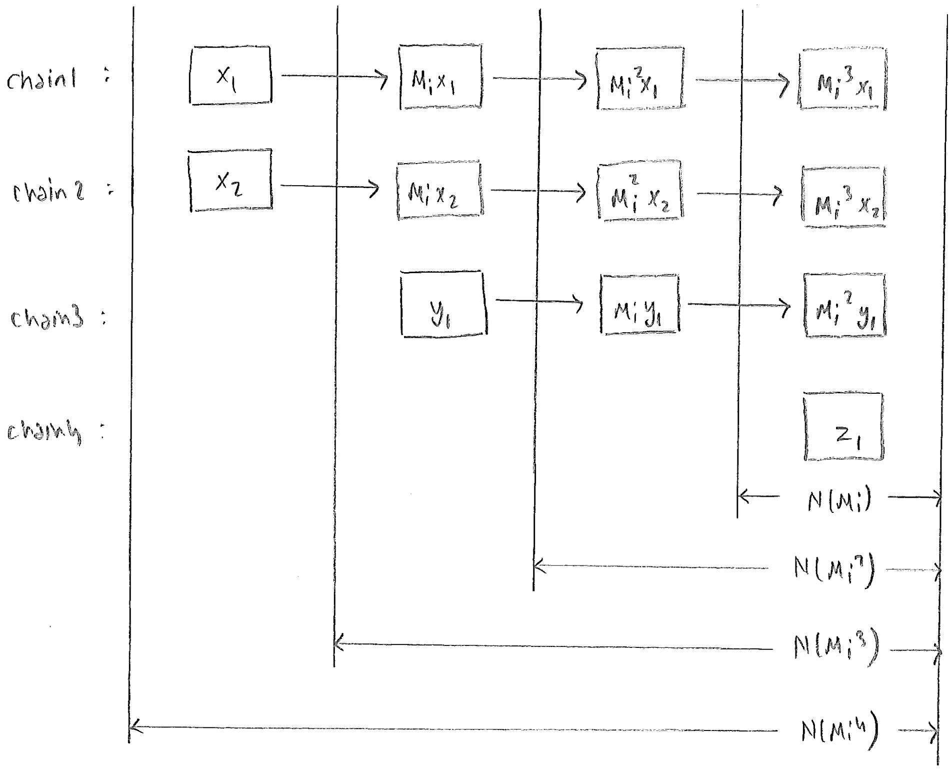
How to construct the chains? Let us switch to shorthand notation $M_i = [A-\lambda_i I]$.

$$\underbrace{N(M_i)}_{\dim = 4} \subset \underbrace{N(M_i^2)}_{\dim=7} \subset \underbrace{N(M_i^3)}_{\dim=10} \subset \underbrace{N(M_i^4)}_{\dim=12} = N(M_i^5) = \cdots$$

$$\xrightarrow{\quad} \qquad \xrightarrow{\quad} \qquad \xrightarrow{\quad}$$

$$3 \qquad\qquad 3 \qquad\qquad 2$$

2017



How to find the "chain starters" x_1, x_2, y_1, z_1 ?

Step1 (Determining x_1 and x_2) Choose some basis $\{v_1, v_2, \cdots, v_{10}\}$ for $N(M_i^3)$. Complete this basis to a basis $\{v_1, v_2, \cdots, v_{10}, x_1, x_2\}$ for $N(M_i^4)$. Then we have

chain 1: $\{x_1, M_i x_1, M_i^2 x_1, M_i^3 x_1\}$

chain 2: $\{x_2, M_i x_2, M_i^2 x_2, M_i^3 x_2\}$

Claim The set $\{x_1, M_i x_1, M_i^2 x_1, M_i^3 x_1, x_2, M_i x_2, M_i^2 x_2, M_i^3 x_2\}$ is lin. ind.

proof Consider $\alpha_0 x_1 + \alpha_1 M_i x_1 + \cdots + \alpha_3 M_i^3 x_1 + \beta_0 x_2 + \cdots + \beta_3 M_i^3 x_2 = 0$

$$\Rightarrow \alpha_0 x_1 + \beta_0 x_2 + M_i(\alpha_1 x_1 + \beta_1 x_2) + \cdots + M_i^3(\alpha_3 x_1 + \beta_3 x_2) = 0 \qquad (1)$$

Multiply (1) by M_i^3:

$$M_i^3(\alpha_0 x_1 + \beta_0 x_2) + \overbrace{M_i^4 \underbrace{(\alpha_1 x_1 + \beta_1 x_2)}_{\in N(M_i^4)}}^{0} + \cdots + \overbrace{M_i^2 M_i^4 \underbrace{(\alpha_3 x_1 + \beta_3 x_2)}_{\in N(M_i^4)}}^{0} = 0$$

$$\Rightarrow M_i^3(\alpha_0 x_1 + \beta_0 x_2) = 0 \Rightarrow \alpha_0 x_1 + \beta_0 x_2 \in N(M_i^3) \Rightarrow -(\alpha_0 x_1 + \beta_0 x_2) \in N(M_i^3)$$

Then we can find $\gamma_1, \gamma_2, \dots, \gamma_{10}$ such that

$$-(\alpha_0 x_1 + \beta_0 x_2) = \sum_{k=1}^{10} \gamma_k v_k \Rightarrow \alpha_0 x_1 + \beta_0 x_2 + \sum_{k=1}^{10} \gamma_k v_k = 0$$

$$\Rightarrow \alpha_0, \beta_0 = 0 \text{ because } \{v_1, \dots, v_{10}, x_1, x_2\} \text{ is lin. ind.}$$

Now, (1) reduces to

$$M_i(\alpha_1 x_1 + \beta_1 x_2) + \dots + M_i^3(\alpha_3 x_1 + \beta_3 x_2) = 0 \qquad (2)$$

Multiply (2) by M_i^2 and repeat the previous procedure. Eventually we establish $\alpha_0, \dots, \alpha_3, \beta_0, \dots, \beta_3$ are all zero. ☒

Step 2 (Determining y_1) Choose some basis $\{w_1, w_2, \dots, w_7\}$ for $N(M_i^2)$. Complete this basis to a basis $\{w_1, \dots, w_7, M_i x_1, M_i x_2, y_1\}$ for $N(M_i^3)$. Then we have the third chain:

Chain 3 : $\{y_1, M_i y_1, M_i^2 y_1\}$

Claim The set $\{\underbrace{x_1, \dots, M_i^3 x_1}_{\text{Chain 1}}, \underbrace{x_2, \dots, M_i^3 x_2}_{\text{Chain 2}}, \underbrace{y_1, \dots, M_i^2 y_1}_{\text{Chain 3}}\}$ is lin. ind.

