ECOE 505 - Linear Systems and Estimation Theory

Course Plan
- Linear and Matrix Theory Review
- Linear System Theory
- Linear Estimation Theory

Instructor: ALPER ERDOGAN

Linear Algebra Review
- Intro
- Vector Space
- Analysis of $A\underset{\sim}{x} = \underset{\sim}{b}$ problem (its existence and uniqueness)

Goal: To use geometric intuition to analyse algebraic problems, especially

$$A\underset{\sim}{x} = \underset{\sim}{b}$$

Historical progress followed the reverse path

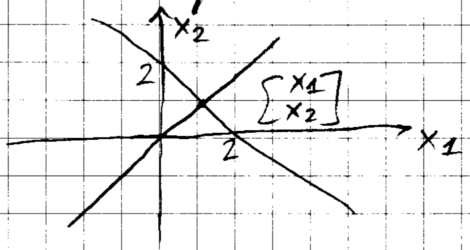
Geometric problem → Algebra involved

In our case we will look at algebra → geometry

Consider a linear system of equations

$$x_1 + x_2 = 2$$

$$x_1 - x_2 = 0$$



Vector Space $(V, R, +, \cdot)$

Definition: A linear vector space V over a set R (set of scalars) is a collection of objects known as vectors together with an additive operation $+$ and scalar multiplicative operation $\cdot$ that satisfy the following properties

$$+ : V \times V \to V$$

$$\cdot : V \times V \to V$$

1. V forms a group under addition $(V, +)$
   - C (closed) For any $\underset{\sim}{x}, \underset{\sim}{y} \in V$ $\quad \underset{\sim}{x} + \underset{\sim}{y} \in V$
   - (Identity element) There is an identity element in V, which is denoted by $\underset{\sim}{0}$, such that for any $\underset{\sim}{x} \in V$ $\quad \underset{\sim}{x} + \underset{\sim}{0} = \underset{\sim}{0} + \underset{\sim}{x} = \underset{\sim}{x}$

- (Inverse) For any element $\underset{\sim}{x} \in V$, there is another element $y \in V$ such that $\underset{\sim}{x} + \underset{\sim}{y} = \underset{\sim}{0}$.
- (Associative) $(\underset{\sim}{x}+\underset{\sim}{y})+\underset{\sim}{z} = \underset{\sim}{x} + (\underset{\sim}{y}+\underset{\sim}{z}) \quad \forall\, \underset{\sim}{x},\underset{\sim}{y},\underset{\sim}{z} \in V$

2. For any $a,b \in R$ $\underset{\sim}{x},\underset{\sim}{y} \in V$

- $a\cdot\underset{\sim}{x} \in V$
- $a\cdot(b\cdot\underset{\sim}{x}) = (ab)\cdot\underset{\sim}{x}$ Multiplication defined in R
- $(a+b)\cdot\underset{\sim}{x} = a\cdot\underset{\sim}{x} + b\cdot\underset{\sim}{x}$
- $a\cdot(\underset{\sim}{x}+\underset{\sim}{y}) = a\cdot\underset{\sim}{x} + b\cdot\underset{\sim}{y}$

3. There is a multiplicative identity $1 \in R$ such that $1.\underset{\sim}{x} = \underset{\sim}{x}$ $\forall\, \underset{\sim}{x} \in V$. There is an element $0 \in R$ s.t. $0.\underset{\sim}{x} = \underset{\sim}{0}$.

Notation: We can use Vector Space and Signal Space interchangeably.

Example N dimensional Euclidean space $\mathbb{R}^N$ where

$$\underset{\sim}{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \qquad x_1, \dots, x_N \in \mathbb{R} \qquad \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \triangleq \begin{bmatrix} x_1+y_1 \\ \vdots \\ x_N+y_N \end{bmatrix} \qquad a.\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} ax_1 \\ \vdots \\ ax_N \end{bmatrix}$$

$$(\overset{V}{\overset{\downarrow}{\mathbb{R}^N}}, \overset{R}{\overset{\downarrow}{\mathbb{R}}}, +, \cdot)$$

Example The set of $M \times N$ matrices

$$(\mathbb{R}^{M\times N}, \mathbb{R}, +, \cdot)$$

Example the set of polynomials up to degree n.

$$\begin{array}{l} \quad a_n x^n + a_{n-1} x^{n-1} + \dots + a_0 \\ \quad b_n x^n + b_{n-1} x^{n-1} + \dots + b_0 \\ + \\ \hline (a_n+b_n)x^n + (a_{n-1}+b_{n-1})x^{n-1} + \dots + (a_0+b_0) \end{array}$$

ECOE 505 - Linear systems and Estimation Theory

Linear Algebra Review

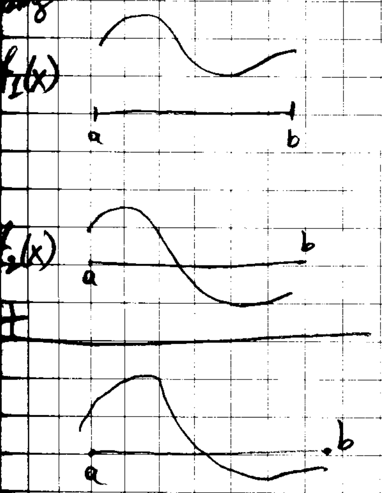
$A\underset{\sim}{x} = \underset{\sim}{b}$

$\boxed{\text{Vector Space}}$ $(V, \mathbb{R}, +, \cdot)$

$\mathbb{R}^2, \mathbb{R}^3$

- $\mathbb{R}^n, \mathbb{R}^{m\times n}$
- Polynomials with degree up to n.
- Example: The set of continuous functions defined over the interval $[a,b]$.

ding



→ also a continuous function

We can talk about 2 functions being orthogonal to each other.

xample: Finite energy signals in the interval $[a,b]$, i.e., the collection

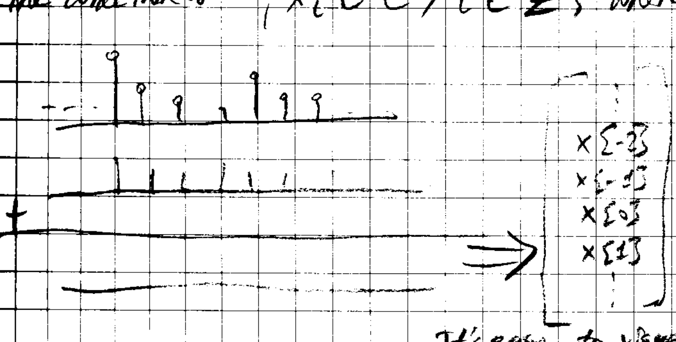
$\{f(x) \in C : x \in [a,b]\}$ where

$$\int_a^b |f(t)|^2 dt < \infty$$ (Energy definition in EE)

Also form a vector space. > Starting vector space for many signal processing tasks.

mple: The set of discrete time sequences with finite energy; i.e.

the collection of $\{x_i \in C, i \in \mathbb{Z}\}$ where $\sum_{i=-\infty}^{\infty} |x_i|^2 < \infty$.

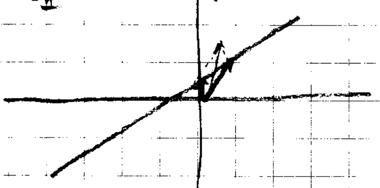


It's easy to visualize discrete-time sequences as vectors.

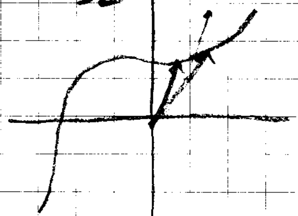
Subspace Let V be a vector space. if $S \subset V$ and if S itself is a vector space then S is a subspace of V.

Example: $V = \mathbb{R}^2$

S_1 not a subspace

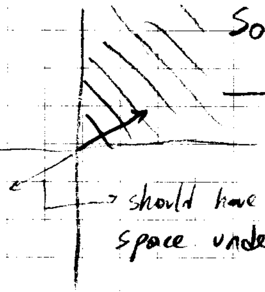


S_2 not a subspace



$S_1 \subset \mathbb{R}^2$

$S_2 \subset \mathbb{R}^2$

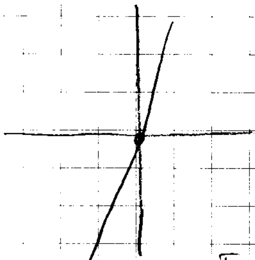


So not a space

→ convex cone. (positive combinations of vectors are closed) weightings

→ should have been part of the space under scalar multiplication.

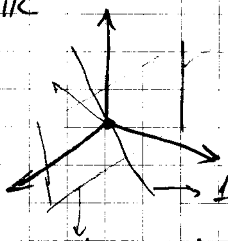
$S_3 (x_1 = x_2)$ ↖ is a subspace



Any line that goes through the origin is a vector space (subspace).?

For a set to qualify as a vector space, it needs to contain the identity element (origin).

Example: $\mathbb{R}^3$



Any line or plane that goes through the origin.

$\rightarrow a_1 x_1 + a_2 x_2 + a_3 x_3 = b_1$

→ 1-dimensional subspace ⟶ 2 equations will give a line.

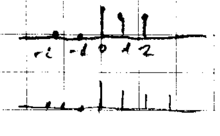
2-dimensional subspace → 3 " " " a point.

For $\mathbb{R}^n$, n-1-dimensional subspaces are called hyperplanes.

We can view a line as the intersection of two planes.

We can view a plane as the vectors orthogonal to the normal of the plane.

Example: the set of causal finite energy discrete time sequences.



↓ all 0 before time 0.

→ subset of finite energy discrete time sequences

→ subspace of finite energy discrete time sequences (closed under +, ·)

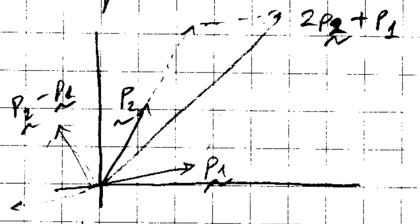
Definition (Linear Combination) Given a vector space V, $\underset{\sim}{x} \in V$ is called a linear combination of vectors $\underset{\sim}{p_1}, \underset{\sim}{p_2}, \ldots, \underset{\sim}{p_m} \in V$ if it can be written as

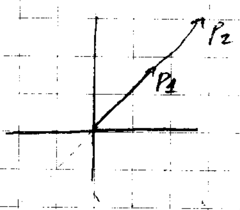
$$\underset{\sim}{x} = \alpha_1 \underset{\sim}{p_1} + \alpha_2 \underset{\sim}{p_2} + \cdots + \alpha_m \underset{\sim}{p_m}$$

Linear combination

~~weighted sum~~

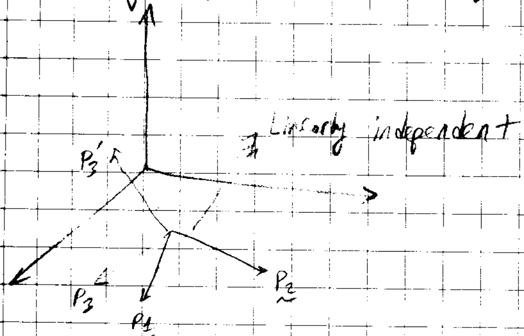
For example $V = \mathbb{R}^2$





Linearly dependent

If linearly independent vectors $\underset{\sim}{p_1}, p_2$ I can represent any point on the $p_1 - p_2$ plane.



Definition (Linear Dependence) Let V be a vector space, and $T \subset V$. The set T is <u>linearly independent</u> if for each nonempty subset of T, $\{\underset{\sim}{p_1}, \ldots, \underset{\sim}{p_m}\}$, the only set of scalors satisfying the equation

$$\alpha_1 \underset{\sim}{p_1} + \alpha_2 \underset{\sim}{p_2} + \cdots + \alpha_m \underset{\sim}{p_m} = \underset{\sim}{0}$$

is the trivial solution $\alpha_1 = \alpha_2 = \cdots = \alpha_\mu = 0$.

$\alpha_1 = 1$ $\quad \underset{\sim}{p_1} - 2\underset{\sim}{p_2} + 2\underset{\sim}{p_3} = \underset{\sim}{0}$

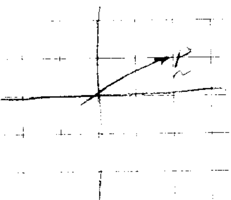
$\alpha_2 = -2$

$\alpha_3 = 2$ $\quad \underset{\sim}{p_1} = 2\underset{\sim}{p_2} - 2\underset{\sim}{p_3}$ $\quad \underset{\sim}{p_1}$ is a linear combination of $\underset{\sim}{p_2}$ and $\underset{\sim}{p_3}$.

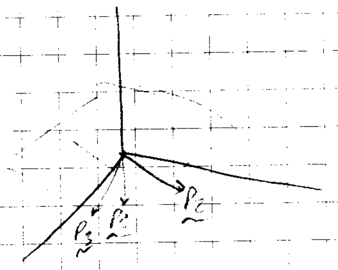
Definition (Span) Let A be a set of vectors in a vector space V. The set of vectors Y that is obtained by taking all possible linear combinations of vectors in A is called the span of A and we use the notation

$$Y = \text{Span}\{A\}$$

$V = \mathbb{R}^2$ $A = \{\underset{\sim}{p}\}$ $V = \mathbb{R}^3$ $A = \{\underset{\sim}{p_1}, \underset{\sim}{p_2}\}$



span{A} is a vector space



$A = \{\underset{\sim}{p_1}, \underset{\sim}{p_2}, \underset{\sim}{p_3}\}$

Span{A} is the smallest linear space that contains A.

If you can eliminate a vector from the set without changing the span, then that vector is linearly dependent.

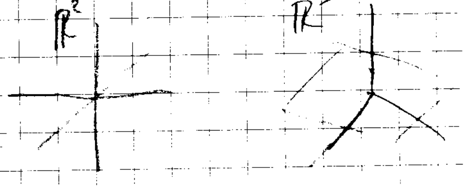
Definition (Basis and Dimension) Let V be a vector space if the set A is linearly independent and V = span{A} then A is a basis for V.
The cardinality (# of elements for finite sets) of A is called the dimension of the vector space V, which is the number of independent vectors to span V.

span{$\underset{\sim}{p_1}, \underset{\sim}{p_2}$} is a 2-dimensional vector space.

Linear Algebra Review

$A\underset{\sim}{x} = \underset{\sim}{b}$

- Vector space $\mathbb{R}^n$, $\mathbb{R}^{m\times n}$, polynomials, functions
- Subspace → Subset of a vector space
  ↘ Vector space



- Linear combination: $(V, R, +, \cdot)$

$$\alpha_1 \underset{\sim}{x_1} + \alpha_2 \underset{\sim}{x_2} + \dots + \alpha_n \underset{\sim}{x_n} = \sum_{i=1}^{n} \alpha_i x_i$$

span is a linear space

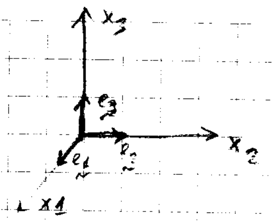
- Span: set A span(A) → 1. All possible linear combinations of elements of set A.
  ↘ 2. The smallest linear space that contains A.
- Linear independence: $\alpha_1 \underset{\sim}{x_1} + \alpha_2 \underset{\sim}{x_2} + \quad + \alpha_n \underset{\sim}{x_n} = \underset{\sim}{0}$ s.t. for $\alpha_1, \alpha_2, \dots, \alpha_n = 0$.
- Basis: A set A is a basis for vector space V iff
  1. span(A) = V
  2. A is linearly independent

# of elements
cardinality of A = dimension of V.

The natural number have the same cardinality with integers. (1-to-1 correspondence)

Example: For $V = \mathbb{R}^3$

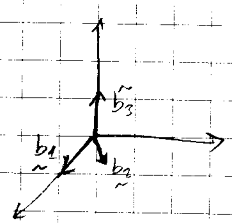
$$\underbrace{\left\{ \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix} \right\}}_{\underset{\sim}{e_1} \quad \underset{\sim}{e_2} \quad \underset{\sim}{e_3}}$$



Standard basis for $\mathbb{R}^3$ — an orthogonal basis — Descartes proposed the orthogonal coordinates.

$$\underset{\sim}{x} = \begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = x_1 \begin{bmatrix}1\\0\\0\end{bmatrix} + x_2 \begin{bmatrix}0\\1\\0\end{bmatrix} + x_3 \begin{bmatrix}0\\0\\1\end{bmatrix} = x_1 \underset{\sim}{e_1} + x_2 \underset{\sim}{e_2} + x_3 \underset{\sim}{e_3}$$

$$\left\{ \underset{\underset{\sim}{b_1}}{\begin{bmatrix}1\\0\\0\end{bmatrix}}, \underset{\underset{\sim}{b_2}}{\begin{bmatrix}1\\1\\0\end{bmatrix}}, \underset{\underset{\sim}{b_3}}{\begin{bmatrix}0\\0\\1\end{bmatrix}} \right\}$$

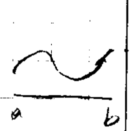


$$\underset{\sim}{x} = \alpha_1 \underset{\sim}{b_1} + \alpha_2 \underset{\sim}{b_2} + \alpha_3 \underset{\sim}{b_3}$$

Example: V = vector space of polynomials

$$A = \{ \underbrace{1, t, t^2, \ldots}_{\text{the set of monomials}} \} \rightarrow \text{not necessarily orthogonal to each other.}$$

Example: For $\mathcal{L}_2[a,b] : \{ f(x) : x \in [a,b], \int_a^b |f(x)|^2 \, dx < \infty \}$



finite energy

Lebeck integration is an alternative for Riemann integration (Analysis course)

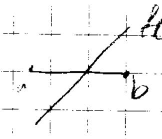
$$A = \{ e^{j \frac{2\pi k t}{b-a}} ; k \in \mathbb{Z} \}$$

$$f(t) = \sum_{k=-\infty}^{\infty} \alpha_k e^{j \frac{2\pi k t}{b-a}} \longrightarrow \text{Fourier series} \rightarrow \text{periodic function}$$

→ Represents a function as a linear combination of basis functions.

↘ Non periodic function

for each k, we have a different basis element.



Why do I need complex exponentials to represent functions?

↓

eigenvectors of linear time invariant systems → ?

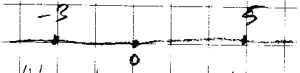
$$\sum_k e^{s_k t} \longrightarrow \boxed{LTI} \longrightarrow \sum_k A e^{s_k t}$$ (form of the function is not going to change)

Eigenfunction concept (I give a waveform, I get a waveform)

Euclidian Norm of a Vector

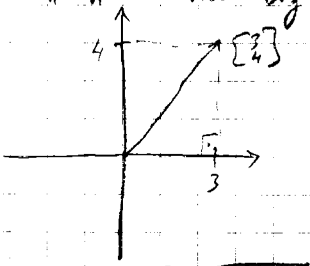
Before that, how big a real number is?

5, -3, 2

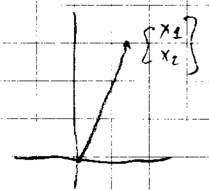


Euclidian distance to the origin

In $\mathbb{R}^2$ How big is $\begin{bmatrix}3\\4\end{bmatrix}$? There are infinitely many ways of saying how big a vector is.



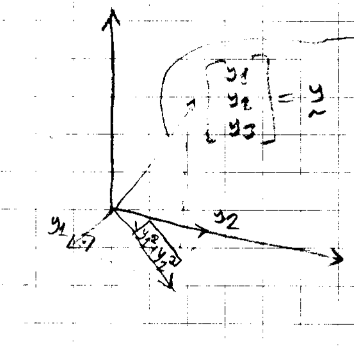
$\sqrt{3^2+4^2} = 5$



$\sqrt{x_1^2+x_2^2}$

$\|\underset{\sim}{x}\| = \sqrt{x_1^2+x_2^2}$

$\mathbb{R}^3$



$\|\underset{\sim}{y}\| = \sqrt{y_1^2+y_2^2+y_3^2}$ In $\mathbb{C}^3$

$\|\underset{\sim}{y}\| = \sqrt{|y_1|^2+|y_2|^2+|y_3|^2}$ → Gives ?