Proof Exercise.

Step 3 (Determining z_1) Choose z_1 such that $\{M_i^3 x_1, M_i^3 x_2, M_i^2 y_1, z_1\}$ is a basis for $N(M_i)$. The last chain then is:

Chain 4: $\{z_1\}$

Now our basis for $N(M_i^4)$ is complete:

$$\{\underbrace{x_1, \dots, M_i^3 x_1}_{\text{Chain 1}}, \underbrace{x_2, \dots, M_i^3 x_2}_{\text{Chain 2}}, \underbrace{y_1, \dots, M_i^2 y_1}_{\text{Chain 3}}, \underbrace{z_1}_{\text{Chain 4}}\}$$

Let us now construct P_i as

$$P_i = \begin{bmatrix} | & & | & | & & | & | & & | & | \\ M_i^3 x_1 & \cdots & x_1 & M_i^3 x_2 & \cdots & x_2 & M_i^2 y_1 & \cdots & y_1 & z_1 \\ | & & | & | & & | & | & & | & | \end{bmatrix}$$

<u>$AP_i = ?$</u>

$$[A - \lambda_i I]\, P_i = \Big[\, \underbrace{M_i^4 x_1}_{0} \cdots M_i x_1 \;\; \underbrace{M_i^4 x_2}_{0} \cdots M_i x_2 \;\; \underbrace{M_i^3 y_1}_{0} \cdots M_i y_1 \;\; \underbrace{M_i z_1}_{0} \,\Big]$$

$$= \begin{array}{c} \begin{array}{cccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \end{array} \\ \begin{bmatrix} 0 & M_i^3 x_1 & M_i^2 x_1 & M_i x_1 & 0 & M_i^3 x_2 & M_i^2 x_2 & M_i x_2 & 0 & M_i^2 y_1 & M_i y_1 & 0 \end{bmatrix} \end{array}$$

$$= P_i \underbrace{\begin{array}{c|cccccccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \hline 1 & 0 & 1 & 0 & 0 & & & & & & & & \\ 2 & 0 & 0 & 1 & 0 & & & & & & & & \\ 3 & 0 & 0 & 0 & 1 & & & & & & & & \\ 4 & 0 & 0 & 0 & 0 & & & & & & & & \\ 5 & & & & & 0 & 1 & 0 & 0 & & & & \\ 6 & & & & & 0 & 0 & 1 & 0 & & & & \\ 7 & & & & & 0 & 0 & 0 & 1 & & & & \\ 8 & & & & & 0 & 0 & 0 & 0 & & & & \\ 9 & & & & & & & & & 0 & 1 & 0 & \\ 10 & & & & & & & & & 0 & 0 & 1 & \\ 11 & & & & & & & & & 0 & 0 & 0 & \\ 12 & & & & & & & & & & & & 0 \end{array}}_{\bar{X}_i}$$

(All empty boxes are zero.)

$$\Rightarrow \; AP_i = P_i \bar{X}_i + \lambda_i P_i = P_i \underbrace{(\bar{X}_i + \lambda_i I)}_{\bar{A}_i}$$

where

$$\bar{A}_i = \begin{bmatrix} J_{i,1} & & & 0 \\ & J_{i,2} & & \\ & & J_{i,3} & \\ 0 & & & J_{i,4} \end{bmatrix}_{r_i \times r_i}$$

with

$$J_{i,1} = J_{i,2} = \begin{bmatrix} \lambda_i & 1 & 0 & 0 \\ 0 & \lambda_i & 1 & 0 \\ 0 & 0 & \lambda_i & 1 \\ 0 & 0 & 0 & \lambda_i \end{bmatrix}_{m_i \times m_i}, \quad J_{i,3} = \begin{bmatrix} \lambda_i & 1 & 0 \\ 0 & \lambda_i & 1 \\ 0 & 0 & \lambda_i \end{bmatrix}, \quad J_{i,4} = [\lambda_i]$$

<u>Remark</u> $J_{i,k}$ is called a Jordan block. The largest Jordan block associated with the eigenvalue λ_i is of size $m_i \times m_i$.

Example (special case 1) Let A have a single eigenvalue λ_1 and $m_1 = r_1$.

Then $A \in \mathbb{C}^{r_1 \times r_1}$ and $m(s) = d(s) = (s-\lambda_1)^{r_1}$. [Recall $M_i = A - \lambda_i I$]

Choose $x \in N(M_1^{m_1})$ but $x \notin N(M_1^{m_1 - 1})$. [Note that $N(M_1^{m_1}) = \mathbb{C}^{r_1}$ (why?)]

The vector x yields the (only) chain

$$\{x, M_1 x, \ldots, M_1^{m_1-1} x\}$$

which has m_1 lin. ind. vectors and therefore is a basis for $N(M_1^{m_1}) = \mathbb{C}^{r_1}$.

Now, let us construct P as

$$P = \begin{bmatrix} M_1^{m_1-1}x & \cdots & M_1 x & x \end{bmatrix}_{r_1 \times r_1}$$

Then

$$[A - \lambda_1 I] P = [0 \;\; M_1^{m_1-1}x \;\; \cdots \;\; M_1 x] = P \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$\Rightarrow AP = P \underbrace{\begin{bmatrix} \lambda_1 & 1 & 0 & \cdots & 0 \\ 0 & \lambda_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & \lambda_1 \end{bmatrix}}_{\bar{A}_1} \qquad \Rightarrow P^{-1} A P = \bar{A}_1$$

Note that

$$N(M_1) \subset N(M_1^2) \subset \cdots \subset N(M_1^{m_1})$$

$$\begin{matrix} \underset{\dim = 1}{\operatorname{span}\{M_1^{m_1-1}x\}} & \underset{\dim = 2}{\operatorname{span}\{M_1^{m_1-1}x, M_1^{m_1-2}x\}} & \underset{\dim = m_1 = r_1}{\operatorname{span}\{M_1^{m_1-1}x, M_1^{m_1-2}x, \ldots, x\}} \end{matrix}$$

Since $\dim N(M_1) = \dim([A - \lambda_1 I]) = 1$ there is only one (lin. ind.) eigenvector, which is $v = M_1^{m_1-1}x$. Observe

$$[A-\lambda_1 I] v = [A - \lambda_1 I] M_1^{m_1-1} x$$

$$= \underbrace{[A - \lambda_1 I]^{m_1}}_{m(A) = 0} x = 0$$

Example (special case 2) Let A have a single eigenvalue λ_1 and $m_1 = 1$.