Extend this definition for $\mathbb{R}^n$

$$\|\underset{\sim}{x}\| = \sqrt{x_1^2+x_2^2+\cdots+x_n^2} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

$$x_1^2+x_2^2+\cdots+x_n^2 = \overset{1\times n}{\underset{\underset{\sim}{x}^T}{\begin{bmatrix}x_1 & x_2 & \cdots & x_n\end{bmatrix}}}\ \overset{n\times 1}{\underset{\underset{\sim}{x}}{\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}}}$$

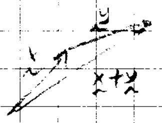
$$= \underset{\sim}{x}^T\underset{\sim}{x}$$

$$\|\underset{\sim}{x}\| = \sqrt{\underset{\sim}{x}^T\underset{\sim}{x}} \Leftrightarrow \|\underset{\sim}{x}\|^2 = \underset{\sim}{x}^T\underset{\sim}{x}$$

Note that Euclidian norm satisfies

- $\|\alpha\underset{\sim}{x}\| = |\alpha|\,\|\underset{\sim}{x}\|$ $\alpha\in\mathbb{R}$ (Homogeneity)
- $\|\underset{\sim}{x}\| \geqslant 0$ (Nonnegative)

  $\|\underset{\sim}{x}\| = 0 \Leftrightarrow \underset{\sim}{x} = \underset{\sim}{0}$ (Definiteness)
- $\|\underset{\sim}{x}+\underset{\sim}{y}\| \leq \|\underset{\sim}{x}\| + \|\underset{\sim}{y}\|$ (Triangle Inequality)



Inner Product

$\underset{\sim}{x},\underset{\sim}{y}\in\mathbb{R}^n$ Def: $\langle\underset{\sim}{x},\underset{\sim}{y}\rangle = \underset{\sim}{x}^T\underset{\sim}{y} = \underset{\sim}{y}^T\underset{\sim}{x} \neq \underset{\sim}{x}\underset{\sim}{y}^T$

$$= \begin{bmatrix}x_1 & x_2 & \cdots & x_n\end{bmatrix}\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix} = x_1y_1+x_2y_2+\cdots+x_ny_n$$

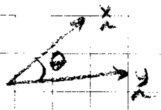
$$= \sum_{k=1}^{n} x_k y_k$$

$$\|\underset{\sim}{x}\| = \sqrt{\langle \underset{\sim}{x}, \underset{\sim}{x} \rangle} \Leftrightarrow \|\underset{\sim}{x}\|^2 = \langle \underset{\sim}{x}, \underset{\sim}{x} \rangle$$

Properties of Inner Product

- $\langle \alpha \underset{\sim}{x}, \underset{\sim}{y} \rangle = \alpha \langle \underset{\sim}{x}, \underset{\sim}{y} \rangle$
- $\langle \underset{\sim}{x} + \underset{\sim}{y}, \underset{\sim}{z} \rangle = (\underset{\sim}{x} + \underset{\sim}{y})^T \underset{\sim}{z} = (\underset{\sim}{x}^T + \underset{\sim}{y}^T)\underset{\sim}{z} = \underset{\sim}{x}^T \underset{\sim}{z} + \underset{\sim}{y}^T \underset{\sim}{z} = \langle \underset{\sim}{x}, \underset{\sim}{z} \rangle + \langle \underset{\sim}{y}, \underset{\sim}{z} \rangle$
- $\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle = \langle \underset{\sim}{y}, \underset{\sim}{x} \rangle$ → in the complex case it matters.
- $\langle \underset{\sim}{x}, \underset{\sim}{x} \rangle \geqslant 0$
- $\langle \underset{\sim}{x}, \underset{\sim}{x} \rangle = 0 \Leftrightarrow \underset{\sim}{x} = \underset{\sim}{0}$

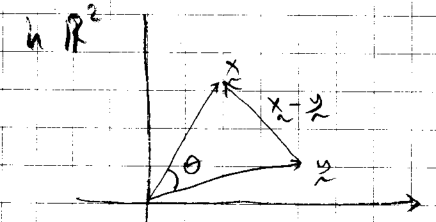
$\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle = \|\underset{\sim}{x}\| \, \|\underset{\sim}{y}\| . \cos\theta$



$$\cos\theta = \frac{\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle}{\|\underset{\sim}{x}\| \, \|\underset{\sim}{y}\|}$$

Span(A) = A if A is a vector space !

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



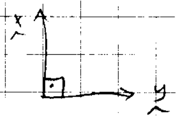
Cosine theorem:

$$\|\underset{\sim}{x} - \underset{\sim}{y}\|^2 = \|\underset{\sim}{x}\|^2 + \|\underset{\sim}{y}\|^2 - 2\|\underset{\sim}{x}\| \, \|\underset{\sim}{y}\| \cos\theta$$

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

$$\cos\theta = \frac{\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle}{\|\underset{\sim}{x}\| \, \|\underset{\sim}{y}\|} = \frac{\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle}{\langle \underset{\sim}{x}, \underset{\sim}{x} \rangle^{1/2} \langle \underset{\sim}{y}, \underset{\sim}{y} \rangle^{1/2}}$$

$\theta = 90^\circ$
-90 $\Rightarrow \underbrace{\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle}_{} = 0 \Leftrightarrow$



orthogonal vectors

$\boxed{|\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle| \leq \|x\| \|y\|}$

Bunyakovski
Cauchy - Schwarz Inequality

$\int f(x) g(x) \, dx \leq \int |f(x)|^2 dx \int |g(x)|^2 dx$

$$\|\underset{\sim}{x} - \underset{\sim}{y}\|^2 = \|\underset{\sim}{x}\|^2 - 2\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle + \|\underset{\sim}{y}\|^2 \geqslant 0$$

$$\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle \leqslant \frac{\|\underset{\sim}{x}\|^2}{2} + \frac{\|\underset{\sim}{y}\|^2}{2}$$

$\underset{\sim}{x} = \frac{\underset{\sim}{u}}{\|\underset{\sim}{u}\|}$ $\quad \underset{\sim}{y} = \frac{\underset{\sim}{v}}{\|\underset{\sim}{v}\|}$ $\Longrightarrow$ Unitnorm vectors $\Rightarrow$ Normalization procedure

$\left\|\frac{\underset{\sim}{v}}{\|\underset{\sim}{v}\|}\right\| = 1.$

$$\left\langle \frac{\underset{\sim}{u}}{\|\underset{\sim}{u}\|}, \frac{\underset{\sim}{v}}{\|\underset{\sim}{v}\|} \right\rangle \leq \frac{\left\|\frac{\underset{\sim}{u}}{\|\underset{\sim}{u}\|}\right\|^2}{2} + \frac{\left\|\frac{\underset{\sim}{v}}{\|\underset{\sim}{v}\|}\right\|^2}{2}$$

$$\left\langle \frac{\underset{\sim}{u}}{\|\underset{\sim}{u}\|}, \frac{\underset{\sim}{v}}{\|\underset{\sim}{v}\|} \right\rangle \leq 1$$

$$\frac{1}{\|\underset{\sim}{u}\|\,\|\underset{\sim}{v}\|} \langle \underset{\sim}{u}, \underset{\sim}{v} \rangle \leq 1 \quad \Rightarrow \quad \langle \underset{\sim}{u}, \underset{\sim}{v} \rangle \leq \|\underset{\sim}{u}\| . \|\underset{\sim}{v}\|$$

Matrices

Notation: A matrix $A = [a_{ij}]_{m \times n}$ is a rectangular array of numbers

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & & & \\ a_{m1} & \cdots & & a_{mn} \end{bmatrix} \qquad A \in \mathbb{R}^{m \times n}$$

- A matrix can be partitioned in many different ways. Two particular useful partitions are

1. $A = [\underset{\sim}{c_1} \ \underset{\sim}{c_2} \ \underset{\sim}{c_3} \ \ldots \ \underset{\sim}{c_n}]$ $\qquad \underset{\sim}{c_i} = \begin{bmatrix} \\ \\ \\ \end{bmatrix} \in \mathbb{R}^{m \times 1}$

$$\underset{\sim}{c_i} = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{bmatrix}$$

2. $A = \begin{bmatrix} \underset{\sim}{r_1}^T \\ \underset{\sim}{r_2}^T \\ \vdots \\ \underset{\sim}{r_m}^T \end{bmatrix}$ $\qquad \underset{\sim}{r_i} = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{bmatrix} \in \mathbb{R}^{n \times 1}$

$$\underset{\sim}{r_i}^T \in \mathbb{R}^{1 \times n}$$

Two alternative ways to perceive matrix vector product $A\underset{\sim}{x}$

1. $A = [\underset{\sim}{c_1}, \ldots, \underset{\sim}{c_n}]$

$$A\underset{\sim}{x} = [\underset{\sim}{c_1}, \ldots \underset{\sim}{c_n}] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

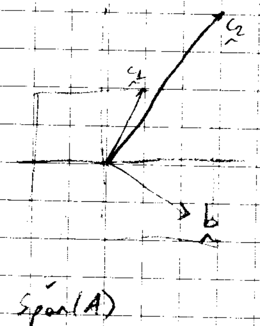
$$= [x_1\underset{\sim}{c_1} + x_2\underset{\sim}{c_2} + \ldots + x_n\underset{\sim}{c_n}]$$ (Linear weighted combination of the colums of A)

$A\underset{\sim}{x} \cong \underset{\sim}{b}$ → Can I find a collection of weights that will give me $\underset{\sim}{b}$ from A?

2. $A = \begin{bmatrix} \underset{\sim}{r_1}^T \\ \vdots \\ \underset{\sim}{r_m}^T \end{bmatrix}$ $\quad A\underset{\sim}{x} = \begin{bmatrix} \underset{\sim}{r_1}^T \\ \vdots \\ \underset{\sim}{r_m}^T \end{bmatrix} \underset{\sim}{x} = \begin{bmatrix} \underset{\sim}{r_1}^T \underset{\sim}{x} \\ \underset{\sim}{r_2}^T \underset{\sim}{x} \\ \vdots \\ \underset{\sim}{r_m}^T \underset{\sim}{x} \end{bmatrix} = \begin{bmatrix} \langle \underset{\sim}{x}, \underset{\sim}{r_1} \rangle \\ \langle \underset{\sim}{x}, \underset{\sim}{r_2} \rangle \\ \vdots \\ \langle \underset{\sim}{x}, \underset{\sim}{r_m} \rangle \end{bmatrix}$

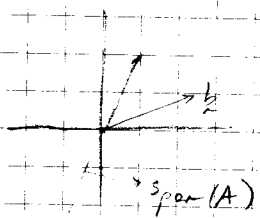
(Taking inner products of $\underset{\sim}{x}$ with the rows of A)

$A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$



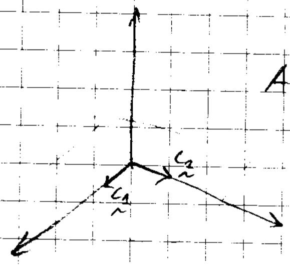
$A\underset{\sim}{x} \cong \underset{\sim}{b}$ → I can find for any $\underset{\sim}{b}$

$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$

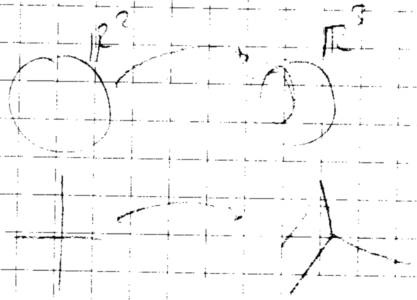


Is $\underset{\sim}{b}$ in the span (columns of A)?

$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$



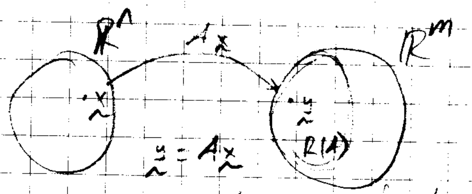
$A\underset{\sim}{x} \overset{?}{\cong} \underset{\sim}{b}$



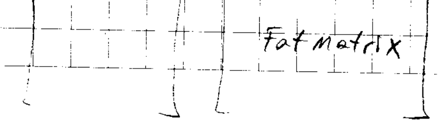
Range Space of a Matrix

- The range of a matrix $A \in \mathbb{R}^{m \times n}$ is the set

$$R(A) = \{ A\underset{\sim}{x} \mid \underset{\sim}{x} \in \mathbb{R}^n \}$$



- $R(A) \subseteq \mathbb{R}^m$
- $R(A) = \text{Span}(\{\underset{\sim}{c_1}, \ldots, \underset{\sim}{c_n}\})$, i.e., the range space is the span of columns of A.
- $R(A)$ can be defined as the collection of $\underset{\sim}{b}$ vectors for which $A\underset{\sim}{x} \cong \underset{\sim}{b}$ has a solution.
- If $R(A) = \mathbb{R}^m$ then that implies
  - $A\underset{\sim}{x} \cong \underset{\sim}{b}$ has solution for any $b \in \mathbb{R}^m$.
  - Span of columns of A is $\mathbb{R}^m$
  - $n \geq m$ (Range space is a span of n vectors) (n: n unknowns, m: equations)



- A has a <u>right inverse</u>, i.e., there exists an $n \times m$ matrix B such that

$$AB = I_m$$ [illegible]

Linear System of Equations

$$\boxed{A\underset{\sim}{x}} \overset{?}{=} \underset{\sim}{b} \begin{matrix} \nearrow \text{Existence} \\ \searrow \text{Uniqueness} \end{matrix}$$

$$A\underset{\sim}{x} \quad A \in \mathbb{R}^{m \times n}$$

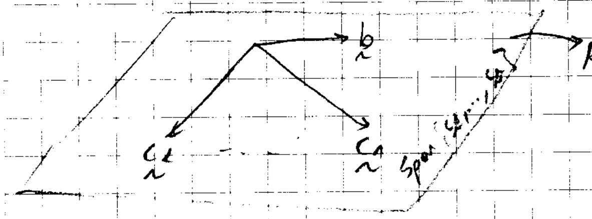
Will be used for the analysis of Existence

$$1. \begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \begin{bmatrix} \vdots \\ x \\ \vdots \end{bmatrix} = \begin{bmatrix} \underset{\sim}{c_1} & \underset{\sim}{c_2} & \cdots & \underset{\sim}{c_n} \end{bmatrix} \begin{bmatrix} \boxed{x_1} \\ \boxed{x_2} \\ \vdots \\ \boxed{x_n} \end{bmatrix} = \boxed{x_1} \begin{bmatrix} \underset{\sim}{c_1} \end{bmatrix} + \boxed{x_2} \begin{bmatrix} \underset{\sim}{c_2} \end{bmatrix} + \cdots + \boxed{x_n} \begin{bmatrix} \underset{\sim}{c_n} \end{bmatrix} = \begin{bmatrix} \\ \\ \end{bmatrix}$$

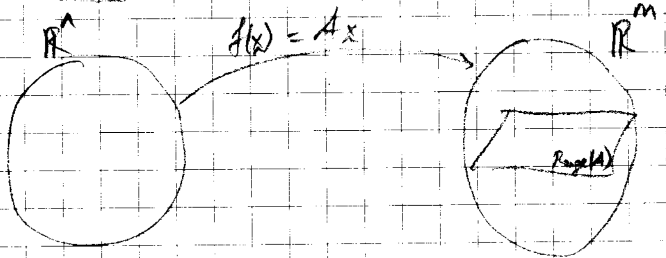
Perception

Uniqueness

$$2. \begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \begin{bmatrix} \\ \underset{\sim}{x} \\ \\ \end{bmatrix} = \begin{bmatrix} \underset{\sim}{r_1}^T \\ \underset{\sim}{r_2}^T \\ \vdots \\ \underset{\sim}{r_m}^T \end{bmatrix} \begin{bmatrix} \\ \underset{\sim}{x} \\ \\ \end{bmatrix} = \begin{bmatrix} \langle \underset{\sim}{r_1}, \underset{\sim}{x} \rangle \\ \vdots \\ \langle \underset{\sim}{r_m}, \underset{\sim}{x} \rangle \end{bmatrix}$$



$$\text{Range}(A) = \{A\underset{\sim}{x} : \underset{\sim}{x} \in \mathbb{R}^n\}$$
$$= \text{Span}\{\underset{\sim}{c_1}, \dots, \underset{\sim}{c_n}\}$$



If $R(A) = \mathbb{R}^m$, we have solution for any $\underset{\sim}{b} \in \mathbb{R}^m$

- $n \geq m$ $\quad \overset{n=m}{\begin{bmatrix} & \\ & \end{bmatrix}} \overset{n>m}{\begin{bmatrix} & & & \\ & & & \end{bmatrix}}$
- $\exists D \in \mathbb{R}^{n \times m}$, right inverse, $AD = I$

$$\underset{\underset{\sim}{d_1}}{A\underset{\sim}{x}} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \underset{\underset{\sim}{d_2}}{A\underset{\sim}{x}} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \cdots \quad \underset{\underset{\sim}{d_m}}{A\underset{\sim}{x}} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

There may be more than one solution $\underset{\sim}{d_i}$ for all tall matrices

$$A\underbrace{[\underset{\sim}{d_1} \ \underset{\sim}{d_2} \ \cdots \ \underset{\sim}{d_m}]}_{D} = [A\underset{\sim}{d_1} \ A\underset{\sim}{d_2} \ \cdots \ A\underset{\sim}{d_m}]$$

$$= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

- $R(A) = \mathbb{R}^m \Longleftrightarrow \det(AA^T) \neq 0.$ (Prove this)

Null Space

$N(A) = \{ \underset{\sim}{x} \in \mathbb{R}^n \mid A\underset{\sim}{x} = \underset{\sim}{0} \}$ $\qquad \underset{\in \mathbb{R}^m}{f(x)} = A\underset{\sim}{x} \rightarrow \in \mathbb{R}^n$

$\underset{\sim}{x} \in N(A)$, $\alpha \underset{\sim}{x}$

$$A(\alpha \underset{\sim}{x}) = \alpha A \underset{\sim}{x} = \underset{\sim}{0}$$

$\underset{\sim}{x_1}, \underset{\sim}{x_2} \in N(A)$

$$A(\underset{\sim}{x_1} + x_2) = A\underset{\sim}{x_1} + A\underset{\sim}{x_2} = \underset{\sim}{0}$$

$A\underset{\sim}{0}' = \underset{\sim}{0}^n$, origin is also included

Vector space

$A\underset{\sim}{z} = \underset{\sim}{b}$, assume $\underset{\sim}{y} \in N(A)$

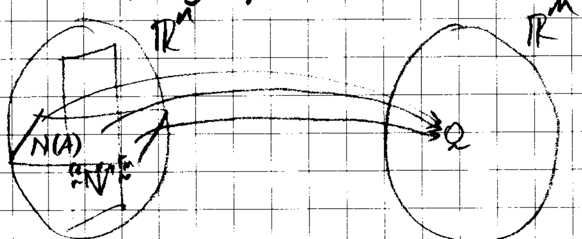
$$A(\underset{\sim}{z} + \underset{\sim}{y}) = A\underset{\sim}{z} + \underset{0}{\cancel{A\underset{\sim}{y}}} = A\underset{\sim}{z} = \underset{\sim}{b}$$

$\underbrace{N(A) = \underset{\sim}{0}}_{\text{Trivial null space}}$ or $\underbrace{N(A) = \{ \ldots \}}_{\text{Non trivial}}$ → Infinitely many solutions (i.e. $N(A) \ni c$ & scaled versions of $\hat{}$ since $N(A)$ is a vector space)

Properties of Null Space:

- $N(A) \subseteq \mathbb{R}^n$ (subspace)

- It can be interpreted as collection of vectors which are mapped to zero under the mapping $\underset{\sim}{y} = A\underset{\sim}{x}$



- Alternative Interpretation: (Due to Perception 2) Collection of vectors in $\mathbb{R}^n$ that are orthogonal to all rows of A. (A null space vector is orthogonal to row space of A)

$$\underset{\sim}{x}^T(\alpha_1 \underset{\sim}{r_1} + \alpha_2 \underset{\sim}{r_2} + \ldots + \alpha_n \underset{\sim}{r_n}) = \alpha_1 \underbrace{\underset{\sim}{x}^T \underset{\sim}{r_1}}_{0} + \alpha_2 \underbrace{\underset{\sim}{x}^T r_2}_{0} + \ldots + \alpha_n \underbrace{\underset{\sim}{x}^T r_n}_{0}$$

$x \in N(A)$

- $N(A) = \{\underset{\sim}{0}\}$ implies

* We can <u>uniquely</u> determine $\underset{\sim}{x}$ from $\underset{\sim}{b}$ if $A\underset{\sim}{x} = \underset{\sim}{b}$ has a solution.

* Columns of A are linearly independent. (Use Perception 1)

$$\underset{\sim}{c_1} \dots \underset{\sim}{c_n} \text{ are linearly independent} \atop \text{iff}$$

$$\alpha_1 \underset{\sim}{c_1} + \alpha_2 \underset{\sim}{c_2} + \dots + \alpha_n \underset{\sim}{c_n} = \underset{\sim}{0} \text{ is } \alpha_1 = \alpha_2 = \dots = 0$$

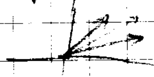
$$\downarrow$$

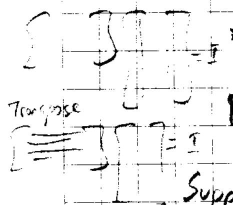
$$[\underset{\sim}{c_1}\ \underset{\sim}{c_2} \dots \underset{\sim}{c_n}] \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \underset{\sim}{0}$$

$n = 10$, $m = 3$ $\Rightarrow$ if $n \geq m$ $\Rightarrow N(A)$ is nontrivial.

* $n \leq m$ $\left(\begin{matrix}\text{Based on the} \\ \text{\# of row} \\ \text{vectors}\end{matrix}\right) \rightarrow$ Second view : Based on the # of column vectors.