Then $m(s) = s - \lambda_1$. This yields

$$0 = m(A) = A - \lambda_1 I \implies A = \lambda_1 I = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_1 & \\ 0 & & \ddots \; \lambda_1 \end{bmatrix}$$

Note that for any nonsingular P $\quad P^{-1}AP = P^{-1}(\lambda_1 I)P = \lambda_1 P^{-1}P = \lambda_1 I = A$.

Therefore the matrix representation is unique regardless of basis. Also,

any (nonzero) vector x is an eigenvector of A since $Ax = \lambda_1 I x = \lambda_1 x$.

Example

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 1 & 0 & 2 \end{bmatrix} \qquad \det(sI - A) = (s-1)^3(s-2)$$

$$\implies \begin{cases} \lambda_1 = 1, \; r_1 = 3, \; m_1 = ? \\ \lambda_2 = 2, \; r_2 = 1, \; m_2 = 1 \end{cases}$$

Step1 Find m_1. Note that m_1 is the smallest integer satisfying $N(M_1^{m_1}) = 3$

$$M_1 = A - \lambda_1 I = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 1 \end{bmatrix}.$$

M_1 has two lin. ind. columns. Hence $\dim R(M_1) = 2$

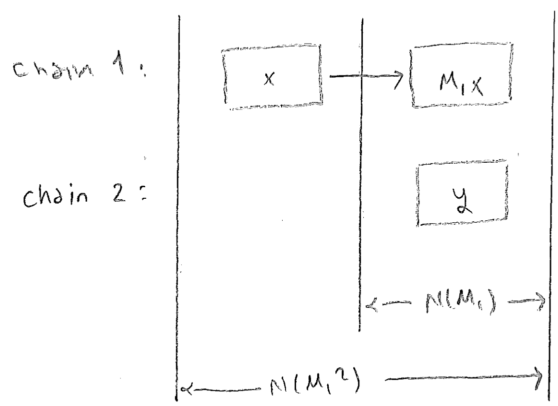
Then $\dim N(M_1) = 4 - \dim R(M_1) = 2$

$$M_1^2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \implies \dim N(M_1^2) = 3 \implies m_1 = 2$$

Step2 Construct P_1.

Q: Number of chains?

A: $\dim N(M_1) = 2$



$\underline{x = ?}$ Choose, for instance, $x = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$. Note that $M_1^2 x = 0$ but $M_1 x = \begin{bmatrix}1\\0\\0\\1\end{bmatrix} \neq 0$

Hence $x \in N(M_1^2)$ & $x \notin N(M_1)$.

$\Rightarrow$ chain 1 $= \{x, M_1 x\}$

$\underline{y = ?}$ Choose, for instance, $y = \begin{bmatrix}0\\0\\2\\0\end{bmatrix}$. Note that $M_1 y = 0$ & $\{y, M_1 x\}$ is lin. ind.

$\Rightarrow$ chain 2 $= \{y\}$

We can now construct $P_1 = [M_1 x \;\; x \;\; y] = \begin{bmatrix}1 & 1 & 0\\0 & 1 & 0\\0 & 1 & 2\\1 & 1 & 0\end{bmatrix}$

<u>Step 3</u> Construct P_2.

$\underbrace{N(M_2)}_{\dim = 1} = N(M_2^{m_2})$. Hence there is only one chain with a single element.

$$M_2 = A - \lambda_2 I = \begin{bmatrix}-1 & 1 & 0 & 0\\0 & -1 & 0 & 0\\0 & 0 & -1 & 0\\-1 & 1 & 0 & 0\end{bmatrix}$$

Choose, for instance, $z = \begin{bmatrix}0\\0\\0\\3\end{bmatrix}$. Then $M_2 z = 0 \Rightarrow z \in N(M_2)$.

$\Rightarrow$ chain 3 $= \{z\}$

This yields $P_2 = [z] = \begin{bmatrix}0\\0\\0\\3\end{bmatrix}$.

Finally, we have $P = [P_1 \;\; P_2] = \begin{bmatrix}1 & 1 & 0 & 0\\0 & 1 & 0 & 0\\0 & 1 & 2 & 0\\1 & 1 & 0 & 3\end{bmatrix}$

$$\Rightarrow P^{-1} A P = \begin{bmatrix}\boxed{\begin{matrix}1 & 1\\0 & 1\end{matrix}} & \begin{matrix}0 & 0\\0 & 0\end{matrix}\\ \begin{matrix}0 & 0\\0 & 0\end{matrix} & \begin{matrix}\boxed{1} & 0\\0 & \boxed{2}\end{matrix}\end{bmatrix}$$

$\begin{bmatrix}\lambda_1 & 1\\0 & \lambda_1\end{bmatrix}$: due to chain $\{x, M_1 x\}$

$[\lambda_1]$: due to chain $\{y\}$

$[\lambda_2]$: due to chain $\{z\}$

2017

Let us summarize our findings in the following theorem.

Theorem Let $A \in \mathbb{C}^{n \times n}$ have

$$d(s) = (s-\lambda_1)^{r_1}(s-\lambda_2)^{r_2}\cdots(s-\lambda_\sigma)^{r_\sigma} \quad \text{(char. poly.) and}$$

$$m(s) = (s-\lambda_1)^{m_1}(s-\lambda_2)^{m_2}\cdots(s-\lambda_\sigma)^{m_\sigma} \quad \text{(min. poly.)}$$

Then we can find $P \in \mathbb{C}^{n \times n}$ such that

$$P^{-1}AP = \bar{A} = \begin{bmatrix} \bar{A}_1 & 0 & \cdots & 0 \\ 0 & \bar{A}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{A}_\sigma \end{bmatrix} \quad \text{where} \quad \bar{A}_i = \begin{bmatrix} J_{i,1} & 0 & \cdots & 0 \\ 0 & J_{i,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & \end{bmatrix}_{r_i \times r_i}$$

where (Jordan block) $$J_{i,k} = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & \lambda_i \end{bmatrix}$$

where the largest Jordan block associated with the eigenvalue λ_i is of size $m_i \times m_i$. The form $\bar{A} = P^{-1}AP$ is called the <u>Jordan Canonical Form.</u>

— o —

Exercise Let $A \in \mathbb{C}^{4 \times 4}$. For each of the below cases write the Jordan Canonical form of the matrix A.

a) $m(s) = s$

b) $m(s) = s^4$

c) $N(A) = R(A)$

Hermitian Matrices

Definition A square matrix A is said to be hermitian if $A^* = A$, i.e., its conjugate transpose equals itself.

Ex

$$A = \begin{bmatrix} 1 & 3+j \\ 3-j & -5 \end{bmatrix}$$

For real matrices $A \in \mathbb{R}^{n\times n}$ being hermitian is equivalent to being symmetric i.e. $A = A^T$. Hermitian matrices enjoy some important properties.

[Note: Henceforth by the inner product $\langle \cdot, \cdot \rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ we mean the standard scalar product $\langle x, y \rangle = y^* x$. Observe that for $A \in \mathbb{C}^{n\times n}$ we have $\langle Ax, y \rangle = y^* A x = (A^* y)^* x = \langle x, A^* y \rangle$.]

Theorem Let A be hermitian. Then $\langle x, Ax \rangle$ is real for all $x \in \mathbb{C}^n$.

proof $\langle x, Ax \rangle = \overline{\langle Ax, x \rangle} = \overline{\langle x, A^* x \rangle} = \overline{\langle x, Ax \rangle}$. (1)

Theorem All eigenvalues of an hermitian matrix are real.

proof Let A be hermitian, λ an eigenvalue of A, and x the corresponding eigenvector, i.e., $Ax = \lambda x$. Then:

$$\begin{aligned}
\bar{\lambda} \langle x, x \rangle &= \langle x, \lambda x \rangle \\
&= \langle x, Ax \rangle \\
&= \overline{\langle x, Ax \rangle} \quad (1) \\
&= \langle Ax, x \rangle \\
&= \langle \lambda x, x \rangle \\
&= \lambda \langle x, x \rangle
\end{aligned}$$

Therefore $\bar{\lambda} = \lambda$. ◻

<u>Theorem</u> Let A be hermitian and λ_i, λ_j be two distinct $(\lambda_i \neq \lambda_j)$ eigenvalues with eigenvectors e_i, e_j. Then $\langle e_i, e_j \rangle = 0$, i.e., e_i and e_j are orthogonal.

<u>proof</u>
$$
\begin{aligned}
0 &= \langle e_i, Ae_j \rangle - \langle e_i, Ae_j \rangle \\
&= \langle e_i, Ae_j \rangle - \langle A^* e_i, e_j \rangle \\
&= \langle e_i, Ae_j \rangle - \langle Ae_i, e_j \rangle \\
&= \langle e_i, \lambda_j e_j \rangle - \langle \lambda_i e_i, e_j \rangle \\
&= \bar{\lambda}_j \langle e_i, e_j \rangle - \lambda_i \langle e_i, e_j \rangle \qquad \lambda_j \text{ is real.} \\
&= \lambda_j \langle e_i, e_j \rangle - \lambda_i \langle e_i, e_j \rangle \\
&= \underbrace{(\lambda_j - \lambda_i)}_{\neq 0} \langle e_i, e_j \rangle
\end{aligned}
$$

Hence $\langle e_i, e_j \rangle = 0$. ▨

<u>Theorem</u> Let $A \in \mathbb{C}^{n \times n}$ be hermitian. Then its minimal polynomial reads $m(s) = (s-\lambda_1)(s-\lambda_2)\cdots(s-\lambda_\sigma)$. That is, $m_i = 1$ for all $i = 1, 2, \dots, \sigma$.

<u>proof</u> Previously we have established

$$
N([A-\lambda_i I]) \underset{\neq}{\subset} N([A-\lambda_i I]^2) \underset{\neq}{\subset} \cdots \underset{\neq}{\subset} N([A-\lambda_i I]^{m_i}) = N([A-\lambda_i I]^{m_i+1}) = \cdots
$$

Hence, to show $m_i = 1$ we need to prove $N([A-\lambda_i I]) = N([A-\lambda_i I]^2)$. This we can do in two steps:

<u>Step 1</u> Show $N([A-\lambda_i I]) \subset N([A-\lambda_i I]^2)$. (This is trivial.)

<u>Step 2</u> Show $N([A-\lambda_i I]) \supset N([A-\lambda_i I]^2)$. Let $x \in N([A-\lambda_i I]^2)$. Then

$$
\begin{aligned}
0 &= \langle x, [A-\lambda_i I]^2 x \rangle \\
&= \langle [A-\lambda_i I]^* x, [A-\lambda_i I] x \rangle \qquad \lambda_i \in \mathbb{R} \\
&= \langle [A-\lambda_i I] x, [A-\lambda_i I] x \rangle \\
&= \| [A-\lambda_i I] x \|^2
\end{aligned}
$$

Hence $x \in N([A-\lambda_i I])$. ▨

2017

The previous theorems allow us to deduce the following. For a hermitian $A \in \mathbb{C}^{n\times n}$ with char. poly. $d(s) = (s-\lambda_1)^{r_1}(s-\lambda_2)^{r_2}\cdots(s-\lambda_\sigma)^{r_\sigma}$ we have

$$\mathbb{C}^n = \underbrace{N([A-\lambda_1 I])}_{\dim = r_1} \overset{\perp}{\oplus} \underbrace{N([A-\lambda_2 I])}_{\dim = r_2} \overset{\perp}{\oplus} \cdots \overset{\perp}{\oplus} \underbrace{N([A-\lambda_\sigma I])}_{\dim = r_\sigma}$$

Whence the below result follows.

Theorem Let $A \in \mathbb{C}^{n\times n}$ be hermitian with $d(s) = (s-\lambda_1)^{r_1}(s-\lambda_2)^{r_2}\cdots(s-\lambda_\sigma)^{r_\sigma}$.

Then there exists a unitary matrix $P \in \mathbb{C}^{n\times n}$ (i.e. $P^{-1} = P^*$) such that

$$P^*AP = \Lambda \quad \text{where} \quad \Lambda = \begin{bmatrix} \Lambda_1 & & & 0 \\ & \Lambda_2 & & \\ & & \ddots & \\ 0 & & & \Lambda_\sigma \end{bmatrix}_{n\times n} \quad \text{with} \quad \Lambda_i = \begin{bmatrix} \lambda_i & & & 0 \\ & \lambda_i & & \\ & & \ddots & \\ 0 & & & \lambda_i \end{bmatrix}_{r_i \times r_i}$$

Proof Exercise. [Hint: see the below example.]