Since $\underset{\sim}{c_1}, \dots, \underset{\sim}{c_n}$ are lin. indep. $n \leq m$





* $\exists D \in \mathbb{R}^{n \times m}$, $DA = I$ (there exists a left inverse) (Prove this)

$R(A) = \mathbb{R}^m$ implies rows of A are linearly independent.

Suppose $R(A) = \mathbb{R}^m$ but rows of A are not linearly independent

$\exists\ \alpha_1, \alpha_2, \dots, \alpha_n$ not all equal to zero

$$\alpha_1 \underset{\sim}{r_1} + \alpha_2 \underset{\sim}{r_2} + \dots + \alpha_m \underset{\sim}{r_m} = \underset{\sim}{0}$$

$$[\underset{\sim}{r_1} \dots \underset{\sim}{r_m}] \begin{bmatrix} \alpha_1 \\ \\ \alpha_m \end{bmatrix} = \underset{\sim}{0} \qquad AD = I$$

$$D^T A^T = I$$
$$\hookrightarrow \text{toll}$$

$$[\alpha_1 \dots \alpha_m] \begin{bmatrix} \underset{\sim}{r_1}^T \\ \\ \underset{\sim}{r_m}^T \end{bmatrix} = \underset{\sim}{0}$$

$$\underset{\sim}{\alpha}^T A = \underset{\sim}{0}$$

$$\underset{\sim}{\alpha}^T [\underset{\sim}{c_1} \dots \underset{\sim}{c_n}] = \underset{\sim}{0}$$

$$[\langle \underset{\sim}{\alpha}, \underset{\sim}{c_1} \rangle \dots \langle \underset{\sim}{\alpha}, \underset{\sim}{c_n} \rangle] = \underset{\sim}{0}$$

Linear Algebra Review

$Ax = b$

Existence   Uniqueness

If solution exists, is it unique?

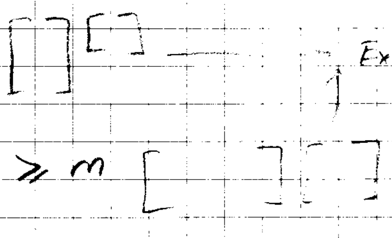
$N(A) = \{0\}$

$n \leq m$

Given (A, b) there a solution? iff $b \in R(A)$

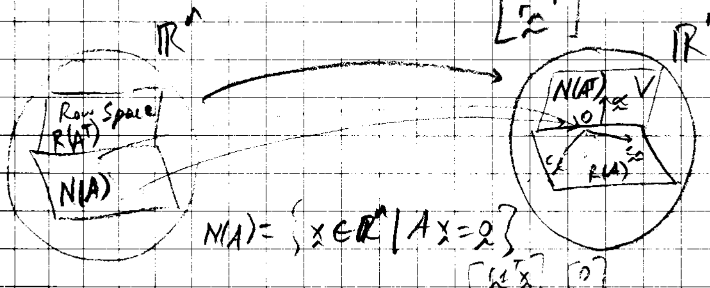
Given A, can we guarantee that we have solution for any b? Yes iff $R(A) = \mathbb{R}^m$

$n \geq m$



Existence & Uniqueness is guarante for square matric

$A \in \mathbb{R}^{m \times n}$ $A = \{c_1, \ldots c_n\} = \begin{bmatrix} r_1^T \\ \\ r_m^T \end{bmatrix}$



$N(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}$

Left null space

If $\alpha \in V$, $\langle \alpha, c_i \rangle = 0$

$$\left.\begin{aligned} \alpha^T c_1 &= 0 \\ \alpha^T c_n &= 0 \end{aligned}\right\} \alpha^T \{c_1 \ \ c_n\} = \{0 \ldots 0$$

$\alpha^T A = 0$

Interpretation of Null Space $\begin{bmatrix} r_1^T x \\ r_m^T x \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

(-) • Suppose we want to measure x, but we have access to $\hat{y} = Ax$.

Let $z \in N(A)$. Since $A(x + z) = Ax$, $N(A)$ characterizes the ambiguity in x given $y = Ax$.

(+) • Alternatively, we want to achieve an output $y = Ax$.

$x \longrightarrow \boxed{Ax} \longrightarrow y$ We could choose any input $x + z$ where $z \in N(A)$.

In this case $N(A)$ characterizes the degree of freedom in choosing x.

Row Space $= \text{Span}\{r_1, \ldots, r_m\}$

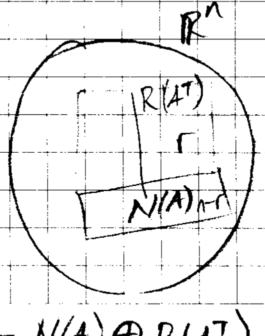
$A^T = \{r_1 \ \cdots \ r_m\}$

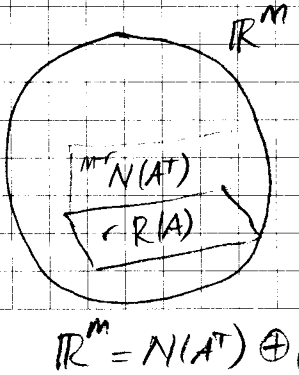
$= R(A^T)$

$\alpha^T A = 0 \Rightarrow A^T \alpha = 0$

Left null space of A = (right) null space of A^T.

4 Fundamental Spaces

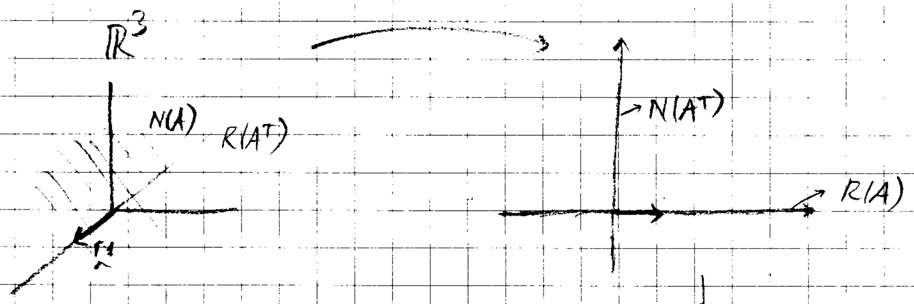




$\mathbb{R}^n = N(A) \oplus R(A^T)$ $\mathbb{R}^m = N(A^T) \oplus R(A)$

↳ Direct sum ≈ span of the union space

Example: $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$



$\underset{\sim}{z} \in \mathbb{R}^3 \rightarrow \underset{\sim}{z} = z_{N(A)} + z_{R(A^T)}$

Solution is not guaranteed to exist.

If it exists, it is not unique.

Rank of a Matrix

· We define the rank of a matrix $A \in \mathbb{R}^{m\times n}$ as

$$\text{rank}(A) = \dim(R(A)) = r$$
$$= \dim(R(A^T))$$

Some facts about rank:

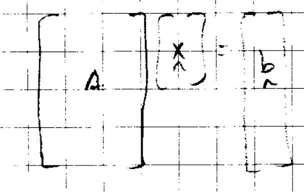
- Rank is the maximum number of linearly independent columns (rows).
- Rank $(A) \leq \min\{m,n\}$ $\dim(R(A)) \leq m$

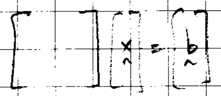
  In case of equality we call "A has full rank". $\dim(R(A^T)) \leq n$

  Otherwise "A is rank deficient".
- $\dim(N(A)) = n-r$
- $\dim(N(A^T)) = m-r$

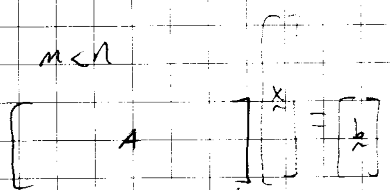
$A\underset{\sim}{x} \cong \underset{\sim}{b}$ Problem

$m > n$ $m = n$ $m < n$





# of eqs = # of unknowns



more unknowns than equations

Uniqueness may be a problem.

More equations than unknowns

Existence may be a problem

Square case $(m=n)$

$$\begin{bmatrix} A \end{bmatrix}\begin{bmatrix} \underset{\sim}{x} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{b} \end{bmatrix}$$

- Rank is equal to n (full rank)
  - $R(A) = \mathbb{R}^n = \mathbb{R}^m \Rightarrow$ We have solution for any $\underset{\sim}{b}$.
  - $\dim(N(A)) = n-r = 0 \Rightarrow N(A) = \{\underset{\sim}{0}\} \Rightarrow$ Solution is unique.
- Rank $< n$ (rank deficient)
  - $R(A) \neq \mathbb{R}^n \Rightarrow \exists \underset{\sim}{b} \notin R(A) \Rightarrow$ There is no solution for some $\underset{\sim}{b}$.
  - $\dim(N(A)) = n-r > 0 \Rightarrow N(A) \neq \{\underset{\sim}{0}\} \Rightarrow$ If solution exists it is not uniqu

If $A \in \mathbb{R}^{n\times n}$ is full rank it has a unique left and right inverse, denoted by A^{-1}, such tha

$$AA^{-1} = A^{-1}A = I$$

$[A][R_1] = [\;]$

$R_n \rightarrow A.R = I$

$LA = I$

$\underbrace{LA}_{I}R = L$

$R = L \Rightarrow$ Unique

Such an A is called invertible or nonsingular.

Equivalent conditions for invertibility:

- Columns (Rows) of A are linearly independent and they form a basis for $\mathbb{R}^n$.
- $\det(A) \neq 0$

$$\det\left(\underbrace{\begin{bmatrix} 1 & & \\ & 1 & \\ \alpha_1 & \cdots & \alpha_n \end{bmatrix}}_{\text{can be } \neq 0}\begin{bmatrix} A \end{bmatrix}\right) = \det\left(\underbrace{\begin{bmatrix} & & \\ 0 & \cdots & 0 \end{bmatrix}}_{0}\right)$$

can be $\neq 0 \Rightarrow \det(A) = 0$

$A\underset{\sim}{x} \overset{?}{=} \underset{\sim}{b}$

$f(\underset{\sim}{x}) = A\underset{\sim}{x}$

$\mathbb{R}^n \quad \mathbb{R}^m$

- If A is full rank, $\exists A^{-1}$ s.t. $A^{-1}A = AA^{-1} = I$ (unique solution for any $\underset{\sim}{b}$)

  $N(A) = \{0\} \qquad R(A) = \mathbb{R}^m$

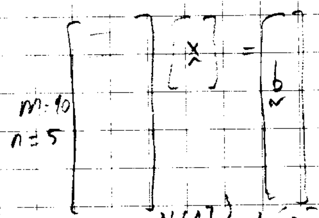
  $\det A \neq 0$

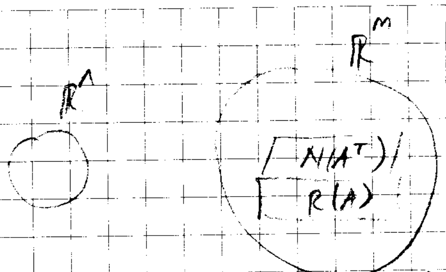
- If A is not full rank (singular)
  - $\det A = 0$

$R(A) \neq \mathbb{R}^m \Rightarrow$ - For some $\underset{\sim}{b}$ there is no solution

$N(A) \neq \{0\} \Rightarrow$ - If there is a solution, it is not unique

Tall A Case $(m > n)$





$N(A^T) \neq \{0\}$ ⇒ for some $\underset{\sim}{b}$ there is no solution (i.e. for $\underset{\sim}{b}$'s $\underset{\sim}{b} \in N(A^T)$, $\underset{\sim}{b} \neq \underset{\sim}{0}$)

- If A is full rank: $r = n$

  ⟺ Columns are linearly independent

  ⟺ $N(A) = \{0\}$

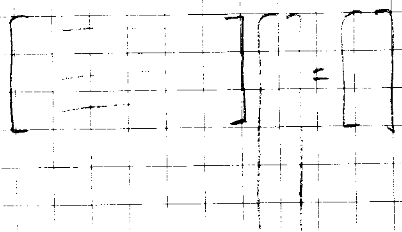
  ⟺ If there is a solution, it is unique.

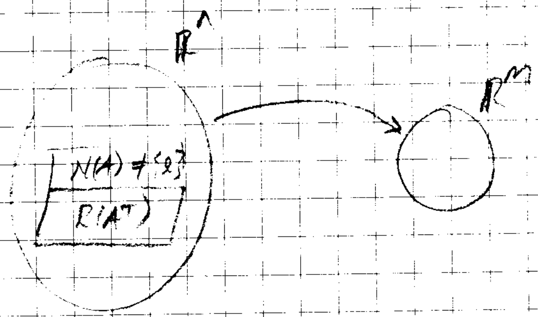
- If A is not full rank: $r < n$

  ⟺ $N(A) \neq \{\underset{\sim}{0}\}$

  If there is a solution, it is not unique!

Fat A Case $(m < n)$





Since $N(A) \neq \{\underset{\sim}{0}\}$ ⟺ If solution exists, it is not unique.

- If A is full rank, i.e., $r = m$, ⟺ $R(A) = \mathbb{R}^m$

  ⟺ Solution exists for any $\underset{\sim}{b} \in \mathbb{R}^m$

- If A is not full rank, i.e., $r < m$, ⟺ $R(A) \neq \mathbb{R}^m$

  ⟺ for some $\underset{\sim}{b}$ there is no solution.

"1000 equations are enough to estimate the parameters" → Ad hoc statement.

Obtaining the solution of $A\underset{\sim}{x} = \underset{\sim}{b}$

Approach: Convert this problem into (multiple) simpler problems:

Simpler problems?

(1)
- A is diagonal

$$\begin{bmatrix} a_{11} & & 0 \\ 0 & & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \Rightarrow \begin{array}{l} x_1 = \dfrac{b_1}{a_{11}} \\ x_2 = \dfrac{b_2}{a_{22}} \\ \vdots \\ x_n = \dfrac{b_n}{a_{nn}} \end{array}$$

(2) - A is (lower or upper) triangular

Gaussian elimination tries to convert $A\underset{\sim}{x} = \underset{\sim}{b}$ into upper triangular system of equations.

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ \vdots \\ b_n \end{bmatrix}$$

No coupling with other variables

Solution by back substitution

$$x_n = \frac{b_n}{a_{nn}}$$

$$a_{n-1,n-1} x_{n-1} + a_{n-1,n} x_n = b_{n-1}$$

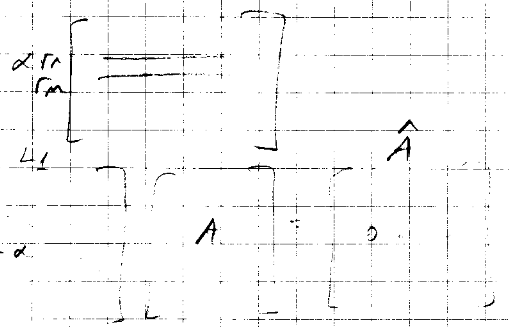
$$x_{n-1} = \frac{1}{a_{n-1,n-1}} \left(b_{n-1} - \frac{a_{n-1,n} b_n}{a_{nn}} \right)$$

(3) - A is orthogonal $\left(\text{i.e. } AA^T = I \atop A^{-1} = A^T \right)$

$$\underset{\sim}{x} = A^T \underset{\sim}{b}$$

Conversion of $A\underset{\sim}{x} = \underset{\sim}{b}$ ⇒ Factorization of A

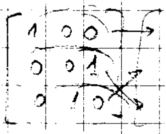
$$PA = LU$$



$$\overset{P}{L_n \cdots L_2 L_1} A = U$$

Inverse of a lower triangular matrix is always a lower triangular matrix.

$$A = \underbrace{L_1^{-1} L_2^{-1} \cdots L_n^{-1}}_{L} U$$



Not all matrices can be written as $A = LU$ so we need P.

$$\boxed{P^{-1} = P^T}$$

$$\boxed{PA = LU} \quad \boxed{A = P^T L U}$$

$$A\underset{\sim}{x} = \underset{\sim}{b}$$

$$P^T L U \underset{\sim}{x} = \underset{\sim}{b}$$

$$L \overset{\underset{\sim}{y}}{U \underset{\sim}{x}} = P \underset{\sim}{b}$$

(a1) $\underset{\sim}{L \underset{\sim}{y} = P \underset{\sim}{b}}$

Obtain $\underset{\sim}{y}$ by forward substitution.

(a2) $U \underset{\sim}{x} = \underset{\sim}{y}$

Obtain $\underset{\sim}{x}$ by back substitution.

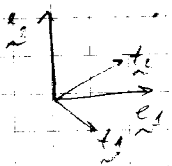
Basis and Coordinate Change

- in $\mathbb{R}^n$, the standard basis is $\{\underset{\sim}{e_1}, \dots, \underset{\sim}{e_n}\}$

$$\underset{\sim}{e_i} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \leftarrow i^{th} \text{ position}$$

- Any vector $\underset{\sim}{x} \in \mathbb{R}^n$ can be written as

$$\underset{\sim}{x} = \overbrace{x_1 \underset{\sim}{e_1} + x_2 \underset{\sim}{e_2} + \dots + x_n \underset{\sim}{e_n}}^{\text{standard coordinates}}$$



Suppose $\{\underset{\sim}{t_1}, \dots, \underset{\sim}{t_n}\}$ is another basis for $\mathbb{R}^n$.

$$\underset{\sim}{x} = \overbrace{\hat{x}_1 \underset{\sim}{t_1} + \hat{x}_2 \underset{\sim}{t_2} + \dots + \hat{x}_n t_n}^{\text{new coordinates w.r.t. basis } \{t_1, \dots, t_n\}}$$

$$= \underbrace{[\underset{\sim}{t_1}\ \underset{\sim}{t_2} \dots \underset{\sim}{t_n}]}_{T} \underbrace{\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \vdots \\ \hat{x}_n \end{bmatrix}}_{\underset{\sim}{\hat{x}}}$$

$$\underset{\sim}{x} = T \underset{\sim}{\hat{x}} \Rightarrow \underset{\sim}{\hat{x}} = T^{-1} \underset{\sim}{x}$$

Suppose we have a linear mapping $\underset{\sim}{y} = A \underset{\sim}{x}$ and we apply a basis change

$$\underset{\sim}{\hat{x}} = T^{-1} \underset{\sim}{x} \Rightarrow \underset{\sim}{x} = T \underset{\sim}{\hat{x}}$$

$$\underset{\sim}{\hat{y}} = \bigcirc\ \underset{\sim}{\hat{x}} \qquad \underset{\sim}{y} = T \underset{\sim}{\hat{y}}$$

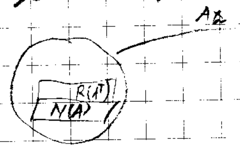
$$T \underset{\sim}{\hat{y}} = A T \underset{\sim}{\hat{x}}$$

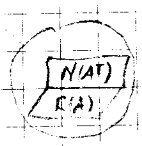
$$\underset{\sim}{\hat{y}} = \boxed{T^{-1} A T} \underset{\sim}{\hat{x}}$$

Question: Can I find a basis $\{\underset{\sim}{t_1}, \dots, \underset{\sim}{t_n}\}$ such that $T^{-1}AT$ is diagonal?

$A\underset{\sim}{x} \doteq \underset{\sim}{b} \rightarrow$ Existence

$\quad\searrow$ Uniqueness





Strategy: Convert it into simple forms ($\approx$ Matrix Factorizations for A)

1. Triangular Form $\quad PA = LU$ $\qquad$ $PA = LU$

2. Diagonal Form $\qquad\qquad$ $A = QR$ ← upper triangular; ← Orthogonal

3. Orthogonal Form $\quad Q\underset{\sim}{x} = \underset{\sim}{b} \quad Q^TQ = I$ $\qquad$ $T^{-1}AT = \Lambda \leftarrow$ diagonal

$\underset{\sim}{x} = Q^T \underset{\sim}{b}$ $\qquad$ $A = U \Sigma V^T$ ← Diagonal; Orthogonal

$\downarrow$

Simple in programming

$Q_{i,j} = Q^T_{j,i}$

Basis Change and Coordinates

$\underbrace{\{t_1, \dots, t_n\}}_{\text{new basis}}$ $\quad x = \tilde{x}_1 t_1 + \tilde{x}_2 t_2 + \dots + \tilde{x}_n t_n$

$$x = \underbrace{[t_1 \dots t_n]}_{T} \underbrace{\begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_n \end{bmatrix}}_{\tilde{x}}$$

$$x = T\tilde{x} \qquad y = Ax$$

$$y = T\tilde{y} \qquad T\tilde{y} = AT\tilde{x}$$

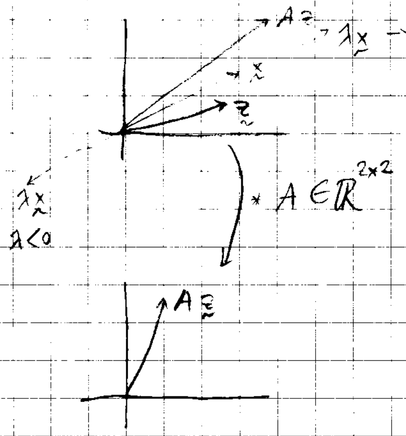
$$\tilde{y} = T^{-1}AT\tilde{x}$$

Eigenvalues – Eigenvectors

Given $A \in \mathbb{R}^{n\times n}$, λ is an eigenvalue of A if $\exists\, x \neq 0$, such that

$$Ax = \lambda x$$

x is called an eigenvector of A corresponding to λ.