Example Let $A \in \mathbb{C}^{6\times 6}$ be hermitian with $d(s) = (s-\lambda_1)(s-\lambda_2)^2(s-\lambda_3)^3$. Choose $\{x_1\}, \{x_2, x_3\}, \{x_4, x_5, x_6\}$ as orthonormal bases for the subspaces $N([A-\lambda_1 I])$, $N([A-\lambda_2 I])$, and $N([A-\lambda_3 I])$, respectively. Note that each x_i is an eigenvector (why?) and that $\langle x_i, x_j\rangle = 0$ for $i \neq j$ (why?). Also, $\langle x_i, x_i \rangle = 1$. Define $P = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_6 \\ | & | & & | \end{bmatrix}$. Then

$$P^*P = \begin{bmatrix} — & x_1^* & — \\ & \vdots & \\ — & x_6^* & — \end{bmatrix} \begin{bmatrix} | & & | \\ x_1 & \cdots & x_6 \\ | & & | \end{bmatrix} = \begin{bmatrix} \langle x_1, x_1\rangle & \langle x_2, x_1\rangle & \cdots & \langle x_6, x_1\rangle \\ \langle x_1, x_2\rangle & \langle x_2, x_2\rangle & \cdots & \langle x_6, x_2\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle x_1, x_6\rangle & \langle x_2, x_6\rangle & \cdots & \langle x_6, x_6\rangle \end{bmatrix} = I$$

Therefore P is unitary. We also have

$$AP = A\begin{bmatrix} | & & | \\ x_1 & \cdots & x_6 \\ | & & | \end{bmatrix} = \begin{bmatrix} | & | & | & | & | & | \\ \lambda_1 x_1 & \lambda_2 x_2 & \lambda_2 x_3 & \lambda_3 x_4 & \lambda_3 x_5 & \lambda_3 x_6 \\ | & | & | & | & | & | \end{bmatrix} = P \underbrace{\begin{bmatrix} \lambda_1 & & & & & 0 \\ & \lambda_2 & & & & \\ & & \lambda_2 & & & \\ & & & \lambda_3 & & \\ & & & & \lambda_3 & \\ 0 & & & & & \lambda_3 \end{bmatrix}}_{\Lambda}$$

Then $P^*AP = \Lambda$.

2017

Theorem Let $A \in \mathbb{C}^{n\times n}$ be hermitian with eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_r$. Let also $\lambda_{min} := \min_i \lambda_i$ and $\lambda_{max} : \max_i \lambda_i$. Then for all $x \in \mathbb{C}^n$ we have

$$\lambda_{min} \langle x, x \rangle \leq \langle x, Ax \rangle \leq \lambda_{max} \langle x, x \rangle$$

Proof Recall $\mathbb{C}^n = N([A-\lambda_1 I]) \overset{\perp}{\oplus} N([A-\lambda_2 I]) \overset{\perp}{\oplus} \cdots \overset{\perp}{\oplus} N([A-\lambda_r I])$. Hence, given x, we can find $x_i \in N([A-\lambda_i I])$ and write $x = x_1 + x_2 + \cdots + x_r$. Then

$$\langle x, Ax \rangle = \left\langle \sum_{i=1}^{r} x_i , A\left(\sum_{j=1}^{r} x_j\right) \right\rangle$$

$$= \left\langle \sum_i x_i , \sum_j A x_j \right\rangle$$

$$= \left\langle \sum_i x_i , \sum_j \lambda_j x_j \right\rangle$$

$$= \sum_i \left\langle x_i , \sum_j \lambda_j x_j \right\rangle$$

$\langle x_i, x_j \rangle = 0$ for $i \neq j$

$$= \sum_i \langle x_i , \lambda_i x_i \rangle$$

λ_i real

$$= \sum_i \lambda_i \langle x_i , x_i \rangle$$

$$\Rightarrow \quad \lambda_{min} \sum_{i=1}^{r} \langle x_i , x_i \rangle \leq \langle x, Ax \rangle \leq \lambda_{max} \sum_{i=1}^{r} \langle x_i , x_i \rangle \qquad (1)$$

Moreover,
$$\langle x, x \rangle = \left\langle \sum_{i=1}^{r} x_i , \sum_{j=1}^{r} x_j \right\rangle$$

$$= \sum_i \left\langle x_i , \sum_j x_j \right\rangle$$

$\langle x_i, x_j \rangle = 0$ for $i \neq j$

$$= \sum_i \langle x_i , x_i \rangle \qquad (2)$$

Combining (1) and (2) yields the result. $\square$

<u>An application</u> For $A \in \mathbb{C}^{m \times n}$ consider the induced matrix norm

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}.$$

A closed-form expression for $\|A\|$ can be obtained as follows.

$$\frac{\|Ax\|_2^2}{\|x\|_2^2} = \frac{\langle Ax, Ax\rangle}{\langle x, x\rangle}$$

$$= \frac{\langle x, A^*Ax\rangle}{\langle x, x\rangle}$$

Observe that $[A^*A]^* = A^*A$. That is, the matrix $[A^*A]$ is hermitian.

$$\leq \frac{\lambda_{max}(A^*A)\langle x, x\rangle}{\langle x, x\rangle}$$

$$= \lambda_{max}(A^*A)$$

[Notation: $\lambda_{max}(A^*A)$ = max eigenvalue of A^*A]

Note that we can attain $\langle x, A^*Ax\rangle = \lambda_{max}(A^*A)\langle x, x\rangle$ by choosing x the eigenvector corresponding to λ_{max}. Hence $\|A\| = \sqrt{\lambda_{max}(A^*A)}$.

— o —

<u>Exercise</u> [Difficult] Show that $\|A\| = \|A^*\|$.

— o —

<u>Definition</u> A hermitian matrix A is said to be <u>positive definite</u> [Notation: $A \succ 0$] if $\langle x, Ax\rangle > 0$ for all $x \neq 0$. (Positive semi-definite if $\langle x, Ax\rangle \geq 0$.)

<u>Theorem</u> Let A be hermitian positive definite. Then all its eigenvalues are positive.

<u>Proof</u> Suppose not. Then A has an eigenvalue $\lambda \leq 0$. Let x be the corresponding eigenvector. Then we have the following contradiction.

$$0 < \langle x, Ax\rangle = \langle x, \lambda x\rangle = \lambda \underbrace{\langle x, x\rangle}_{>0} \leq 0$$

☑

2017

<u>Example</u> Let $P, Q \in \mathbb{R}^{n\times n}$ be symmetric pos. def. matrices and $A \in \mathbb{R}^{n\times n}$ satisfies the "Lyapunov Equation"

$$A^T P + PA + Q = 0.$$

Show that the eigenvalues λ_i of A satisfy $Re(\lambda_i) < 0$.

<u>Sol'n</u> Let $v_i \in \mathbb{C}^n$ be the eigenvector for λ_i, i.e., $Av_i = \lambda_i v_i$. We have

$$0 = v_i^*(A^T P + PA + Q)v_i$$

$$= v_i^* A^T P v_i + v_i^* PAv_i + v_i^* Q v_i$$

$$= (Av_i)^* P v_i + v_i^* P(Av_i) + v_i^* Q v_i$$

$$= (\lambda_i v_i)^* P v_i + v_i^* P(\lambda_i v_i) + v_i^* Q v_i$$

$$= \bar{\lambda}_i [v_i^* P v_i] + \lambda_i [v_i^* P v_i] + v_i^* Q v_i$$

$$= \underbrace{(\lambda_i + \bar{\lambda}_i)}_{2Re(\lambda_i)} \underbrace{[v_i^* P v_i]}_{>0} + \underbrace{v_i^* Q v_i}_{>0}$$

$$\Rightarrow \quad Re(\lambda_i) = -\frac{1}{2} \frac{v_i^* Q v_i}{v_i^* P v_i} < 0$$

$\blacksquare$

FUNCTION of a MATRIX

Motivation LTI differential equations

$$(1)\quad \begin{cases} \text{diff. eqn.} & \dot{x}(t) = Ax(t) \\ \text{init. cond.} & x(0) = x_0 \end{cases} \qquad \Bigg| \quad x \in \mathbb{C}^n,\ A \in \mathbb{C}^{n \times n}$$

sol'n $x(t) = ?$

Guess The form of the sol'n:

$$(2) \qquad x(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots \qquad \text{with } \alpha_i \in \mathbb{C}^n.$$

$\alpha_0 = ?$

$$\left.\begin{aligned} (1) &\Rightarrow x(0) = x_0 \,! \\ (2) &\Rightarrow x(0) = \alpha_0 \end{aligned}\right\} \quad \boxed{\alpha_0 = x_0}$$

$\alpha_1 = ?$

$$\left.\begin{aligned} (1) &\Rightarrow \dot{x}(0) = Ax(0) = Ax_0 \\ (2) &\Rightarrow \dot{x}(0) = \alpha_1 \end{aligned}\right\} \quad \boxed{\alpha_1 = Ax_0}$$

$\alpha_2 = ?$

$$\left.\begin{aligned} (1) &\Rightarrow \ddot{x}(0) = A\dot{x}(0) = A^2 x(0) = A^2 x_0 \\ (2) &\Rightarrow \ddot{x}(0) = 2\alpha_2 \end{aligned}\right\} \quad \boxed{\alpha_2 = \frac{1}{2} A^2 x_0}$$

$\vdots$

Hence, in general, $\alpha_k = \frac{1}{k!} A^k x_0$

Then $x(t) = \left[\sum_{k=0}^{\infty} \frac{t^k}{k!} A^k \right] x_0$

Define $\boxed{e^{At} := \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k}$ ↝ "state transition matrix"

Solution to (1) is $\boxed{x(t) = e^{At} x(0)}$

Exact discretization of $\dot{x} = Ax$ reads

$$x[(k+1)T] = e^{AT} x[kT]$$

T = sampling period

$k = 0, 1, 2, \ldots$ discrete time variable.

Without loss of generality take $T=1$. Then

$$x[k+1] = e^{A} x[k]$$: discrete-time system

$e^A = ?$

Now, $e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k \qquad (A^0 = I)$

Q: How to compute e^A (without infinite summation)?

A: Cayley-Hamilton Thm.

Recall: $A^n = \gamma_0 I + \gamma_1 A + \gamma_2 A^2 + \cdots + \gamma_{n-1} A^{n-1}$ for some scalars $\gamma_0, \gamma_1, \ldots, \gamma_{n-1}$

Hence, there should exist a polynomial $p(s) = c_0 + c_1 s + \cdots + c_{n-1} s^{n-1}$ such that $e^A = p(A) = c_0 I + c_1 A + \cdots + c_{n-1} A^{n-1}$.

Q: What are $c_0, c_1, \ldots, c_{n-1}$?

A: Eigenvalue - eigenvector eqn.

Assume A has n distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ with eigenvectors $e_1, e_2, \ldots, e_n$.

$$\Rightarrow e^A = p(A)$$

$$\Rightarrow \sum_{k=0}^{\infty} \frac{A^k}{k!} = c_0 I + c_1 A + \cdots + c_{n-1} A^{n-1}$$

$$\Rightarrow \left(\sum_{k=0}^{\infty} \frac{A^k}{k!}\right) e_i = \left(c_0 I + c_1 A + \cdots + c_{n-1} A^{n-1}\right) e_i$$

$A e_i = \lambda_i e_i \Rightarrow A^k e_i = \lambda_i^k e_i$

$$\Rightarrow \underbrace{\left(\sum_{k=0}^{\infty} \frac{\lambda_i^k}{k!}\right)}_{e^{\lambda_i}} e_i = \left(c_0 + c_1 \lambda_i + \cdots + c_{n-1} \lambda_i^{n-1}\right) e_i$$

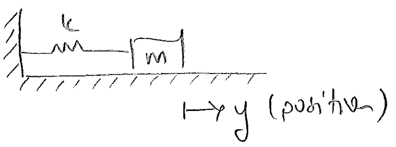
$$\Rightarrow e^{\lambda_i} = \begin{bmatrix} 1 & \lambda_i & \cdots & \lambda_i^{n-1} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{bmatrix}$$

Hence,

$$\underbrace{\begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_n & \lambda_n^2 & \cdots & \lambda_n^{n-1} \end{bmatrix}}_{\Lambda} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{bmatrix} = \begin{bmatrix} e^{\lambda_1} \\ e^{\lambda_2} \\ \vdots \\ e^{\lambda_n} \end{bmatrix}$$

Λ^{-1} exists if $\lambda_1, \lambda_2, \ldots, \lambda_n$ are distinct $(\lambda_i \neq \lambda_j)$. This allows us to compute the coefficients $c_0, c_1, \ldots, c_{n-1}$.

<u>Example</u>



Assume : no friction

For the state choice $x_1 = y$, $x_2 = \dot{y}$ (velocity) :

a) obtain the state space representation $\dot{x} = Ax$.

b) Take $k=1, m=1$. obtain the discretization $x[(k+1)T] = [e^{AT}]\, x[kT]$.

<u>Sol'n</u> a) Diff. eqn. $m\ddot{y} = -ky$

$$\left.\begin{aligned} \Rightarrow \dot{x}_1 &= \dot{y} = x_2 \\ \dot{x}_2 &= \ddot{y} = -\frac{k}{m} y = -\frac{k}{m} x_1 \end{aligned}\right\} \quad \dot{x} = \underbrace{\begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & 0 \end{bmatrix}}_{A} x \qquad \left(x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right)$$

b) $$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \Rightarrow AT = \begin{bmatrix} 0 & T \\ -T & 0 \end{bmatrix} =: F$$

$$\det(sI - F) = s^2 + T^2 = (s - jT)(s + jT) \Rightarrow \lambda_1 = jT \;\&\; \lambda_2 = -jT$$

$$e^F = c_0 I + c_1 F \quad \& \quad \begin{bmatrix} 1 & jT \\ 1 & -jT \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} e^{jT} \\ e^{-jT} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \frac{1}{-j2T} \begin{bmatrix} -jT & -jT \\ -1 & 1 \end{bmatrix} \begin{bmatrix} e^{jT} \\ e^{-jT} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}(e^{jT} + e^{-jT}) \\ \frac{1}{T} \cdot \frac{1}{j2}(e^{jT} - e^{-jT}) \end{bmatrix} = \begin{bmatrix} \cos T \\ \frac{1}{T} \sin T \end{bmatrix}$$

$$\Rightarrow e^F = \cos T \cdot I + \frac{1}{T} \sin T \cdot F = \begin{bmatrix} \cos T & \sin T \\ -\sin T & \cos T \end{bmatrix} = [e^{AT}]$$

↳ rotation matrix !