How do we find eigenvalues of a given matrix?
eigenvectors

$$Ax = \lambda x$$

$$Ax - \lambda x = 0$$

$$Ax - (\lambda I)x = 0$$

$$(A - \lambda I)x = 0 \quad \Rightarrow N(A-\lambda I) \neq \{0\}$$

$\Rightarrow A - \lambda I$ is not full rank
is singular
is not invertible

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

$$\underset{\substack{\uparrow \\ \text{characteristic polynomial of } A}}{P_A(\lambda)} = \lambda^n + c_{n-1}\lambda^{n-1} + \dots + c_0 = 0$$

$\rightarrow$ Gauss : "n^{th} degree polynomial has at most n roots"
Fundamental Theorem of Algebra

by F.T.A. $P_A(\lambda)$ has n roots.

eigenvalues of $A \longleftrightarrow$ roots of $P_A(\lambda)$

If λ_i is an eigenvalue then

$(A - \lambda_i I)$ has a non-trivial null space.

Eigenspace of A corresponding to λ_i $\leftarrow$ $\boxed{N(A-\lambda_i I)}$ $\Longleftrightarrow$ eigenvectors of A corresponding to λ_i.

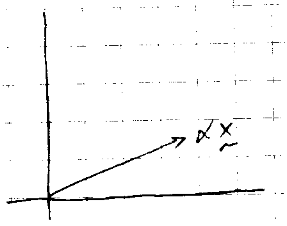
if $\underset{\sim}{x} \in N(A-\lambda_i I)$

$$(A-\lambda_i I)\underset{\sim}{x} = 0$$

$$A\underset{\sim}{x} = \lambda_i \underset{\sim}{x}$$

$$\begin{bmatrix} x & & & \\ & x & & \\ & & x & x \end{bmatrix} \quad x_3 + x_2 = ?$$

How you obtain a null Space of a matrix?



$$A\underset{\sim}{x} = \lambda \underset{\sim}{x}$$

$$A(\alpha \underset{\sim}{x}) = \alpha A \underset{\sim}{x} = \alpha \lambda \underset{\sim}{x}$$

$$A(\alpha \underset{\sim}{x}) = \lambda(\alpha \underset{\sim}{x})$$

- Suppose $\underset{\sim}{v_1}, \underset{\sim}{v_2}, \dots, \underset{\sim}{v_n}$ are eigenvectors of A which are <u>linearly independent</u>

such that $A\underset{\sim}{v_i} = \lambda_i \underset{\sim}{v_i}$

$$A[\underset{\sim}{v_1}\ \underset{\sim}{v_2}\ \dots\ \underset{\sim}{v_n}] = [A\underset{\sim}{v_1}\ A\underset{\sim}{v_2}\ \dots\ A\underset{\sim}{v_n}]$$

$$V \in \mathbb{R}^{n\times n} \quad = [\lambda_1 \underset{\sim}{v_1}\ \lambda_2 \underset{\sim}{v_2}\ \dots\ \lambda_n \underset{\sim}{v_n}]$$

$$= [\underset{\sim}{v_1}\ \underset{\sim}{v_2}\ \dots\ \underset{\sim}{v_n}] \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}}_{\Lambda}$$

$$AV = V\Lambda$$

v_i's are <u>linearly independent</u> $\Leftrightarrow$ V is invertible

$$V^{-1}AV = \Lambda \quad \Leftrightarrow \quad A = V\Lambda V^{-1}$$

- $A \in \mathbb{R}^{n\times n}$ is <u>diagonalizable</u> $\Leftrightarrow$ You can find n linearly independent eigenvectors of A.
- if A has $\overbrace{n \text{ distinct eigenvalues}}^{\text{sufficient condition not a necessary condition.}}$ then it is diagonalizable (prove this in HW).

The reverse is not necessarily true. For example $A = I$.

If A is diagonalizable and if we have $A\underset{\sim}{x} = \underset{\sim}{b}$ → Convolution is similar. You convert a complex convolution to a multiplication in the new basis.

Suppose $V = [\underset{\sim}{v_1} \dots \underset{\sim}{v_n}]$ contains eigenvectors of A

check:

$$\left.\begin{matrix} V\underset{\sim}{\tilde{x}} = \underset{\sim}{x} \\ V\underset{\sim}{\tilde{b}} = \underset{\sim}{b} \end{matrix}\right\} \quad \begin{matrix} V^{-1}AV\underset{\sim}{\tilde{x}} = \underset{\sim}{\tilde{b}} \\ \Lambda \underset{\sim}{\tilde{x}} = \underset{\sim}{\tilde{b}} \end{matrix}$$

Not necessarily: span{eigenvectors} = R(A) →

[illegible] = R(A) ↗

Signal subspace: In Hermitian matrices you can get a diagonal set of eigenvalues. If you remove the ones that are zero or small, you get a signal subspace.

- A, B are called simultaneously diagonalizable iff $\exists T$ such that

$$A = T\Lambda_1 T^{-1} \qquad B = T\Lambda_2 T^{-1}$$

$$\begin{aligned} AB &= T\Lambda_1 T^{-1} T\Lambda_2 T^{-1} \\ &= T\Lambda_1\Lambda_2 T^{-1} \\ &= T\Lambda_2\Lambda_1 T^{-1} \\ &= T\Lambda_2 T^{-1} T\Lambda_1 T^{-1} \end{aligned}$$

$AB = BA$ → commute in terms of matrix multiplication.

Orthogonal (Unitary) Matrices

Small Note about N dimensional Complex Space C^N

- C^N contains all vectors $\underset{\sim}{x}$ with N complex entries

$$\underset{\sim}{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \qquad x_k \in C \qquad x_k = a_k + i b_k \quad a_k, b_k \in \mathbb{R}$$

$$\bar{x}_k = x_k^* = a_k - i b_k$$ (complex conjugate of a given complex number)

$\underset{\sim}{x}^H = \underset{\sim}{x}^* = [x_1^* \; x_2^* \ldots x_N^*]$ Hermitian-Transpose Operation

if $A \in C^{m \times n}$

$$A^H = A^* = \underbrace{\begin{bmatrix} a_{11}^* & a_{21}^* & \cdots & a_{N1}^* \\ \vdots & & & \vdots \\ a_{1N}^* & \cdots & & a_{NN}^* \end{bmatrix}}_{\text{Transpose of } A \text{ with elements complex conjugated.}}$$

Linear Dependence:

The vectors $\underset{\sim}{v_1}, \ldots, \underset{\sim}{v_k} \in C^N$ are linearly independent if some nontrivial combination

$$c_1\underset{\sim}{v_1} + c_2\underset{\sim}{v_2} + \cdots + c_k\underset{\sim}{v_k} = \underset{\sim}{0}$$ where all c_k's are all equal to zero.

Now c_k's may be complex numbers.

$(V, R, +, \cdot)$

$(C^N, C, +, \cdot)$ → Complex vector space

- Inner Product

$$[y_1^* \cdots y_N^*]\begin{bmatrix}x_1\\ \vdots \\ x_N\end{bmatrix}$$

$$\parallel$$

For $\underset{\sim}{x}, \underset{\sim}{y} \in C^N$, $\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle = \underset{\sim}{y}^* \underset{\sim}{x} = \sum_{k=1}^{N} x_k y_k^*$

$$\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle = (\langle \underset{\sim}{y}, \underset{\sim}{x} \rangle)^*$$

$a \in C$ $\langle a\underset{\sim}{x}, \underset{\sim}{y} \rangle = a \langle \underset{\sim}{x}, \underset{\sim}{y} \rangle$

$$\langle \underset{\sim}{x}, a\underset{\sim}{y} \rangle = a^* \langle \underset{\sim}{x}, \underset{\sim}{y} \rangle$$

$$\underset{\sim}{x} \in C^N \Rightarrow \| \underset{\sim}{x} \| = \sqrt{\underset{\sim}{x}^* \underset{\sim}{x}}$$

$$= \sqrt{\langle \underset{\sim}{x}, \underset{\sim}{x} \rangle} = \sqrt{\sum_{k=1}^{n} \underset{\substack{\downarrow \\ x_k x_k^*}}{|x_k|^2}} \text{ = magnitude}$$

$$A\underset{\sim}{x} = \underset{\sim}{b}$$

Easy cases:

1. Diagonal
2. Upper Triangular
3. Orthogonal Matrix

$$A^{-1} = A^T, \quad \underset{\sim}{x} = A^T \underset{\sim}{b}$$

Harder

- Orthogonal (Unitary)
- Special matrices (Normal matrices)
  (Unitary, Hermitian)

Complex Inner Product

- $\underset{\sim}{x}, \underset{\sim}{y} \in C^n$

$$\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle = \underset{\sim}{y}^{*H} \underset{\sim}{x} = \sum_{k=1}^{n} x_k y_k^* \qquad \underset{\sim}{y}^* = [y_1^* \cdots y_n^*]$$

- $\underset{\sim}{x}, \underset{\sim}{y} \in C^n$ is orthogonal $\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle = 0$
- $A \in C^{m \times n}$ if

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{m1} & \cdots & \cdots & a_{mn} \end{bmatrix}$$

$$A^H = A^* = \begin{bmatrix} a_{11}^* & a_{21}^* & \cdots & a_{m1}^* \\ a_{12}^* & & & \\ & & & \\ a_{1n}^* & & & a_{mn}^* \end{bmatrix}$$

Orthogonal Vectors

- (Definition): A pair of vectors $x, y \in \mathbb{C}^n$ are orthogonal if

$$x^* y = y^* x = 0 = \langle x, y \rangle = \langle y, x \rangle$$

In general, they are not equal but they are 0

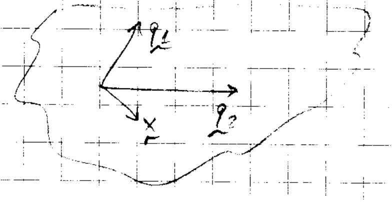
- In the real 2-3 dimensional vectors orthogonal $\Longleftrightarrow$ right angle between the vectors.
- Two sets $X \subset \mathbb{C}^n$, $Y \subset \mathbb{C}^n$ are orthogonal if every $x \in X$ is orthogonal to every $y \in Y$.
- A set of vectors $S \subset \mathbb{C}^n$ is orthogonal iff its elements are pairwise orthogonal.

$$x, y \in \mathbb{C}^n, \; x \neq y \Rightarrow \langle x, y \rangle = 0$$

- A set of vectors $S \subset \mathbb{C}^n$ is orthonormal iff S is orthogonal and every $x \in S$ has $\|x\| = 1$. (Remember $\|x\| = \sqrt{x^* x}$)
- If $x \in \operatorname{Span}\{q_1, \dots, q_n\}$ where $x, q_1, \dots, q_n \in \mathbb{C}^m$ and $\{q_1, \dots, q_n\}$ is orthonormal.

$$\langle q_i, q_j \rangle = \delta_{i-j}$$

Kronecker delta function



$$x = \alpha_1 q_1 + \alpha_2 q_2 + \dots + \alpha_n q_n = [q_1 \dots q_n] \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$$

Take the inner product of both sides with q_1

$$\langle x, q_1 \rangle = \langle \alpha_1 q_1 + \dots + \alpha_n q_n, q_1 \rangle$$

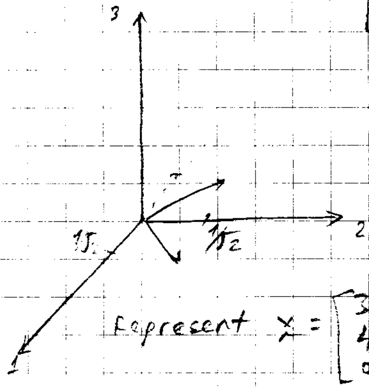
$$= \langle \alpha_1 q_1, q_1 \rangle + \langle \alpha_2 q_2, q_1 \rangle + \dots + \langle \alpha_n q_n, q_1 \rangle$$

$$= \alpha_1 \langle q_1, q_1 \rangle$$

$$\alpha_1 = \frac{\langle x, q_1 \rangle}{\langle q_1, q_1 \rangle} = \langle x, q_1 \rangle$$

Similarly, $\alpha_i = \langle x, q_i \rangle$

Example: Let $S = \left\{ \overset{\underset{\sim}{q_i}}{\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}}, \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \right\}$



$$\left.\begin{array}{l} \langle \underset{\sim}{q_1}, \underset{\sim}{q_2} \rangle = -\frac{1}{2} + \frac{1}{2} = 0 \\ \|\underset{\sim}{q_i}\| = \sqrt{\frac{1}{2} + \frac{1}{2}} = 1 \end{array}\right\} S \text{ is orthonormal}$$

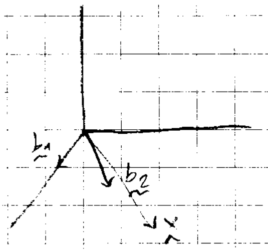
Represent $\underset{\sim}{x} = \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix} = \alpha_1 \underset{\sim}{q_1} + \alpha_2 \underset{\sim}{q_2}$

$$\alpha_1 = \langle \underset{\sim}{x}, \underset{\sim}{q_1} \rangle = 3 \frac{1}{\sqrt{2}} + 4 \frac{1}{\sqrt{2}} = \frac{7}{\sqrt{2}}$$

$$\alpha_2 = \langle \underset{\sim}{x}, \underset{\sim}{q_2} \rangle = 3 \left(-\frac{1}{\sqrt{2}}\right) + 4 \frac{1}{\sqrt{2}} = \frac{1}{\sqrt{2}}$$

$$\underset{\sim}{x} = \frac{7}{\sqrt{2}} \underset{\sim}{q_1} + \frac{1}{\sqrt{2}} \underset{\sim}{q_2}$$

What if we had basis $\left\{ \overset{\underset{\sim}{b_1}}{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}}, \overset{\underset{\sim}{b_2}}{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}} \right\}$



$$\underset{\sim}{x} = \beta_1 \underset{\sim}{b_1} + \beta_2 \underset{\sim}{b_2}$$

Note: $\beta_i \neq \dfrac{\langle \underset{\sim}{x}, \underset{\sim}{b_i} \rangle}{\langle \underset{\sim}{b_i}, \underset{\sim}{b_i} \rangle}$

$$\underset{\sim}{x} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}$$

$$\beta_2 = 4, \quad \beta_1 = -1$$

$$\langle \underset{\sim}{x}, \underset{\sim}{b_1} \rangle = 3 \qquad \langle \underset{\sim}{x}, \underset{\sim}{b_2} \rangle = 7$$

$$\langle \underset{\sim}{x}, \underset{\sim}{b_1} \rangle = \beta_1 \langle \underset{\sim}{b_1}, \underset{\sim}{b_1} \rangle + \beta_2 \langle \underset{\sim}{b_2}, \underset{\sim}{b_1} \rangle$$

$$\langle \underset{\sim}{x}, \underset{\sim}{b_2} \rangle = \beta_2 \langle \underset{\sim}{b_1}, \underset{\sim}{b_2} \rangle + \beta_2 \langle \underset{\sim}{b_2}, \underset{\sim}{b_2} \rangle$$

$$3 = \beta_1 \cdot 1 + \beta_2 \cdot 1$$

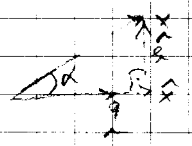
$$7 = \beta_2 \cdot 1 + \beta_2 \cdot 2$$

$$\Rightarrow \begin{bmatrix} 3 \\ 7 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}}_{A} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

Using Cramer's rule: $\dfrac{\det [illegible]}{\det \begin{bmatrix} 1 & 1 \\ [illegible] & [illegible] \end{bmatrix}} = \beta_1 = -1$

$\beta_2 = 4$

– Orthogonal projection of a vector over another vector



$$\hat{x} = \frac{\underset{\sim}{q}}{\|\underset{\sim}{q}\|}\left(\|\underset{\sim}{x}\| . \cos\alpha\right) \qquad \cos\alpha = \frac{\langle \underset{\sim}{x}, \underset{\sim}{q} \rangle}{\|\underset{\sim}{x}\|\,\|\underset{\sim}{q}\|}$$

$$\hat{x} = \frac{\underset{\sim}{q}}{\|\underset{\sim}{q}\|}\left(\frac{\cancel{\|\underset{\sim}{x}\|}\langle \underset{\sim}{x}, \underset{\sim}{q} \rangle}{\cancel{\|\underset{\sim}{x}\|}\,\|\underset{\sim}{q}\|}\right)$$

$$\hat{x} = \underset{\sim}{q}\,\frac{\langle \underset{\sim}{x}, \underset{\sim}{q} \rangle}{\langle \underset{\sim}{q}, \underset{\sim}{q} \rangle}$$

When we write (when $\{\underset{\sim}{q_1}, \dots, q_n\}$ is an orthogonal set)

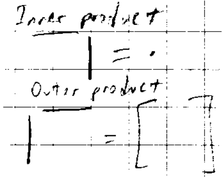
$$\underset{\sim}{x} = \underbrace{\frac{\langle \underset{\sim}{x}, \underset{\sim}{q_1} \rangle}{\langle \underset{\sim}{q_1}, \underset{\sim}{q_1} \rangle}\underset{\sim}{q_1}}_{\text{Orthogonal projection of } \underset{\sim}{x} \text{ to } \underset{\sim}{q_1}} + \underbrace{\frac{\langle \underset{\sim}{x}, \underset{\sim}{q_2} \rangle}{\langle \underset{\sim}{q_2}, \underset{\sim}{q_2} \rangle}\underset{\sim}{q_2}}_{\text{Orthogonal projection of } \underset{\sim}{x} \text{ to } \underset{\sim}{q_2}} + \dots \underbrace{\frac{\langle \underset{\sim}{x}, \underset{\sim}{q_n} \rangle}{\langle \underset{\sim}{q_n}, \underset{\sim}{q_n} \rangle}\underset{\sim}{q_n}}_{\text{Orthogonal projection of } \underset{\sim}{x} \text{ to } \underset{\sim}{q_n}}$$

-- --

In the orthonormal case

$$\begin{aligned}
\underset{\sim}{x} &= \langle \underset{\sim}{x}, \underset{\sim}{q_1} \rangle \underset{\sim}{q_1} + \langle \underset{\sim}{x}, \underset{\sim}{q_2} \rangle \underset{\sim}{q_2} + \dots + \langle \underset{\sim}{x}, \underset{\sim}{q_n} \rangle \underset{\sim}{q_n} \\
&= (q_1 \underset{\sim}{q_1^*})\underset{\sim}{x} + (q_2 \underset{\sim}{q_2^*})\underset{\sim}{x} + \dots + (q_n q_n^*)\underset{\sim}{x} \\
&= \sum_{k=1}^{n} (q_k q_k^*)\,\underset{\sim}{x} = \sum_{k=1} Q_k \underset{\sim}{x} \quad \text{where } Q_k = \underset{\sim}{q_k}\underset{\sim}{q_k^*}
\end{aligned}$$

$$Q_k \underset{\sim}{x} = \underbrace{\underset{\sim}{q_k}(\underset{\sim}{q_k^*}\underset{\sim}{x})}_{\text{Projection of } \underset{\sim}{x} \text{ over } \underset{\sim}{q_k}}$$



$$Q_k = \underset{\sim}{q_k}\,\underset{\sim}{q_k^*}$$

$$= \begin{bmatrix} q_{k1} \\ \vdots \\ q_{km} \end{bmatrix} \begin{bmatrix} q_{k1}^* & \cdots & q_{km}^* \end{bmatrix} = \underbrace{\begin{bmatrix} |q_{k1}|^2 & q_{k1} q_{k2}^* & \\ q_{k2} q_{k1}^* & |q_{k2}|^2 & \\ & & \ddots \end{bmatrix}}_{\text{Hermitian matrix}}$$

→ covariance matrix is also a Hermitian matr

Q_k has rank 1

Actually any outer matrix $\underset{\sim}{x}\underset{\sim}{y}^*$ has rank 1.

$$= \begin{bmatrix} q_{k1} \\ \vdots \\ q_{km} \end{bmatrix} \underset{\sim}{q_k^*} = \begin{bmatrix} q_{k1}\,\underset{\sim}{q_k^*} \\ q_{k2}\,\underset{\sim}{q_k^*} \\ \vdots \\ q_{km}\,\underset{\sim}{q_k^*} \end{bmatrix}$$

$$= \begin{bmatrix} q_{k1}^*\underset{\sim}{q_k} & q_{k2}^*\underset{\sim}{q_k} & \cdots & q_{km}^*\underset{\sim}{q_k} \end{bmatrix}$$

Midterm: Nov. 15th Thursday

Orthogonal and Unitary Matrices

A set $\{q_1, \dots, q_k\}$ is orthogonal iff

$$\langle q_i, q_j \rangle = D_i \delta_{i-j} \qquad \delta_k = \begin{cases} 1 & k=0 \\ 0 & \text{otherwise} \end{cases}$$

$$D_i = \|q_i\|^2$$

If $D_i = 1$, the set is orthonormal

Unitary Matrices

A matrix $U \in C^{n \times n}$ is unitary iff

$$U^* U = U U^* = I \Leftrightarrow U^* = U^{-1}$$

$Ux = b$ * It is easy to solve linear system of equations with

$x = U^* b$ unitary matrices.