Total Energy : $E = \frac{1}{2}kx_1^2 + \frac{1}{2}m x_2^2 = [x_1 \; x_2] \underbrace{\begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}}_{P} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$\Rightarrow \quad E = x^T P x$

$$\Delta E = E(k+1) - E(k) = x(k+1)^T P x(k+1) - x(k)^T P x(k)$$

$$= x(k)^T [e^F]^T P [e^F] x(k) - x(k)^T P x(k)$$

$$= x(k)^T \underbrace{\{ [e^F]^T P [e^F] - P \}}_{?} x(k)$$

$$\underbrace{[e^F]^T P [e^F]}_{\frac{1}{2}I} - P = \frac{1}{2}\left([e^F]^T [e^F] - I \right)$$

$$= \frac{1}{2}\left(\underbrace{\begin{bmatrix} \cos^2 T + \sin^2 T & 0 \\ 0 & \cos^2 T + \sin^2 T \end{bmatrix}}_{I} - I \right)$$

$$= 0$$

$\Rightarrow \quad \Delta E = 0 \quad \Rightarrow$ energy conserved. (as expected since there's no friction)

2017

Computing the function of a matrix $A \in \mathbb{C}^{n\times n}$

Given : $f(s) = \sum_{k=0}^{N} \alpha_k s^k$ (N can be ∞)

Want : $f(A) = \sum_{k=0}^{N} \alpha_k A^k = ?$

ex $\cos(s) = 1 - \frac{s^2}{2} + \frac{s^4}{4!} - \frac{s^6}{6!} + \dots$

$\cos(A) = I - \frac{1}{2}A^2 + \frac{1}{4!}A^4 - \frac{1}{6!}A^6 + \dots$

Approach We can write $f(s) = \underbrace{d(s)}_{\text{char. poly.}} q(s) + \underbrace{r(s)}_{\text{poly. of degree} \le n-1}$

$\Rightarrow f(A) = \underbrace{d(A)}_{=0} q(A) + r(A) = r(A)$

Let $r(s) = \sum_{k=0}^{n-1} c_k s^k$. Then

"computing $f(A)$" $\Leftrightarrow$ "finding $c_0, c_1, \dots, c_{n-1}$"

To compute the coefficients $c_0, c_1, \dots, c_{n-1}$ we need n equations.

Let $d(s) = (s-\lambda_1)^{r_1}(s-\lambda_2)^{r_2}\dots(s-\lambda_\sigma)^{r_\sigma}$

$\Rightarrow f(s) = d(s)q(s) + r(s) = (s-\lambda_i)^{r_i}\underbrace{[\text{some poly.}]}_{\text{say } b(s)} + r(s)$

$\Rightarrow f'(s) = r_i(s-\lambda_i)^{r_i-1}b(s) + (s-\lambda_i)^{r_i}b'(s) + r'(s)$

$= (s-\lambda_i)^{r_i-1}[\text{some poly.}] + r'(s)$

$\Rightarrow f''(s) = (s-\lambda_i)^{r_i-2}[\text{some poly.}] + r''(s)$

$\vdots$

$\Rightarrow f^{(r_i-1)}(s) = (s-\lambda_i)[\text{some poly.}] + r^{(r_i-1)}(s)$

Now, let e_i be the eigenvector for λ_i, i.e., $Ae_i = \lambda_i e_i$. Then

$$f(\lambda_i)e_i = \left(\sum \alpha_k A^k\right)e_i = f(A)e_i = [\text{some poly. of } A]\underbrace{[A-\lambda_i I]^{r_i} e_i}_{=0} + \underbrace{r(A)e_i}_{r(\lambda_i)e_i}$$

$$= r(\lambda_i)e_i \quad \Rightarrow \quad \boxed{r(\lambda_i) = f(\lambda_i)}$$

$$f'(\lambda_i)e_i = [\text{some poly. of } A]\underbrace{[A-\lambda_i I]^{r_i-1} e_i}_{=0} + r'(A)e_i$$

$$= r'(\lambda_i)e_i \quad \Rightarrow \quad \boxed{r'(\lambda_i) = f'(\lambda_i)}$$

$$f''(\lambda_i)e_i = r''(\lambda_i)e_i \quad \Rightarrow \quad \boxed{r''(\lambda_i) = f''(\lambda_i)} \quad \text{and so on.}$$

Hence each λ_i will give us r_i many equations:

$$\underbrace{\begin{array}{c} r(\lambda_1) = f(\lambda_1) \\ r'(\lambda_1) = f'(\lambda_1) \\ \vdots \\ r^{(r_1-1)}(\lambda_1) = f^{(r_1-1)}(\lambda_1) \end{array}}_{r_1 \text{ eqn.}} + \underbrace{\begin{array}{c} r(\lambda_2) = f(\lambda_2) \\ r'(\lambda_2) = f'(\lambda_2) \\ \vdots \\ r^{(r_2-1)}(\lambda_2) = f^{(r_2-1)}(\lambda_2) \end{array}}_{r_2 \text{ eqn}} + \cdots + \underbrace{\begin{array}{c} r(\lambda_\sigma) = f(\lambda_\sigma) \\ r'(\lambda_\sigma) = f'(\lambda_\sigma) \\ \vdots \\ r^{(r_\sigma-1)}(\lambda_\sigma) = f^{(r_\sigma-1)}(\lambda_\sigma) \end{array}}_{r_\sigma \text{ eqn.}} = n \text{ eqn.}$$

Remark If we know the minimal poly. $m(s) = (s-\lambda_1)^{m_1}\cdots(s-\lambda_\sigma)^{m_\sigma}$, we can write $f(s) = m(s)q(s) + r(s)$ with $\deg r < \deg m =: l = m_1 + m_2 + \cdots + m_\sigma$

$\Rightarrow f(A) = r(A)$ and we need only l equations.