$U = [u_1 \cdots u_n]$

$$U^* = \begin{bmatrix} u_1^* \\ \vdots \\ u_n^* \end{bmatrix} \qquad U^* U = \begin{bmatrix} u_1^* \\ \vdots \\ u_n^* \end{bmatrix} \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} \langle v_1, v_1 \rangle & \langle v_2, v_1 \rangle & \\ \langle v_1, v_2 \rangle & \langle v_2, v_2 \rangle & \\ & & \ddots \end{bmatrix} = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}$$

$\langle v_i, v_j \rangle = \delta_{i-j} \Leftrightarrow$ the columns of U is an orthonormal set

Actually $\underbrace{\{u_1 \quad u_n\}}_{\text{columns (rows)}}$ forms an orthonormal basis for C^n.

So for any $x \in C^n$

$$x = \alpha_1 u_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$$

$$= [u_1 \quad u_n] \underbrace{\begin{bmatrix} \alpha_1 \\ \\ \alpha_n \end{bmatrix}}_{\hat{x}}$$

$$x = U \hat{x}$$

$$\hat{x} = U^* x = \begin{bmatrix} u_1^* \\ \\ u_n^* \end{bmatrix} x = \begin{bmatrix} \langle x, u_1 \rangle \\ \langle x, u_2 \rangle \\ \vdots \\ \langle x, u_n \rangle \end{bmatrix} \qquad \alpha_i = \langle x, u_i \rangle$$

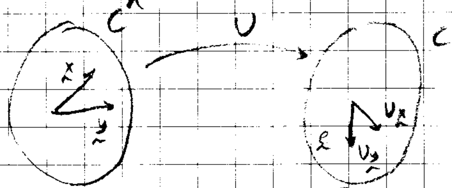
$$\langle x, u_i \rangle = \alpha_1 \langle u_1, u_i \rangle + \alpha_2 \langle v_2, u_i \rangle + \cdots + \alpha_n \langle u_n, u_i \rangle$$

$$= \alpha_i$$

Some Properties

- Let $x, y \in C^n$ and U is unitary

$$\langle Ux, Uy\rangle = (Uy)^* Ux = y^* U^* U x = y^* x = \langle x, y\rangle$$

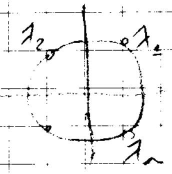


Angles are preserved under the unitary mapping.

- $\|Ux\| = \sqrt{\langle Ux, Ux\rangle} = \sqrt{\langle x, x\rangle} = \|x\|$ Lengths are preserved under the unitary mapping.

For general linear mapping, e can be small in the target domain but large in the original domain. → Important thing to check for Stretch

- Eigenvalue λ of U has the property that $|\lambda| = 1$.



Proof: Let λ be an eigenvalue of U and let v be an eigenvector corresponding to λ, i.e.,

$$Uv = \lambda v$$

$$\|Uv\| = \|\lambda v\|$$

$$\cancel{\|v\|} = |\lambda| \cancel{\|v\|}$$

$$|\lambda| = 1.$$

Orthogonal Matrices (⊂ Unitary Matrices)

$Q \in \mathbb{R}^{n\times n}$ is real orthogonal iff

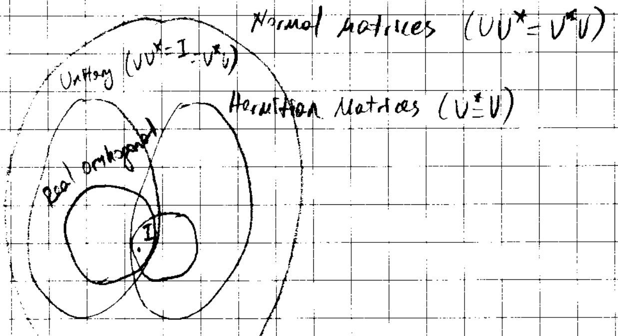
$$Q^T Q = Q Q^T = I$$

Orthogonal Matrix = Real unitary Matrix

Columns (rows) of Q form an orthonormal basis for $\mathbb{R}^n$.

General orthogonal matrices do not necessarily have unit norm. → $Q^T Q = Q Q^T = \Lambda$

A diagonal



(Real) Symmetric Matrices

$A \in \mathbb{R}^{n\times n}$ is Symmetric iff

$$A = A^T$$

Hermitian Matrices

$A \in \mathbb{C}^{n\times n}$ is Hermitian iff $A = A^*$.

- Properties $\quad -[\;]\,1$ $\qquad$ Covariance matrices are Hermitian.

* For all $x \in \mathbb{C}^n$ $\quad \underbrace{\overset{1\times n}{x^*}\overset{n\times n}{A}\overset{n\times 1}{x}}_{\text{quadratic in } x} \in \mathbb{R}$

(quadratic functions) A real number's complex conjugate is itself.

A generalization of $y = ax^2+bx+c$

Proof: $(x^*Ax)^* = x^* A^* x$

$$= x^* A x \quad \blacksquare$$

* Every eigenvalue of a Hermitian matrix is real.

Proof: λ, v $\qquad$ λ_1 | λ_2 ?

$Av = \lambda v$

$\underbrace{v^*Av}_{\text{real}} = \overline{\lambda}\underbrace{v^*v}_{\text{real}}$ — should be real

* Eigenvectors of a Hermitian matrix, if they correspond to different eigenvalues are orthogonal to one another.

Proof: $Ax = \lambda_1 x \quad Ay = \lambda_2 y \quad A = A^*$

$y^*Ax = \lambda_1 \langle x, y\rangle \quad x^*Ay = \lambda_2 \langle y, x\rangle$

$x^*Ay = \lambda_1 \langle y, x\rangle$

$\lambda_2 \langle y, x\rangle = \lambda_1 \langle y, x\rangle$ so when $\lambda_1 \neq \lambda_2 \Rightarrow \langle x, y\rangle = 0$

Normal Matrices

· $A \in \mathbb{C}^{n\times n}$ is normal iff $AA^* = A^*A$

- Since $U^*U = UU^* = I$, unitary matrices are normal.
- Since $A^*A = AA^*$ for $A = A^*$, Hermitian matrices are normal
- If $A^* = -A$, A is called skew-Hermitian. Skew-Hermitian matrices are normal.

$$\begin{bmatrix} j & -1+i \\ 1+i & 2j \end{bmatrix} \begin{bmatrix} -j & 1-i \\ -1-i & -2j \end{bmatrix}$$

Skew-Hermitian $\quad Av = \lambda v$

$v^*Av = \lambda v^*v$

$v^*A^*v = \lambda^* v^*v$

$-v^*Av = \lambda^* v^*v \rightarrow \lambda^* = -\lambda$

Thus, skew-Hermitian matrices have purely imaginary eigenvalues.

- If $A \in C^{n\times n}$ has eigenvalues $\lambda_1, \ldots, \lambda_n$, the following statements are equivalent:

(a) A is normal

(b) A is unitarily diagonalizable $\quad U^{-1}AU = D \Rightarrow U^*AU = D$

$A = UDU^*$

Unitary matrices are unitarily diagonalizable

(c) $\sum_{i,j} |a_{i,j}|^2 = \sum_{i=1}^{n} |\lambda_i|^2$

(d) There is an orthonormal set of eigenvectors

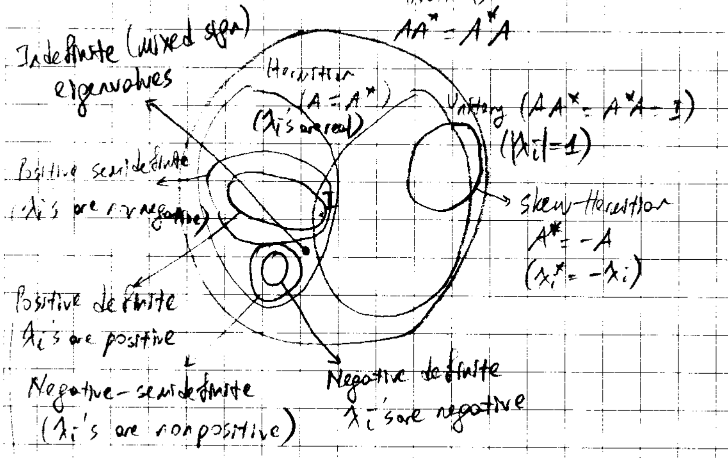
$$AU = UD$$

$$A[\underset{\sim}{v_1} \ldots \underset{\sim}{v_n}] = [\underset{\sim}{v_1} \ldots \underset{\sim}{v_n}]D \qquad D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

$$A\underset{\sim}{v_i} = \lambda_i \underset{\sim}{v_i}$$

$$\sum_{i,j} |a_{i,j}|^2 = Tr(A^*A)$$

Normal Matrices



Normal Matrices

1) A is unitarily diagonalizable

$A = UDU^*$ $\quad$ D → diagonal

$UU^* = I$

2) A has orthonormal set of eigenvectors

$$AU = UD$$

$$A[\underset{\sim}{v_1} \ldots \underset{\sim}{v_n}] = [\underset{\sim}{v_1} \ldots \underset{\sim}{v_n}]\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

$$A\underset{\sim}{v_i} = \lambda_i \underset{\sim}{v_i}$$

3) $\sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|^2 = \sum_{i=1}^{n} |\lambda_i|^2$

To prove 3: Use the following lemma

* Quadratic functions with parabola up

Lemma: For any $A \in C^{n\times n}$, $Tr(A^*A) = \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|^2$

Proof: $A = [\underset{\sim}{a_1} \ldots \underset{\sim}{a_n}]$

$$A^* = \begin{bmatrix} \underset{\sim}{a_1}^* \\ \vdots \\ \underset{\sim}{a_n}^* \end{bmatrix} \qquad A^*A = \begin{bmatrix} \underset{\sim}{a_1}^* \\ \vdots \\ \underset{\sim}{a_n}^* \end{bmatrix} [\underset{\sim}{a_1} \ldots \underset{\sim}{a_n}]$$

$$= \begin{bmatrix} \underset{\sim}{a_1}^*\underset{\sim}{a_1} & & X \\ & \underset{\sim}{a_2}^*\underset{\sim}{a_2} & \\ X & & \ddots \\ & & & \underset{\sim}{a_n}^*\underset{\sim}{a_n} \end{bmatrix}$$

$$Tr(A^*A) = \underset{\sim}{a_1}^*\underset{\sim}{a_1} + \ldots + \underset{\sim}{a_n}^*\underset{\sim}{a_n} = \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|^2$$

Show that $1 \Rightarrow 3$

$$A = UDU^*$$

$$\sum_{i,j}^{n} |a_{ij}|^2 = Tr(A^*A) = Tr(\underbrace{UD^*U^*}_{A^*}\underbrace{UDU^*}_{A})$$

$$= Tr(\underbrace{UD^*}_{X}\underbrace{DU^*}_{Y})$$

$$D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_n \end{bmatrix} \quad D^* = \begin{bmatrix} \lambda_1^* & 0 \\ 0 & \lambda_n^* \end{bmatrix}$$

$$D^*D = \begin{bmatrix} |\lambda_1|^2 & 0 \\ 0 & |\lambda_1|^2 \end{bmatrix}$$

Trace Property

If the sizes are compatible

$$Tr(XY) = Tr(YX)$$

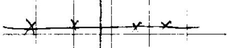
$$= Tr(U^*UD^*D)$$

$$= Tr(D^*D) = Tr\left(\begin{bmatrix} |\lambda_1|^2 & 0 \\ 0 & |\lambda_n|^2 \end{bmatrix}\right)$$

$$= \sum_{k=1}^{n} |\lambda_k|^2$$

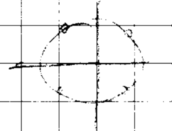
A is normal $\Longleftrightarrow A = UDU^*$

→ If λ_i's are real $A^* = UD^*U^*$



$= UDU^* = A$ (Hermitian)

→ If $|\lambda_i| = 1$ $AA^* = UDU^*UD^*U^*$



$= U\underbrace{DD^*}_{I}U^* = I$ (Unitary)

→ If $\lambda_i^* = -\lambda_i$ $A = -A^*$ $\Longleftrightarrow$ skew hermitian

Analogy between $C^{n\times n} \rightarrow C$

* Hermitian matrices are analogous to real numbers. Any $A \in C^{n\times n}$ can be written

as $A = R + iJ$ $R = R^*$, $J = J^*$

$$R = \frac{A + A^*}{2} \quad J = \frac{A - A^*}{2i} \quad R + iJ = A$$

* Unitary matrices are analogous to complex numbers on unit circle

Remember any complex number $x \in C$, $x \neq 0$ can be written as

$$x = \underbrace{r}_{\text{Positive real}}\ \underbrace{e^{j\theta}}_{\text{unit circle}} \quad \text{(Polar form)}$$

Similarly any $A \in C^{n\times n}$ (full rank)

$$A = \underbrace{R}_{\text{Hermitian matrix with positive eigenvalues}}\ \underbrace{U}_{\text{unitary matrix}}$$

= Positive definite Hermitian matrix

} Polar form

* A Hermitian matrix $A \in \mathbb{C}^{n \times n}$ is positive (negative) definite if all eigenvalues of A are strictly positive (negative)

following are equivalent

* A is positive definite

* $\lambda_i > 0 \quad i = 1, \dots, n$

* For any $\underset{\sim}{x} \neq 0 \in \mathbb{C}^n$, $\underset{\sim}{x}^* A \underset{\sim}{x} > 0$ → positive real number

$A = UDU^*$

$$\underset{\sim}{x}^* A \underset{\sim}{x} = \underset{\sim}{x}^* UD \underbrace{U^* \underset{\sim}{x}}_{\underset{\sim}{y}} = \underset{\sim}{y}^* D \underset{\sim}{y} = [y_1^* \dots y_n^*] \begin{bmatrix} \lambda_1 y_1 \\ \vdots \\ \lambda_n y_n \end{bmatrix}$$

$$= \lambda_1 |y_1|^2 + \lambda_2 |y_2|^2 + \dots + \lambda_n |y_n|^2 > 0$$

U is always nonsingular since $\det(U) \neq 0$. ⇒ Always invertible.

$\underset{\sim}{y} = U^* \underset{\sim}{x}$

* A Hermitian matrix $A \in \mathbb{C}^{n \times n}$ is positive (negative) semidefinite if all eigenvalues of A are nonnegative (nonpositive)

Quadratic Function (Single variable → multivariable)

* Single variable quadratic functions

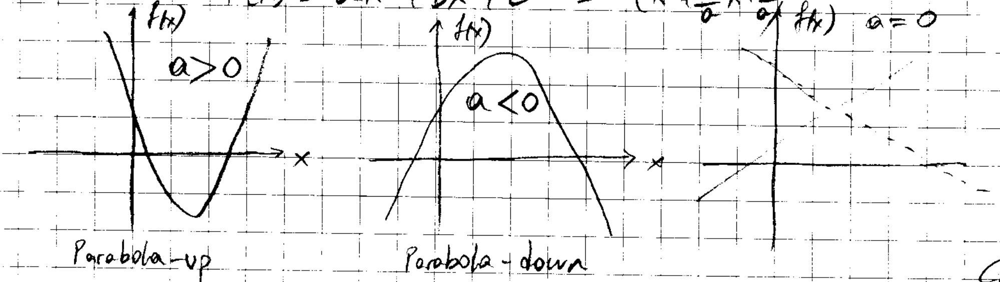
$$f(x) = ax^2 + bx + c = a\left(x^2 + \frac{b}{a}x + \frac{c}{a}\right) = a\left(x^2 + \frac{b}{a}x + \frac{b^2}{4a^2} + \frac{c}{a} - \frac{b^2}{4a^2}\right)$$

$$= a\left(x + \frac{b}{2a}\right)^2 + c - \frac{b^2}{4a}$$

$$= \left(x + \frac{b}{2a}\right)^T a \left(x + \frac{b}{2a}\right) + c -$$

⇓

Completion to square



* Two variable quadratic functions

$$f(x,y) = f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = ax^2 + bxy + cy^2 + d$$

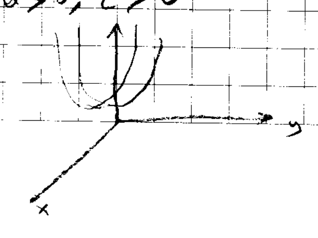
Let's assume $b = 0$ for simplicity

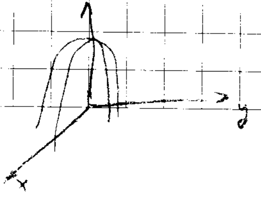
$$f(x,y) = ax^2 + cy^2 + d = \underset{\underset{\sim}{z}^T}{[x \ y]} \underset{A}{\begin{bmatrix} a & 0 \\ 0 & c \end{bmatrix}} \underset{\underset{\sim}{z}}{\begin{bmatrix} x \\ y \end{bmatrix}} + d$$

$a > 0,\ c > 0$ $\qquad$ $a < 0,\ c < 0$

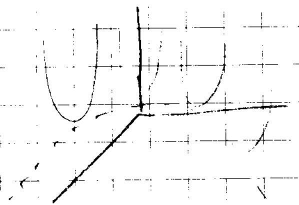
↘ If A is positive definite, Parabola-up

↘ If A is negative definite Parabola-down





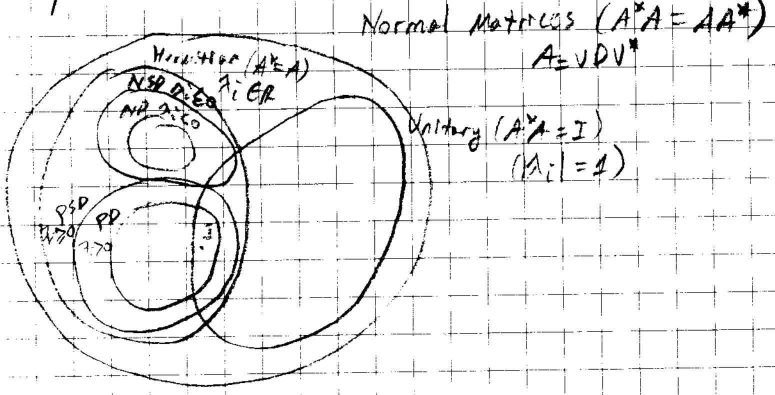
$a > 0$, $c < 0$ → A is indefinite



→ Matlab's mesh function ile çizilebilir.

Saddle structure

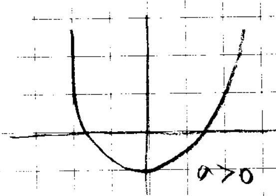
Special Matrices

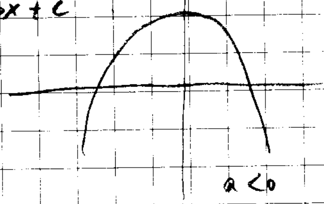


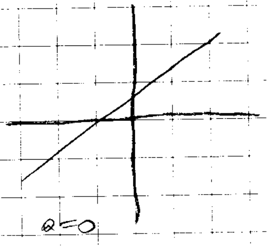
Quadratic Function

- Single variable case

$f(x) = ax^2 + bx + c$







- Two variable case

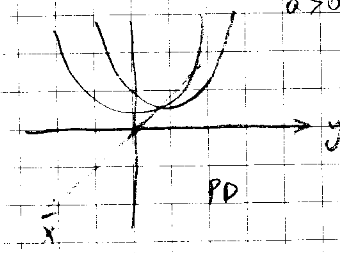
$$f(x,y) = f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = ax^2 + bxy + cy^2 + d$$

Assuming $b = 0$

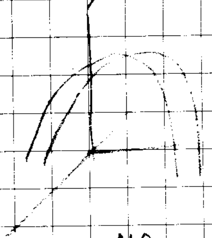
$$f(x,y) = ax^2 + cy^2 + d$$

→ Draw this with matlab

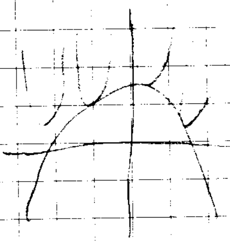
$a > 0$, $c > 0$



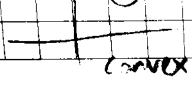
$a < 0$, $c < 0$

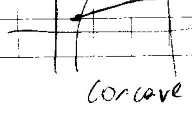


$a > 0$, $c < 0$



$$f(x,y) = [x \ \ y]\begin{bmatrix} a & 0 \\ 0 & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + d$$





The case $b \neq 0$

$$f(x,y) = ax^2 + bxy + cy^2 + d = [x\ y]\begin{bmatrix} a & \alpha \\ \beta & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + d$$

$$ax^2 + cy^2 + \alpha xy + \beta xy + d$$

$$\alpha + \beta = b$$

$$= [x\ y]\underbrace{\begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}}\underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\underset{\sim}{x}} + d$$

Symmetric matrix ⟹ PD or ND or Indefinite

$$= \underset{\sim}{x}^T A \underset{\sim}{x} + b \qquad A = A^T$$

Side note: For $\underset{\sim}{x}^T A \underset{\sim}{x}$ where $A \neq A^T$, we can show that $\underset{\sim}{x}^T A \underset{\sim}{x} = \underset{\sim}{x}^T \left(\frac{A + A^T}{2}\right)\underset{\sim}{x}$

Proof: $A = \underbrace{\frac{A + A^T}{2}}_{\text{Symmetric}} + \underbrace{\frac{A - A^T}{2}}_{\text{Skew symmetric}}$

Symmetric component,

$$\underset{\sim}{x}^T A \underset{\sim}{x} = \underset{\sim}{x}^T\left(\frac{A + A^T}{2}\right)\underset{\sim}{x} + \underset{\sim}{x}^T\left(\frac{A - A^T}{2}\right)\underset{\sim}{x}$$

$$A = \begin{bmatrix} a & b/4 \\ 3b/4 & c \end{bmatrix} \qquad \frac{A + A^T}{2} = \begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}$$

Scalar, so its transpose should be equal to itself.

$$\frac{\underset{\sim}{x}^T A \underset{\sim}{x}}{2} - \frac{\underset{\sim}{x}^T A^T \underset{\sim}{x}}{2}$$

$$\cancel{\frac{\underset{\sim}{x}^T A^T \underset{\sim}{x}}{2}} - \cancel{\frac{\underset{\sim}{x}^T A^T \underset{\sim}{x}}{2}}$$

$$= 0$$

* You can work with the symmetric component of A in $\underset{\sim}{x}^T A \underset{\sim}{x}$.

$$f(\underset{\sim}{x}) = f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \underset{\sim}{x}^T \underbrace{\begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}}\underset{\sim}{x} + d$$

Symmetric, Real Hermitian matrix ⟹ I can diagonalize it.

Since $A = A^T$, $A = Q \Lambda Q^T$ $\quad \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$ $\quad \lambda_1, \lambda_2 \in \mathbb{R}$

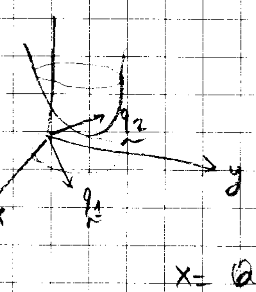
$$QQ^T = I$$

$$f(\underset{\sim}{x}) = \underset{\sim}{x}^T Q \Lambda Q^T \underset{\sim}{x} + d$$

$$\underset{\sim}{z} = Q^T \underset{\sim}{x}$$

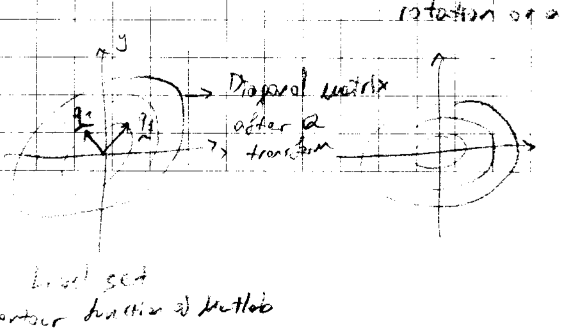
$$f(\underset{\sim}{x}) = f(\underset{\sim}{z}) = \underset{\sim}{z}^T \Lambda \underset{\sim}{z} + d = [z_1\ z_2]\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + d$$

$$= \lambda_1 z_1^2 + \lambda_2 z_2^2 + d$$



$Q = [\underset{\sim}{q_1}\ \underset{\sim}{q_2}]$ → orthonormal matrix → It's always a rotation or a reflection

$$\underset{\sim}{x} = Q\underset{\sim}{z} = z_1 \underset{\sim}{q_1} + z_2 \underset{\sim}{q_2}$$



→ Diagonal matrix after Q transform

Level set

Contour function of Matlab

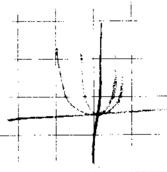
If A is PD $\Leftrightarrow f(\underset{\sim}{x})$ is parabola up (convex)

A is ND $\Leftrightarrow f(\underset{\sim}{x})$ is parabola down

A is indefinite $\Leftrightarrow f(\underset{\sim}{x})$ is saddle structure

n-variable quadratic functions (centered around origin)

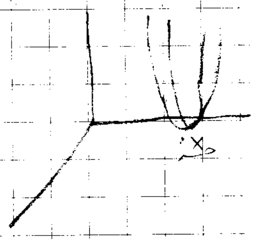
$\underset{\sim}{x} \in \mathbb{R}^n$



$$f(\underset{\sim}{x}) = \underset{\sim}{x}^T A \underset{\sim}{x} + d$$

quadratic function centered around $\underset{\sim}{x}_0$

$$f(\underset{\sim}{x}) = (\underset{\sim}{x} - \underset{\sim}{x}_0)^T A (\underset{\sim}{x} - \underset{\sim}{x}_0) + d$$



Level Sets

$$L_\alpha = \{\underset{\sim}{x} \mid f(\underset{\sim}{x}) = \alpha\}$$

$$f(\underset{\sim}{x}) = \alpha \Rightarrow \underset{\sim}{x}^T A \underset{\sim}{x} + d = \alpha$$

$\alpha \geq d$

In the two variable case

$$\underset{\sim}{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad ax_1^2 + bx_1x_2 + cx_2^2 + d = \alpha$$

if A is PD

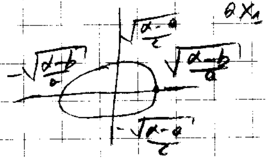
so $\underset{\sim}{x}^T A \underset{\sim}{x} + d = \alpha$

is an ellipse when A is PD.

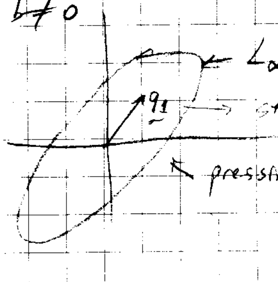
if $b = 0$

$a > 0,\ c > 0$

$$ax_1^2 + cx_2^2 + d = \alpha$$



if $b \neq 0$

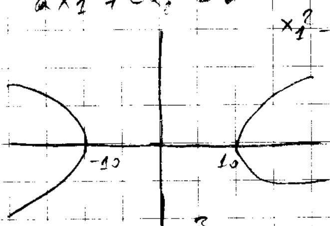


Suppose $b = 0$

$a > 0,\ c < 0$

$ax_1^2 + cx_2^2 = \alpha'$

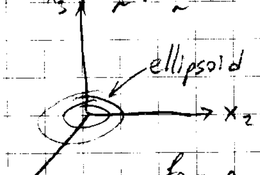
$x_1^2 - 2x_2^2 = 100$



← Try this in MATLAB

when $\underset{\sim}{x} \in \mathbb{R}^3$

$\underset{\sim}{x}^T A \underset{\sim}{x} + d = \alpha$ when A is PD



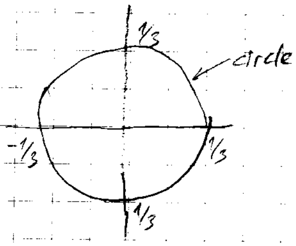
for n dimensional case we have hyperellipse

Examples

1. Let $A = \begin{bmatrix} 9 & 0 \\ 0 & 9 \end{bmatrix}$

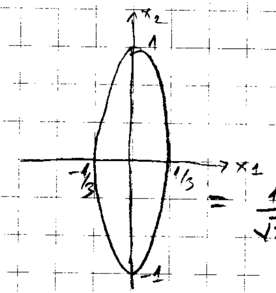
$\underset{\sim}{x}^T A \underset{\sim}{x} = 1$

$[x_1 \ x_2] \begin{bmatrix} 9 & 0 \\ 0 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 1 \qquad 9x_1^2 + 9x_2^2 = 1$



2. $A = \begin{bmatrix} \overset{\lambda_1}{9} & 0 \\ 0 & 1 \end{bmatrix} \leftarrow \lambda_2$

$\underset{\sim}{x}^T A \underset{\sim}{x} = 1 \qquad 9x_1^2 + x_2^2 = 1$



$= \frac{1}{\sqrt{\lambda_1}}$

3. $A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$ eigenvectors Let's find eigenvalues and eigenvectors

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

$\det\left(\begin{bmatrix} \lambda-5 & -4 \\ -4 & \lambda-5 \end{bmatrix}\right) = 0 \qquad \lambda^2 - 10\lambda + 25 - 16 = 0$

$\lambda^2 - 10\lambda + 9 = 0$

$\lambda_1 = 9 \quad \lambda_2 = 1$

Eigenspace for $\lambda_1 = 9 = N(9I - A)$

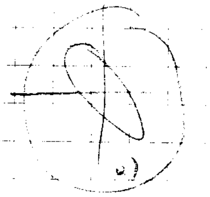
$= N\left(\begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix}\right) = \left\{ t \begin{bmatrix} 1 \\ 1 \end{bmatrix} \middle| t \in \mathbb{R} \right\}$

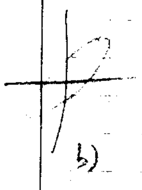
$\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \rightarrow$ unit norm in that direction

Eigenspace for $\lambda_2 = 1 = N(I - A)$

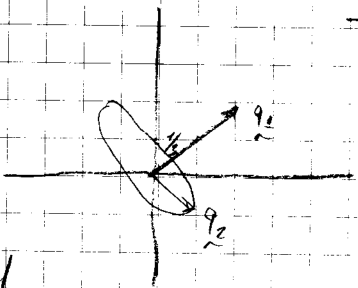
$= N\left(\begin{bmatrix} -4 & -4 \\ -4 & -4 \end{bmatrix}\right) = \left\{ t \begin{bmatrix} 1 \\ -1 \end{bmatrix} \middle| t \in \mathbb{R} \right\}$

$\begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$





$Q = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \qquad \Lambda = \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix}$



$A = Q \Lambda Q^T$

Eigenvectors of A defines the principal axes of the ellipse.

Special Matrices (Positive Definite Matrices)

A matrix $A = A^* \in \mathbb{C}^{n \times n}$ is PD if for all $\underset{\sim}{x} \neq \underset{\sim}{0} \in \mathbb{C}^n$

$\underset{\sim}{x}^* A \underset{\sim}{x} > 0$

- A is PD $\Longleftrightarrow$ $\lambda_i(A) \in \mathbb{R} > 0$, $A = A^*$

$$A \text{ is PD} \Longleftrightarrow A \overset{\prime}{\underset{\succ}{>}} 0$$ generalized inequality

$A \geq B$ means $A-B$ is positive semidefinite

$\leq$ negative - - -

$A > B$ means $A-B$ is positive definite

$<$ negative - - -

If $A > 0$, then $f(\underset{\sim}{x}) = \underset{\sim}{x}^* A \underset{\sim}{x}$ is convex (up) We have seen that the level sets of $f(\underset{\sim}{x})$ are ellipsoids.

Some properties of PD matrices

- If $A \in C^{n \times n}$ is positive semidefinite, then so are all the powers A^k, $k = 1, 2, \ldots$

Proof: if $A > 0$

$$A = UDU^* \qquad D = \begin{bmatrix} \lambda_1 & & \\ & & \\ & & \lambda_n \end{bmatrix} \qquad \lambda_i \geq 0$$

$$A^2 = AA = UDU^*UDU^*$$

$$= U\underset{\text{nonnegative diagonal}}{\underset{\sim}{D^2}}U^* \quad \geq 0$$

Similarly $A^k = UD^kU^*$

- If $A = [a_{ij}] \in C^{n \times n}$ is Hermitian and strictly diagonally dominant, i.e.,

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|, \text{ and}$$

$$\begin{bmatrix} a_{11} & a_{12} & & a_{1n} \\ a_{21} & a_{22} & & \\ \vdots & & & \\ a_{n1} & & & a_{nn} \end{bmatrix}$$

if $\underline{a_{ii} > 0}$ for all $i = 1, \ldots, n$ then A is PD.

$$\begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} \qquad \text{if } \underset{\sim}{x} = \underset{\sim}{e_i} \quad \underset{\sim}{e_i}^* A \underset{\sim}{e_i} > 0$$

$$a_{ii} > 0$$

- A is PD $\Rightarrow$ $a_{ii} > 0$

Any principal submatrix of A should be PD. (Prove this) Similarly

removing the same columns and rows in the Hermitian case

Congruence and Positive Definite Matrices

Definition: Let $A, B \in C^{n\times n}$ be given. If there exists a nonsingular matrix S such that

a) $B = SAS^*$ then B is said to be $*$-congruent to A.

b) $B = SAS^T$ then B is said to be T-congruent to A.

Similarity Transformation

$$B = SAS^{-1}$$

characteristic polynomial of B

$$\det(\lambda I - B) = \det(\lambda I - SAS^{-1}) = \det(S(\lambda S^{-1}S - A)S^{-1})$$
$$= \det(S(\lambda S^{-1} - AS^{-1})) = \det(S(\lambda I - A)S^{-1})$$
$$= \det S \, \det(\lambda I - A) \det S^{-1}$$
$$= \det(\lambda I - A)$$

$\Rightarrow$ A and B has same eigenvalues.

Congruence is an equivalence relation.

i.e. if A is $\overset{T}{*}$- congruent to B

B is $\overset{T}{*}$- congruent to C

$\Rightarrow$ A is $\overset{T}{*}$- congruent to C

So, congruent matrices form a set.

Definition: (Inertia) Let $A \in C^{n\times n}$ be Hermitian. The inertia of A is the ordered triple

$$i(A) = (i_+(A), i_-(A), i_0(A))$$

where

$i_+(A)$ is the number of positive eigenvalues of A

$i_-(A)$ negative

$i_0(A)$ zero —

Example: $A = \begin{bmatrix} 5 & & & 0 \\ & 2 & & \\ & & -1 & \\ 0 & & & 0 \end{bmatrix}$ $\quad i(A) = (2, 1, 1)$

Theorem: Let $A, B \in \mathbb{C}^{n\times n}$ be Hermitian matrices. There is a nonsingular matrix $S \in \mathbb{C}^{n\times n}$ such that $A = SBS^*$ if and only if A and B have the same inertia

A, B are $*$-congruent $\Longleftrightarrow i(A) = i(B)$

Let's assume $i(A) = i(B)$

Proof: Since $A = A^* \Rightarrow A = U_A D_A U_A^*$ $U_A U_A^* = I$ $D_A = \begin{bmatrix} \lambda_{A,1} & \\ & \lambda_{A,n} \end{bmatrix}$

$\lambda_{A,k} > 0 \quad k = 1, \dots, i_+(A)$

$\lambda_{A,k} < 0 \quad k = i_+(A)+1, \dots, i_+(A) + i_-(A)$

$\lambda_{A,k} = 0 \quad k > i_+(A) + i_-(A)$

Similarly $B = U_B D_B U_B^*$ $D_B = \begin{bmatrix} \lambda_{B,1} & & \\ & \ddots & \\ & & \lambda_{B,n} \end{bmatrix}$

$A = U_A D_A U_A^* \Leftrightarrow D_A = U_A^* A U_A$

define $Y_A = \begin{bmatrix} \frac{1}{\sqrt{|\lambda_{A,1}|}} & & & & \\ & \ddots & & & \\ & & \frac{1}{\sqrt{|\lambda_{A,i_+(A)+i_-(A)}|}} & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix}$

For example if $D_A = \begin{bmatrix} 4 & & & & \\ & 16 & & & \\ & & -2 & & \\ & & & 0 & \\ & & & & 0 \end{bmatrix}$

then $Y_A = \begin{bmatrix} \frac{1}{2} & & & & \\ & \frac{1}{4} & & & \\ & & \frac{1}{\sqrt{2}} & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix}$

$Y_A D_A Y_A^* = \begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & -1 & & & \\ & & & -1 & & \\ & & & & 0 & \\ & & & & & 0 \end{bmatrix} = J = Y_B D_B Y_B^*$ where Y_B is defined similarly

$Y_B D_B Y_B^* = Y_A D_A Y_A^*$

$D_A = Y_A^{-1} Y_B D_B Y_B^* Y_A^{-*}$

Define $S = U_B Y_B^{-1} Y_A U_A^*$

$$SAS^* = U_B Y_B Y_A^{-1} \underbrace{U_A^* U_A}_{} D_A \underbrace{U_A^* U_A}_{} Y_A^{-*} Y_B^* U_B^*$$

$$= U_B \underbrace{Y_B \underbrace{Y_A^{-1} D_A Y_A^{-*}}_{J} Y_B^*}_{D_B} U_B^*$$

$$= U_B D_B U_B^* = B$$

$$A = [\underset{\sim}{v_1}\ \underset{\sim}{v_2}\ \underset{\sim}{v_3}]\begin{bmatrix}1 & & \\ & -1 & \\ & & 1\end{bmatrix}\begin{bmatrix}\underset{\sim}{v_1}^* \\ \underset{\sim}{v_2}^* \\ \underset{\sim}{v_3}^*\end{bmatrix}$$ ⟹ You can always change the order of eigenvalues.

$$= [\underset{\sim}{v_1}\ \underset{\sim}{v_3}\ \underset{\sim}{v_2}]\begin{bmatrix}1 & & \\ & 1 & \\ & & -1\end{bmatrix}\begin{bmatrix}\underset{\sim}{v_1}^* \\ \underset{\sim}{v_3}^* \\ \underset{\sim}{v_2}^*\end{bmatrix}$$

Property: A Hermitian matrix A is positive definite iff it is $*$-congruent to I.

$\Rightarrow \exists$ invertible S

$$\boxed{A = SIS^* = \underset{\nwarrow \text{a square root of } A}{S}S^*}$$

If S is a square root of $A > 0$, i.e. SS^*, S is invertible then for any U, $UU^* = I$

$Y = SU$ is also a square root of A.

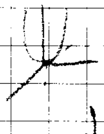
$$YY^* = SU(SU)^* = SUU^*S^* = SS^* = A.$$

Theorem (Cholesky Decomposition): A matrix $A \in \mathbb{C}^{n\times n}$ is positive definite iff there exists a nonsingular lower triangular matrix $L \in \mathbb{C}^{n\times n}$ with positive diagonal entries such that $A = LL^*$. If A is real then L may be taken as real.

P.D.

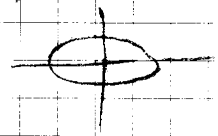
$$A = A^*, A > 0 = \begin{cases} \underset{\sim}{x}^* A \underset{\sim}{x} > 0 & \forall \underset{\sim}{x} \neq \underset{\sim}{0} \\ \lambda_i(A) > 0 \end{cases}$$

$f(\underset{\sim}{x}) = \underset{\sim}{x}^* A \underset{\sim}{x}$



convex function

$\underset{\sim}{x}^* A \underset{\sim}{x} = c$ defines ellipse



For Hermitian matrices A, B

if $A = SBS^*$, where S invertible

$\Rightarrow i(A) = i(B)$

For any positive definite matrix $A \in \mathbb{C}^{n\times n}$

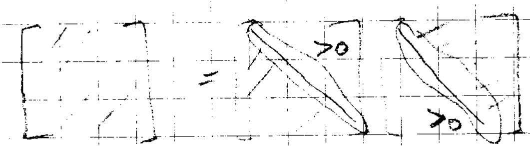
$i(A) = i(B)$ $\quad \exists S \leftarrow$ invertible

$A = SS^* \rightarrow S$ is a square root of A

$Y = SU$ for any $UU^* = I$ is also a square root of A.

Special Case: (Cholesky)

$A = LL^*$ $\quad L \Rightarrow$ Lower triangular with positive diagonal entries



Applications of Cholesky Factorization

1. To solve $H\underset{\sim}{x} = \underset{\sim}{c}$ where $H > 0$

$H = LL^*$

$\underbrace{LL^*\underset{\sim}{x}}_{\underset{\sim}{y}} = \underset{\sim}{c}$

$\left.\begin{array}{l} L\underset{\sim}{y} = \underset{\sim}{c} \\ L^*\underset{\sim}{x} = \underset{\sim}{y} \end{array}\right\}$

Due to Condition number

(Numerically, not a good idea!)

For a general overdetermined $m \geq n$

$A\underset{\sim}{x} = \underset{\sim}{b}$

$\boxed{A^*A\underset{\sim}{x} = A^*\underset{\sim}{b}}$ Normal equations

We'll see this in Least Squares discussion.