Example $\quad A = \begin{bmatrix} 0 & 0 & -2 \\ 0 & 1 & 0 \\ 1 & 0 & 3 \end{bmatrix}$. Find e^A.

$$\det(sI - A) = (s-1)^2(s-2) = d(s)$$

$$A - \lambda_1 I = \begin{bmatrix} -1 & 0 & -2 \\ 0 & 0 & 0 \\ 1 & 0 & 2 \end{bmatrix} \Rightarrow \dim \mathcal{N}[A - \lambda_1 I] = 2 \Rightarrow m(s) = (s-1)(s-2)$$

soln with $d(s)$

$$f(s) = e^s, \qquad f(s) = \underbrace{d(s)}_{deg=3} q(s) + \underbrace{r(s)}_{deg=2}, \qquad r(s) = c_2 s^2 + c_1 s + c_0$$

$\lambda_1 = 1 \quad (r_1 = 2)$

$$\left.\begin{array}{l} r(\lambda_1) = f(\lambda_1) \\ r'(\lambda_1) = f'(\lambda_1) \end{array}\right\} \quad \begin{array}{ll} c_2 + c_1 + c_0 = e & (1) \\ 2c_2 + c_1 = e & (2) \end{array}$$

$\lambda_2 = 2 \quad (r_2 = 1)$

$$r(\lambda_2) = f(\lambda_2) \Rightarrow 4c_2 + 2c_1 + c_0 = e^2 \quad (3)$$

$$(1),(2),(3) \Rightarrow \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 0 \\ 4 & 2 & 1 \end{bmatrix} \begin{bmatrix} c_2 \\ c_1 \\ c_0 \end{bmatrix} = \begin{bmatrix} e \\ e \\ e^2 \end{bmatrix} \Rightarrow \begin{bmatrix} c_2 \\ c_1 \\ c_0 \end{bmatrix} = \begin{bmatrix} 1.9525 \\ -1.1867 \\ 1.9525 \end{bmatrix}$$

soln with $m(s)$

$$f(s) = \underbrace{m(s)}_{deg=2} \hat{q}(s) + \underbrace{\hat{r}(s)}_{deg=1}, \qquad \hat{r}(s) = \hat{c}_1 s + \hat{c}_0$$

$\lambda_1 = 1 \quad (m_1 = 1)$

$$\hat{r}(\lambda_1) = f(\lambda_1) \Rightarrow \hat{c}_1 + \hat{c}_0 = e \quad (\hat{1})$$

$\lambda_2 = 2 \quad (m_2 = 1)$

$$\hat{r}(\lambda_2) = f(\lambda_2) \Rightarrow 2\hat{c}_1 + \hat{c}_0 = e^2 \quad (\hat{2})$$

$$(\hat{1}),(\hat{2}) \Rightarrow \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} \hat{c}_1 \\ \hat{c}_0 \end{bmatrix} = \begin{bmatrix} e \\ e^2 \end{bmatrix} \Rightarrow \begin{bmatrix} \hat{c}_1 \\ \hat{c}_0 \end{bmatrix} = \begin{bmatrix} 4.6708 \\ -1.9525 \end{bmatrix}$$

Exercise: check whether $r(A) = \hat{r}(A)$.

Example $A = \begin{bmatrix} 1.5 & 0.5 \\ -0.5 & 0.5 \end{bmatrix}$ Find A^{100}.

$f(s) = s^{100}$

$d(s) = (s-1)^2$ & $A \neq I \Rightarrow m(s) = (s-1)^2$

$f(s) = \underbrace{m(s)}_{deg=2} q(s) + \underbrace{r(s)}_{deg=1} \Rightarrow r(s) = as + b$

$$\left.\begin{array}{l} r(1) = f(1) \\ r'(1) = f'(1) \end{array}\right\} \quad \left.\begin{array}{l} a+b = 1^{100} \\ a = 100 \cdot 1^{99} \end{array}\right\} \quad \left.\begin{array}{l} a = 100 \\ b = -99 \end{array}\right\} \quad r(s) = 100s - 99$$

$$\Rightarrow A^{100} = r(A) = 100 \cdot A - 99 \cdot I = \begin{bmatrix} 150 & 50 \\ -50 & 50 \end{bmatrix} - \begin{bmatrix} 99 & 0 \\ 0 & 99 \end{bmatrix} = \boxed{\begin{bmatrix} 51 & 50 \\ -50 & -49 \end{bmatrix}}$$

— o —

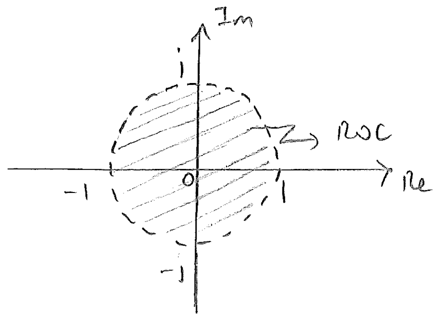
Let $f : \mathbb{C} \to \mathbb{C}$ be a complex-valued function of a complex variable. Provided that $s = 0$ is not a singularity of f, the function has a power series representation around $s=0$ as

$$f(s) = \sum_{k=0}^{\infty} c_k s^k \qquad (1)$$

The region of convergence (ROC) is the subset of $\mathbb{C}$ where eq. (1) holds.

Example $f(s) = \dfrac{1}{1-s}$

$$\Rightarrow f(s) = \sum_{k=0}^{\infty} s^k \quad \text{with} \quad ROC = \{ s \in \mathbb{C} : |s| < 1 \}$$



Given $f : \mathbb{C} \to \mathbb{C}$ with power series representation

$$f(s) = \sum_{k=0}^{\infty} c_k s^k$$

valid in some ROC, we can define $f(A)$ as

$$f(A) = \sum_{k=0}^{\infty} c_k A^k$$

provided that all eigenvalues λ_i of A belong to ROC.

Example For $f(s) = (1-s)^{-1}$

We can define $f(A) = \sum_{k=0}^{\infty} A^k$ provided that $\lambda_i \in \underbrace{\{s \in \mathbb{C} : |s| < 1\}}_{\text{ROC}}$

Example For $f(s) = e^s$

We can always write $f(A) = \sum_{k=0}^{\infty} \frac{1}{k!} A^k$ since ROC for e^s is the entire $\mathbb{C}$.

Exercise Let $f(s) = \sum_{k=0}^{\infty} c_k s^k$ with some ROC and $A \in \mathbb{C}^{n \times n}$ be such that $\lambda_i \in$ ROC and distinct for $i = 1, 2, \ldots, n$. Let $A = P \Lambda P^{-1}$ with $\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$

Then show that

$$f(A) = P \underbrace{\begin{bmatrix} f(\lambda_1) & & & 0 \\ & f(\lambda_2) & & \\ & & \ddots & \\ 0 & & & f(\lambda_n) \end{bmatrix}}_{f(\Lambda)} P^{-1} = P f(\Lambda) P^{-1}$$