A^*A is Hermitian if A is full rank

$\Rightarrow A^*A$ is invertible (tall & full rank)

Lemma: Given any $A \in \mathbb{C}^{m \times n}$, $A^*A \geq 0$

We should show for any $\underset{\sim}{x}$ $\quad \underset{\sim}{x}^*A^*A\underset{\sim}{x} \geq 0$

$(A\underset{\sim}{x})^*A\underset{\sim}{x}$

$\|A\underset{\sim}{x}\|^2 \geq 0$

so $A^*A \geq 0$

When is $A^*A > 0$

$\underset{\sim}{x} \neq \underset{\sim}{0} \quad \underset{\sim}{x}^*A^*A\underset{\sim}{x} > 0$

$\Rightarrow \|A\underset{\sim}{x}\|^2 > 0$

$\Rightarrow A\underset{\sim}{x} \neq \underset{\sim}{0} \quad \Rightarrow \underbrace{N(A) = \{\underset{\sim}{0}\}}$

$m \geq n$, A is full rank

2. Computer Simulations of Random Vectors

Goal: Generate random vectors with correlation matrix $R > 0$ in MATLAB.

$$R = E(\underset{\sim}{x}\underset{\sim}{x}^*) = E\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\begin{bmatrix} x_1^* & \cdots & x_n^* \end{bmatrix}\right) = E\left(\begin{bmatrix} x_1x_1^* & x_1x_2^* & \cdots & x_1x_n^* \\ x_2x_1^* & x_2x_2^* & & \vdots \\ \vdots & & \ddots & \vdots \\ x_nx_1^* & \cdots & & x_nx_n^* \end{bmatrix}\right)$$

$$= \begin{bmatrix} E(x_1 x_1^*) & E(x_1 x_2^*) & \cdots & E(x_1 x_n^*) \\ E(x_2 x_1^*) & E(x_2 x_2^*) & & \\ & & & \\ E(x_n x_1^*) & \cdots & & E(x_n x_n^*) \end{bmatrix}$$

$R_{ij} \rightarrow$ correlation of x_i with x_j

If $\underset{\sim}{x}$ is $\underline{\text{uncorrelated}}$ then R is diagonal.

elements of $\underset{\sim}{x}$ are mutually uncorrelated.

$R \rightarrow$ Hermitian $(R = R^*)$

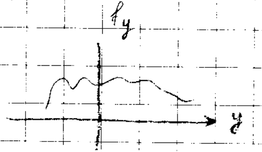
$R \rightarrow R \geq 0 \rightarrow$ P.S.D

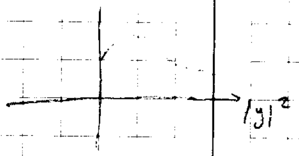
$$R \geq 0 \iff \forall \underset{\sim}{a} \in \mathbb{C}^n \quad \underset{\sim}{a}^* R \underset{\sim}{a} \geq 0$$

$$\underset{\sim}{a}^* R \underset{\sim}{a} = \underset{\sim}{a}^* E(\underset{\sim}{x}\underset{\sim}{x}^*) \underset{\sim}{a}$$

$$= E(\underbrace{\underset{\sim}{a}^* \underset{\sim}{x}}_{y} \underbrace{\underset{\sim}{x}^* \underset{\sim}{a}}_{y^*}) = E(yy^*)$$

$$= \underbrace{E(|y|^2)}_{\geq 0}$$





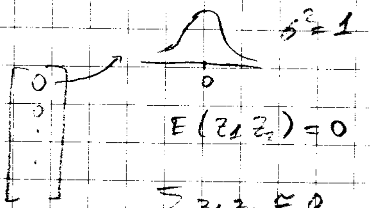
so $R \geq 0$.

If R is invertible $R > 0$

In MATLAB if you use

`z = randn (n, 1);` ← Generates an uncorrelated vector

$E(\underset{\sim}{z}\underset{\sim}{z}^*) = I$



$E(z_1 z_2) = 0 \rightarrow$ independent

$\sum z_1 z_2 \cong 0$

`L = chol(R);` $(R = LL^*)$

`x = L * z;` $E(\underset{\sim}{x}\underset{\sim}{x}^*) = E(L \underset{\sim}{z}\underset{\sim}{z}^* L^*)$

I make the elements of $\underset{\sim}{z}$ correlated.

$$= L\, E(\underset{\sim}{z}\underset{\sim}{z}^*)\, L^*$$

$$= LL^* = R$$

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} l_{11} & & & \\ l_{21} & l_{22} & & \\ & & & \\ l_{n1} & \cdots & & l_{nn} \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$$

← Coloring matrix

$x_1 = l_{11} z_1$

$x_2 = l_{21} z_1 + l_{22} z_2$

$\underset{\sim}{z} \rightarrow$ "white" = uncorrelated

$\underset{\sim}{x} \rightarrow$ "colored" = correlated rank($\underset{\sim}{x}\,\underset{\sim}{y}^*$) = 1 but rank($E(\underset{\sim}{x}\,\underset{\sim}{y}^*)$) may be full.

$$E\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\begin{bmatrix} x_1^* & \cdots & x_n^* \end{bmatrix}\right) = \sum_{x_1 \ldots x_n} \underset{\sim}{x}\,\underset{\sim}{x}^* p_{\underset{\sim}{x}}(\underset{\sim}{x})$$

→ We can also use this to prove that R is always PSD.

i.e.
$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\text{rank}=1} + \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\text{rank}=1} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\text{rank}=2}$$

Algorithm for Computing Cholesky Factorization

Let $B > 0$, $B \in \mathbb{C}^{n \times n}$, $B = B^*$

$$B = \begin{bmatrix} \alpha & \underset{\sim}{v}^* \\ \underset{\sim}{v} & B_{n-1} \end{bmatrix} \qquad \alpha \in \mathbb{R},\ \alpha > 0$$

$$\begin{array}{c} \\ n-1 \end{array}\begin{bmatrix} 1 & \overset{1 \times n-1}{\underset{\sim}{0}} \\ -\frac{\underset{\sim}{v}}{\alpha} & I \end{bmatrix}\begin{bmatrix} \alpha & \underset{\sim}{v}^* \\ \underset{\sim}{v} & B_{n-1} \end{bmatrix} = \begin{bmatrix} \alpha & \underset{\sim}{v}^* \\ \underset{\sim}{0} & B_{n-1} - \frac{\underset{\sim}{v}\underset{\sim}{v}^*}{\alpha} \end{bmatrix}$$

$$\begin{bmatrix} 1 & \underset{\sim}{0} \\ -\frac{\underset{\sim}{v}}{\alpha} & I \end{bmatrix}\begin{bmatrix} \alpha & \underset{\sim}{v}^* \\ \underset{\sim}{v} & B_{n-1} \end{bmatrix}\begin{bmatrix} 1 & -\frac{\underset{\sim}{v}^*}{\alpha} \\ \underset{\sim}{0} & I \end{bmatrix} = \begin{bmatrix} \alpha & \underset{\sim}{v}^* \\ \underset{\sim}{0} & B_{n-1} - \frac{\underset{\sim}{v}\underset{\sim}{v}^*}{\alpha} \end{bmatrix}\begin{bmatrix} 1 & -\frac{\underset{\sim}{v}^*}{\alpha} \\ 0 & I \end{bmatrix}$$

$$= \begin{bmatrix} \alpha & \underset{\sim}{0} \\ \underset{\sim}{0} & B_{n-1} - \frac{\underset{\sim}{v}\underset{\sim}{v}^*}{\alpha} \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 \\ \frac{\underset{\sim}{v}}{\alpha} & I \end{bmatrix}\begin{bmatrix} 1 & 0 \\ -\frac{\underset{\sim}{v}}{\alpha} & I \end{bmatrix} = I$$

$$\underbrace{\begin{bmatrix} \alpha & \underset{\sim}{v}^* \\ \underset{\sim}{v} & B_{n-1} \end{bmatrix}}_{B} = \begin{bmatrix} 1 & 0 \\ \frac{\underset{\sim}{v}}{\alpha} & I \end{bmatrix}\begin{bmatrix} \overset{\sqrt{\alpha}\sqrt{\alpha}}{\alpha} & \underset{\sim}{0} \\ \underset{\sim}{0} & B_{n-1} - \frac{\underset{\sim}{v}\underset{\sim}{v}^*}{\alpha} \end{bmatrix}\begin{bmatrix} 1 & \frac{\underset{\sim}{v}^*}{\alpha} \\ 0 & I \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} \sqrt{\alpha} & 0 \\ \frac{\underset{\sim}{v}}{\alpha} & I \end{bmatrix}}_{S \text{ invertible}}\underbrace{\begin{bmatrix} 1 & 0 \\ 0 & \underbrace{B_{n-1} - \frac{\underset{\sim}{v}\underset{\sim}{v}^*}{\alpha}}_{\substack{\text{Hermitian} \\ >0}} \end{bmatrix}}_{B'}\underbrace{\begin{bmatrix} \sqrt{\alpha} & \frac{\underset{\sim}{v}^*}{\alpha} \\ 0 & I \end{bmatrix}}_{S^*}$$

$$B_{n-1} - \frac{\underset{\sim}{v}\underset{\sim}{v}^*}{\alpha} \rightarrow L_{n-1}L_{n-1}^*$$

B and B′ are *-congruent.

$$i(B) = i(B') \iff B > 0 \Rightarrow B' > 0$$

$$= \begin{bmatrix} \sqrt{\alpha} & 0 \\ \frac{\underset{\sim}{v}}{\sqrt{\alpha}} & L_{n-1} \end{bmatrix}\begin{bmatrix} \sqrt{\alpha} & \frac{\underset{\sim}{v}^*}{\sqrt{\alpha}} \\ 0 & L_{n-1}^* \end{bmatrix}$$

QR Factorization (Construction of Orthogonal Basis)

$A = QR \rightarrow$ Upper triangular matrix

$[a_1 \cdots a_n]$ $\downarrow$ Matrix with orthonormal columns

Given

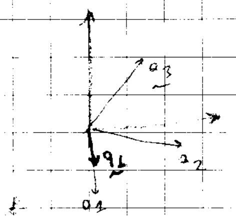
$\text{Span}\{a_1, \dots, a_n\} = \text{Span}\{q_1, \dots, q_{n'}\}$ $\quad \langle q_i, q_j \rangle = \delta_{ij}$ (Elements have unit norm)

$n' \leq n$ $\quad$ Orthonormal basis for the span of $\{a_1, \dots, a_n\}$

$n' = n$ if the set is linearly independent.

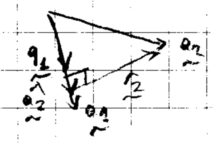
Let's look at construction of orthogonal basis problem:

One particular solution: Gram-Schmidt procedure



Let's assume $\{a_1, \dots, a_n\}$ is linearly independent.

1. Choose $q_1 = \dfrac{a_1}{\|a_1\|}$



2. Define $r_2 = a_2 - \hat{a}_2 \leftarrow$ projection of a_2 over q_1

$$\hat{a}_2 = \frac{\langle a_2, q_1 \rangle}{\underbrace{\langle q_1, q_1 \rangle}_{1}} q_1 \rightarrow \text{Alternatively}$$

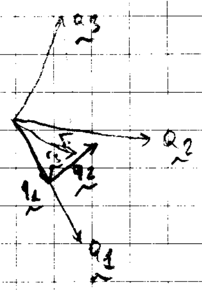
$$\hat{a}_2 = \frac{\langle a_2, a_1 \rangle}{\langle a_1, a_1 \rangle} a_1$$

$$r_2 = a_2 - \langle a_2, q_1 \rangle q_1$$

$\text{Span}\{q_1\} = \text{Span}\{a_1\}$

$$q_2 = \frac{r_2}{\|r_2\|} \qquad \text{Span}\{q_1, q_2\} = \text{Span}\{a_1, a_2\}$$

$\text{Span}\{q_1, q_2, q_3\} = \text{Span}\{a_1, a_2, a_3\}$



3. $r_3 = a_3 - \hat{a}_3 \leftarrow$ Projection of a_3 on $\text{Span}\{a_1, a_2\}$

$\text{Span}\{q_1, q_2\}$

$$\hat{a}_3 = \langle a_3, q_1 \rangle q_1 + \langle a_3, q_2 \rangle q_2$$

$$\neq \frac{\langle a_3, a_1 \rangle}{\langle a_1, a_1 \rangle} a_1 + \frac{\langle a_3, a_2 \rangle}{\langle a_2, a_2 \rangle} a_2$$

$$q_3 = \frac{r_3}{\|r_3\|}$$

K^{th} step: $r_k = a_k - \underbrace{\sum_{n=1}^{k'-1} \langle a_k, q_n \rangle q_n}_{\hat{a}_k}$

$$q_{k'} = \frac{r_k}{\|r_k\|}$$

$k' = k' + 1.$

If $a_k \in \text{Span}\{a_1, \dots, a_{k-1}\}$

$\Rightarrow r_k = 0 \quad (\hat{a}_k = a_k)$

$\Rightarrow$ Skip defining a new q vector.

$$\underset{\sim}{q_1} = \frac{\underset{\sim}{a_1}}{\|\underset{\sim}{a_1}\| = r_{11}}$$

$$\underset{\sim}{a_1} = \underset{\sim}{q_1} \cdot r_{11} \qquad r_{11} = \|\underset{\sim}{a_1}\|$$

$$\underset{\sim}{a_2} = \underbrace{\underset{\sim}{q_2}\, \|\underset{\sim}{r_2}\|}_{\underset{\sim}{r_2} \quad r_{22}} + \underbrace{\langle \underset{\sim}{a_2}, \underset{\sim}{q_1} \rangle}_{r_{12}} \underset{\sim}{q_1}$$

$$\underset{\sim}{a_3} = \underbrace{\underset{\sim}{q_3}\, \|\underset{\sim}{r_3}\|}_{\underset{\sim}{r_3}} + \underbrace{\langle \underset{\sim}{a_3}, \underset{\sim}{q_1} \rangle}_{r_{13}} \underset{\sim}{q_1} + \underbrace{\langle \underset{\sim}{a_3}, \underset{\sim}{q_2} \rangle}_{r_{23}} \underset{\sim}{q_2}$$

$$3 \updownarrow \overset{\xleftrightarrow{3}}{\begin{bmatrix} \underset{\sim}{a_1} & \underset{\sim}{a_2} & \underset{\sim}{a_3} \end{bmatrix}} = 3 \updownarrow \overset{\xrightarrow{3}}{\begin{bmatrix} \underset{\sim}{q_1} & \underset{\sim}{q_2} & \underset{\sim}{q_3} \end{bmatrix}} \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{bmatrix}$$

→ Since linear combination of columns, multiply from right.

Suppose

$\underset{\sim}{a_3} \in \text{Span}\{\underset{\sim}{a_1}, \underset{\sim}{a_2}\}$

$$\begin{bmatrix} \underset{\sim}{a_1} & \underset{\sim}{a_2} & \underset{\sim}{a_3} \end{bmatrix} = \underset{\hat{Q}}{\begin{bmatrix} \underset{\sim}{q_1} & \underset{\sim}{q_2} \end{bmatrix}} \underset{\hat{R}}{\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \end{bmatrix}}$$

← Reduced QR ⟺ If A is full rank

$\text{size}(A) = \text{size}(Q)$

o.w. $\text{size}(A) > \text{size}(Q)$

//

Suppose $\underset{\sim}{a_k} \in C^{3\times 1}$

$$\begin{bmatrix} \underset{\sim}{a_1} & \underset{\sim}{a_2} & \underset{\sim}{a_3} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{q_1} & \underset{\sim}{q_2} & \underset{\text{orthogonal to } \underset{\sim}{q_1}, \underset{\sim}{q_2}}{\underset{\sim}{q_3}'} \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & 0 \end{bmatrix}$$

← Full QR

→ added column is silenced.

$$= \begin{bmatrix} \hat{Q} & \underset{\sim}{q_3}' \end{bmatrix} \begin{bmatrix} \hat{R} \\ 0 \end{bmatrix} = \hat{Q}\hat{R} + \underset{0}{\cancel{\hat{q}_3' \underset{\sim}{0}}}$$

Suppose $\underset{\sim}{a_1}, \underset{\sim}{a_2}, \underset{\sim}{a_3} \in C^{10,1}$ and they are linearly independent

$$\underset{10\times 3}{\begin{bmatrix} \underset{\sim}{a_1} & \underset{\sim}{a_2} & \underset{\sim}{a_3} \end{bmatrix}} = \overset{10\times 3}{\underset{\hat{Q}}{\begin{bmatrix} \underset{\sim}{q_1} & \underset{\sim}{q_2} & \underset{\sim}{q_3} \end{bmatrix}}} \overset{3\times 3}{\underset{\hat{R}}{\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{bmatrix}}}$$

← Reduced QR

$$\underset{10\times 3}{\begin{bmatrix} \underset{\sim}{a_1} & \underset{\sim}{a_2} & \underset{\sim}{a_3} \end{bmatrix}} = \underset{10\times 10}{\begin{bmatrix} \underset{\sim}{q_1} & \underset{\sim}{q_2} & \underset{\sim}{q_3} & \underset{\sim}{q_4}' & \cdots & \underset{\sim}{q_{10}}' \end{bmatrix}} \begin{array}{r} \\ 7 \updownarrow \end{array} \begin{bmatrix} \hat{R} \\ \hline 0 \end{bmatrix} \begin{array}{l} 3 \\ \\ \end{array}$$

← Full QR $\quad \text{size}(A) = \text{size}(R)$

$R \to 10 \times 3$

If A is full rank square matrix

⇒ Full QR = Reduced QR

square matrices

$\hat{A}\hat{R}_1\hat{R}_2\cdots\hat{R}_n=\hat{Q}$

Obtaining QR Factorization (Focus of MATH504)

Algorithms:

Triangular orthogonalization
1. Gram-Schmidt
2. Modified Gram-Schmidt

$[\underset{\sim}{a_1}\ \underset{\sim}{a_2} \cdots \underset{\sim}{a_n}]$ $[\underset{\sim}{q_1}$

$[\underset{\sim}{a_1}\ \underset{\sim}{v_2} \cdots \underset{\sim}{v_n}]$

↳ projections on q_i are subtracted.

Orthogonal Triangularization
3. Householder Transformations ← Reflections
4. Givens Rotations ← Rotations

[illegible]

$Q_n\cdots Q_2 Q_1 A = R$

Useful Matrix Identities (Block Matrices)

Consider a block matrix

$$\mathcal{M}=\begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

$$\begin{bmatrix} I & 0 \\ X & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix}=\begin{bmatrix} A & B \\ \underbrace{XA+C}_{0} & XB+D \end{bmatrix}$$

$$X=-CA^{-1}$$

$$\det\left(\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix}\right)=\det\left(\begin{bmatrix} A & B \\ 0 & \underbrace{D-CA^{-1}B}_{\Delta_A} \end{bmatrix}\right)$$

Δ_A → Schur complement of A.

$$\det \mathcal{M} = \det A \cdot \det \Delta_A$$

Similarly

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix}=\begin{bmatrix} A & 0 \\ C & \underbrace{D-CA^{-1}B}_{\Delta_A} \end{bmatrix}$$

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix}=\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & 0 \\ C & \underbrace{D-CA^{-1}B}_{\Delta_A} \end{bmatrix}$$

$$=\begin{bmatrix} A & 0 \\ 0 & \Delta_A \end{bmatrix}$$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}=\begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & 0 \\ 0 & \Delta_A \end{bmatrix}\begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}$$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1}=\begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix}\begin{bmatrix} A^{-1} & 0 \\ 0 & \Delta_A^{-1} \end{bmatrix}\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}$$

HW#3, Q2

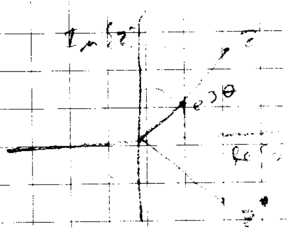
A nonsingular, $A = \underset{>0}{R}\,U$ $\quad UU^* = I$ $\qquad z = \overset{>0}{r}\, e^{j\theta}$

$AA^* = RU(RU)^*$

$= RR^*$

$zz^* = re^{j\theta}\, re^{-j\theta}$

$= r^2$



closest point in the unit circle to z.

AA^* nonsingular $\Rightarrow$ PD $\Rightarrow QDQ^*$

$\quad \hookrightarrow QQ^* = I$

$R = QD^{1/2}Q^*$

$AA^* = RR^*$

$\Rightarrow$ Both are the square roots of the same thing

$\Rightarrow A = RU$

$\hookrightarrow$ U is the closest unitary matrix among all matrices

Useful Matrix Identities

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & \Delta_A \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}$$

$\to$ Schur complement of

$$\det M = \det A \cdot \det \Delta_A = \det A \det(D - CA^{-1}B)$$

$$M^{-1} = \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & \Delta_A^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}$$

$$M = \begin{bmatrix} I & X \\ 0 & I \end{bmatrix} \begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix} \begin{bmatrix} I & 0 \\ X & I \end{bmatrix}$$

$\qquad\qquad\qquad D \qquad\qquad L$

$$= \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \overset{\Delta_D}{A - BD^{-1}C} & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}$$

Schur complement of D

$$\det M = \det D \cdot \det \Delta_D$$

$$M^{-1} = \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} \begin{bmatrix} \Delta_D^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix}$$

Memorize

$$\boxed{(A+BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}}$$ Matrix Inversion Lemma

Ex: $(R + \underset{\sim}{x}\underset{\sim}{y}^*)^{-1} = R^{-1} - R^{-1}\underset{\sim}{x}(1 + \underbrace{\underset{\sim}{y}^* R^{-1}\underset{\sim}{x}}_{\text{scalar}})^{-1}\underset{\sim}{y}^* R^{-1}$

$$= R^{-1}\left(I - \frac{\underset{\sim}{x}\underset{\sim}{y}^* R^{-1}}{(1 + \underset{\sim}{y}^* R^{-1}\underset{\sim}{x})}\right)$$

$$R_n = \frac{1}{n}\sum_{k=1}^{n} \underset{\sim}{x_k}\, \underset{\sim}{y_k}^*$$

↙ Update of your covariance matrix with new data

$$R_n = R_{n-1} + \underset{\sim}{x}_n \underset{\sim}{y}_n^*$$

You may be using the inverse of R_n in your application.

Simple minded approach: $R_n = R_{n-1} + \underset{\sim}{x}_n \underset{\sim}{y}_n^*$

R_n^{-1}

Instead:

$$R_n^{-1} = R_{n-1}^{-1}\left(I - \frac{\underset{\sim}{x}_n \underset{\sim}{y}_n^* R_{n-1}^{-1}}{(1 + \underset{\sim}{y}_n^* R_{n-1}^{-1} \underset{\sim}{x})}\right)$$

Linear (Dynamical) Systems

- Linear Algebra & Matrix Theory
- Linear Systems ←
- Linear Estimation

When you press the ~~brake~~ gas pedal in a car it does not change its state immediately since previous states effect the current state dynamically.

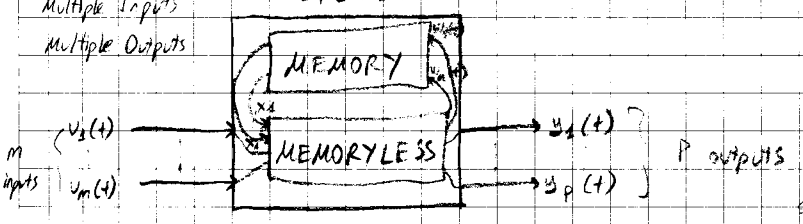
- State Space Representation: Excellent way to model a dynamical system with memory.

Logical Structure of State Space Representation

SSR + Kalman Filtering (Linear Estimation) enabled us to land on moon.

(MIMO)
Multiple Inputs
Multiple Outputs

SYSTEM



↳ Antennas → Signal is WSS r.p. → sample on different times?

$x(t) \to$ input
$y(t) \to$ output
System (memoryless)
$y(t) = \alpha\, x(t)$

System with memory

$$y(t) = \int_{-\infty}^{\infty} x(\tau)\, d\tau$$

$$y_1(t) = g_1(x_1(t), \ldots, x_n(t), u_1(t), \ldots, u_m(t), t)$$

↳ order of my system = # of memory elements used

$$y_p(t) = g_p(x_1(t), \ldots, x_n(t), u_1(t), \ldots, u_m(t), t)$$

g_i's → memoryless

$$\underset{\sim}{y}(t) = \begin{bmatrix} g_1(t) \\ \vdots \\ y_p(t) \end{bmatrix} \quad \underset{\sim}{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_p(t) \end{bmatrix} \quad \underset{\sim}{u}(t) = \begin{bmatrix} u_1(t) \\ \vdots \\ u_m(t) \end{bmatrix}$$

$x_1(t-1)$ … previous states can also be added but ~~that~~ would be a different system (because the philosophy of this system)

$$\underset{\sim}{y}(t) = g(\underset{\sim}{x}(t), u(t), t)$$

$$w_1(t) = f_1(x_1(t), x_2(t), \ldots, x_n(t), u_1(t), \ldots, u_m(t), t)$$

$$w_n(t) = f_n(x_1(t), \ldots, x_n(t), u_1(t), \ldots, u_m(t), t)$$

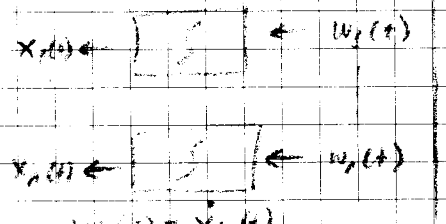
$$\underset{\sim}{w}(t) = \begin{bmatrix} w_1(t) \\ \vdots \\ w_n(t) \end{bmatrix}$$

$$\left.\begin{array}{l} \rightarrow \underset{\sim}{w}(t) = f(\underset{\sim}{x}(t), \underset{\sim}{u}(t), t) \\ \rightarrow \underset{\sim}{y}(t) = g(\underset{\sim}{x}(t), \underset{\sim}{u}(t), t) \end{array}\right]$$

How do we use $\underset{\sim}{w}(t)$ to update $\underset{\sim}{x}(t)$?

Basic idea: keep memory part as simple as possible

C.T.

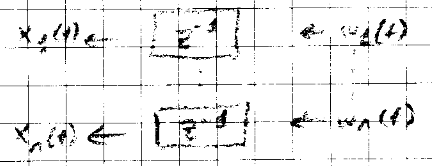


$w_k(t) = \dot{x}_k(t)$

derivative.

$\underset{\sim}{w}(t) = \dot{\underset{\sim}{x}}(t)$

D.T.



$x_k(t+1) = w_k(t)$

C.T.S.S.

$\dot{\underset{\sim}{x}}(t) = f(\underset{\sim}{x}(t), \underset{\sim}{u}(t), t)$

$\underset{\sim}{y}(t) = g(\underset{\sim}{x}(t), \underset{\sim}{u}(t), t)$

D.T.S.S.

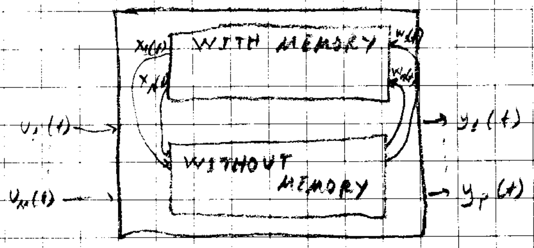
$\underset{\sim}{x}(k+1) = f(\underset{\sim}{x}(k), \underset{\sim}{u}(k), k)$

$\underset{\sim}{y}(k) = g(\underset{\sim}{x}(k), \underset{\sim}{u}(k), k)$

f, g are general nonlinear functions. f, g are memoryless functions. They do not determine the state. They just determine the change in the state.

Nonlinear – time varying

Linear Dynamical Systems



C.T.

$\dot{\underset{\sim}{x}}(t) = f(\underset{\sim}{x}(t), \underset{\sim}{u}(t), t)$

$\underset{\sim}{y}(t) = g(\underset{\sim}{x}(t), \underset{\sim}{u}(t), t)$

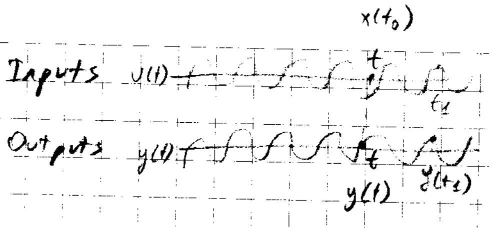
D.T.

$\underset{\sim}{x}_{k+1} = f(\underset{\sim}{x}_k, \underset{\sim}{u}_k, k)$

$\underset{\sim}{y}_k = g(\underset{\sim}{x}_k, \underset{\sim}{u}_k, k)$

Concept of State

State is a set of quantities for which the present state is sufficient in predicting the current and future output given the present and future outputs.



To calculate $y(t_1)$ you need

$\{u(t),\ t \le t_1\}$ or

$x(t_0)$ and $\{u(t);\ t_0 \le t \le t_1\}$

Example:

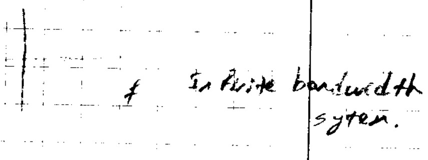
Perfect Delay Line

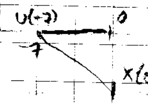
$y(t) = u(t-7)$ (Continuous Time System)

$H(t) = e^{-j2\pi f7}$ $h(t) = \delta(t-7)$

How many states does this system have?

a) 0 b) 1 c) 7 [✓ d) ∞] e) none





We need infinitely many states.

$y(0) = u(-7)$

$y(0.1) = u(-6.9)$

If this was a discrete time system: Answer = 7

LINEARITY AND TIME INVARIANCE

a) Linear Time Varying

$$\underset{\sim}{x}(t) \in \mathbb{R}^n \qquad \dot{\underset{\sim}{x}}(t) = \overset{n\times n}{A}(t)\,\underset{\sim}{x}(t) + \overset{n\times m}{B}(t)\,\underset{\sim}{u}(t) \qquad \underset{\sim}{u}(t) \in \mathbb{R}^m$$

$$\underset{\sim}{y}(t) = \overset{p\times n}{C}(t)\,\underset{\sim}{x}(t) + \overset{p\times m}{D}(t)\,\underset{\sim}{u}(t) \qquad \underset{\sim}{y}(t) \in \mathbb{R}^p$$

D.T.

$$\underset{\sim}{x}_{k+1} = A(k)\,\underset{\sim}{x}_k + B(k)\,\underset{\sim}{u}_k$$

$$\underset{\sim}{y}_k = C(k)\,\underset{\sim}{x}_k + D(k)\,\underset{\sim}{u}_k$$

b) Linear Time Invariant (LTI)

C.T.

$$\dot{\underset{\sim}{x}}(t) = A\underset{\sim}{x}(t) + B\underset{\sim}{u}(t)$$

$$\underset{\sim}{y}(t) = C\underset{\sim}{x}(t) + D\underset{\sim}{u}(t)$$

D.T.

$$\underset{\sim}{x}_{k+1} = A\underset{\sim}{x}_k + B\underset{\sim}{u}_k$$

$$\underset{\sim}{y}_k = C\underset{\sim}{x}_k + D\underset{\sim}{x}_k$$

$$\begin{bmatrix}\dot{\underset{\sim}{x}}(t)\end{bmatrix} = {}_{n}\overset{n}{\begin{bmatrix} \quad A \quad \end{bmatrix}}\begin{bmatrix}\underset{\sim}{x}(t)\end{bmatrix} + {}_{n}\overset{m}{\begin{bmatrix} \quad B \quad \end{bmatrix}}\begin{bmatrix}\underset{\sim}{u}(t)\end{bmatrix}$$

$$\begin{bmatrix}\underset{\sim}{y}(t)\end{bmatrix} = {}_{p}\overset{n}{\begin{bmatrix} \quad C \quad \end{bmatrix}}\begin{bmatrix}\underset{\sim}{x}(t)\end{bmatrix} + {}_{p}\overset{m}{\begin{bmatrix} \quad D \quad \end{bmatrix}}\begin{bmatrix}\underset{\sim}{u}(t)\end{bmatrix}$$

$$\begin{bmatrix}\dot{\underset{\sim}{x}}(t) \\ \underset{\sim}{y}(t)\end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix}\underset{\sim}{x}(t) \\ \underset{\sim}{u}(t)\end{bmatrix}$$

$$\begin{bmatrix}\underset{\sim}{x}_{k+1} \\ \underset{\sim}{y}_k\end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix}\underset{\sim}{x}_k \\ \underset{\sim}{u}_k\end{bmatrix}$$

SOLUTION OF STATE SPACE EQUATIONS

A. DISCRETE-TIME LTI CASE

Starting with $k = k_0$ (we assume we know $\underline{x}(k_0)$)

$$\underline{x}(k_0+1) = A\underline{x}(k_0) + B\underline{u}(k_0)$$

$$\underline{x}(k_0+2) = A(A\underline{x}(k_0) + B\underline{u}(k_0)) + B\underline{u}(k_0+1)$$

$$= A^2\underline{x}(k_0) + AB\underline{u}(k_0) + B\underline{u}(k_0+2)$$

$$\underline{x}(k_0+3) = A^3\underline{x}(k_0) + A^2B\underline{u}(k_0) + AB\underline{u}(k_0+1) + B\underline{u}(k_0+3)$$

for $\ell > k_0$ $\quad \underline{x}(\ell) = A^{\ell-k_0}\underline{x}(k_0) + \sum_{j=0}^{\ell-k_0-1} A^{j}B\underline{u}(\ell-j-1)$

$$\underline{x}(\ell) = A^{\ell-k_0}\underline{x}(k_0) + \begin{bmatrix} B & AB & A^2B & \cdots & A^{\ell-k_0-1}B \end{bmatrix}\begin{bmatrix} \underline{u}(\ell-1) \\ \underline{u}(\ell-2) \\ \vdots \\ \underline{u}(k_0) \end{bmatrix}$$

$$\underline{y}(\ell) = C\underline{x}(\ell) + D\underline{u}(\ell)$$

$$= \underset{\text{(state dependent)}}{CA^{\ell-k_0}\underline{x}(k_0)} + \begin{bmatrix} \underset{h(0)}{\overset{p\times m}{D}} & \underset{h(1)}{\overset{p\times m}{\overset{p\times n}{C}\overset{n\times m}{B}}} & \underset{h(2)}{CAB} & \cdots & \underset{h(\ell-k_0)}{CA^{\ell-k_0-1}B} \end{bmatrix}\begin{bmatrix} \underline{u}(\ell) \\ \underline{u}(\ell-1) \\ \vdots \\ \underline{u}(k_0) \end{bmatrix}$$

$$= CA^{\ell-k_0}\underline{x}(k_0) + \sum_{j=0}^{\ell-k_0} \underset{\to\, p\times m \text{ matrix}}{h(j)}\,\overset{m\times 1}{u}(\ell-j) \leftarrow \text{(input dependent)}$$

What is the impulse response?

Single Input case $\quad \underline{x}(0) = 0 \quad \underline{u}(n) = \delta(n) = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$

$$\underline{y}(0) = C\underline{x}(0) + Du(0) = D$$

$$\underline{x}(1) = C\underline{x}(0) + Bu(0) = B$$

$$\underline{y}(1) = C\underline{x}(1) + D\underline{u}(1) = CB$$

$$\underline{x}(2) = A\underline{x}(1) + B\underline{u}(1) = AB$$

$$\underline{y}(2) = C\underline{x}(2) + D\underline{u}(2) = CAB$$

$$\underline{y}(3) = \qquad CA^2B$$

B. CONTINUOUS TIME CASE

$$\dot{\underset{\wedge}{x}}(t) = A\underset{\wedge}{x}(t) + B\underset{\wedge}{u}(t) \quad ?$$

vector differential equation (first order)

Let's first look at $\underset{\wedge}{u}(t) = 0$ (Homogeneous case)

$$\dot{\underset{\wedge}{x}}(t) = A\underset{\wedge}{x}(t)$$

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

Diagonal matrix

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} a_{11} & \\ & a_{22} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

$$x_1(t) = e^{a_{11}t} x_1(0)$$

Like the chicken-and-egg problem

$$\boxed{\dot{x}_1(t) - a_{11}x_1(t) = a_{12}x_2(t)}$$

Diff. Eq. for $x_1(t)$. If we know $x_2(t)$ we could solve for $x_1(t)$

$$\boxed{\dot{x}_2(t) - a_{22}x_2(t) = a_{21}x_1(t)}$$

If $\underset{\wedge}{x}(t)$ is smooth

$$\underset{\wedge}{x}(t) = \sum_{k=0}^{\infty} \frac{x^{(k)}(0)\,t^k}{k!}$$

Taylor series expansion

$$f(t) = \sum \frac{f^{(k)}(t_0)}{k!}(t-t_0)^k$$

$$= \underset{\wedge}{x}(0) + \dot{\underset{\wedge}{x}}(0)t + \ddot{\underset{\wedge}{x}}(0)\frac{t^2}{2} + \cdots$$

$$\underset{\wedge}{x}(t) = \underset{\wedge}{x}(0) + A\underset{\wedge}{x}(0)t + A^2\underset{\wedge}{x}(0)\frac{t^2}{2} + \cdots$$

$$= \left(I + At + \frac{(At)^2}{2} + \frac{(At)^3}{3!} + \cdots\right)\underset{\wedge}{x}(0)$$

$$= e^{At}\underset{\wedge}{x}(0)$$

$$e^{At} \triangleq \left(I + At + \frac{(At)^2}{2} + \cdots\right)$$

$$\frac{de^{At}}{dt} = A + A^2t + \frac{A^3t^2}{2!} + \cdots$$

$$= A\left(I + At + \frac{(At)^2}{2!} + \cdots\right)$$

$$= Ae^{At}$$

$$x^{(3)}(t) + a_2x^{(2)}(t) + a_1x^{(1)}(t) + a_0X(t) = 0$$

$$(s^3 + a_2s^2 + a_1s + a_0)X(s) = 0$$ Laplace transform

$$x(t) = c_1e^{s_1t} + c_2e^{s_2t} + c_3e^{s_3t}$$

Oppenheim

$s \leftrightarrow$ [illegible]

Linear System Theory

C.T. | D.T.

$$\dot{\underset{\sim}{x}}(t) = A\underset{\sim}{x}(t) + B\underset{\sim}{u}(t) \qquad \underset{\sim}{x}_{k+1} = A\underset{\sim}{x}_k + B\underset{\sim}{u}_k$$

$$\underset{\sim}{y}(t) = C\underset{\sim}{x}(t) + D\underset{\sim}{u}(t) \qquad \underset{\sim}{y}_k = C\underset{\sim}{x}_k + D\underset{\sim}{u}_k$$

$$y(t) = F(\underset{0}{x(t_0)}, \underset{0}{\{u(t), t \geq t_0\}})$$

$u(t) = 0$

$$\dot{\underset{\sim}{x}}(t) = A\underset{\sim}{x}(t)$$

$$\underset{\sim}{x}(t) = e^{At}\underset{\sim}{x}(0) \rightarrow e^{A(t-t_0)}\underset{\sim}{x}(t_0)$$

$$e^{At} = (I + tA + \frac{t^2A^2}{2} + \frac{t^3A^3}{3} + \dots)$$

Example: $A = \begin{bmatrix} 0 & \alpha \\ 0 & 0 \end{bmatrix}$ $\quad A^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ $\quad A^k = 0$, $k \geq 2$

$$e^{At} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & \alpha t \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & \alpha t \\ 0 & 1 \end{bmatrix}$$

$$x(t) = \begin{bmatrix} 1 & \alpha t \\ 0 & 1 \end{bmatrix} x(0)$$

Example: Diagonal $A = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}$ $\quad A^k = \begin{bmatrix} \alpha^k & 0 \\ 0 & \beta^k \end{bmatrix}$

$$e^{At} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{I} + \underbrace{\begin{bmatrix} \alpha t & 0 \\ 0 & \beta t \end{bmatrix}}_{At} + \underbrace{\begin{bmatrix} \frac{\alpha^2t^2}{2} & 0 \\ 0 & \frac{\beta^2t^2}{2} \end{bmatrix}}_{\frac{A^2t^2}{2}} + \begin{bmatrix} \frac{\alpha^3t^3}{3!} & 0 \\ 0 & \frac{\beta^3t^3}{3!} \end{bmatrix} + \dots$$

$$= \begin{bmatrix} 1 + \alpha t + \frac{\alpha^2t^2}{2} + \frac{\alpha^3t^3}{3!} + \dots & 0 \\ 0 & 1 + \beta t + \frac{\beta^2t^2}{2} + \frac{\beta^3t^3}{3!} + \dots \end{bmatrix} = \begin{bmatrix} e^{\alpha t} & 0 \\ 0 & e^{\beta t} \end{bmatrix}$$

For an nxn diagonal matrix $A = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$

$$e^{At} = \begin{bmatrix} e^{\lambda_1 t} & & O \\ & e^{\lambda_2 t} & \\ & \ddots & \\ O & & e^{\lambda_n t} \end{bmatrix}$$

If A is diagonalizable $\Rightarrow A = T\Lambda T^{-1}$ $\quad \Lambda = \begin{bmatrix} \lambda_1 & \\ & \lambda_n \end{bmatrix}$ $\quad A^k = T\Lambda^k T^{-1}$

$$e^{At} = I + At + \frac{A^2t^2}{2} + \frac{A^3t^3}{3!} + \dots$$

$$= TT^{-1} + T\Lambda T^{-1}t + T\Lambda^2T^{-1}\frac{t^2}{2} + \dots = T(I + \Lambda t + \frac{\Lambda t^2}{2} + \dots)T^{-1} = Te^{\Lambda t}T^{-1}$$

$$e^{At} = T \begin{bmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_n t} \end{bmatrix} T^{-1}$$

- block diagonal form
- exists for all square matrices

If A is not diagonalizable $\Rightarrow$ One can use Jordan Form instead of EVD.

$u(t) = 0$

$$T = [\underline{p}_1 \; \cdots \; \underline{p}_n] \qquad T^{-1} = \begin{bmatrix} \underline{q}_1^T \\ \vdots \\ \underline{q}_n^T \end{bmatrix}$$

$\dot{\underline{x}}(t) = A\underline{x}(t)$

$\underline{x}(t) = e^{A(t-t_0)}\underline{x}(t_0)$

$$e^{At} = [\underline{p}_1 \; \cdots \; \underline{p}_n] \begin{bmatrix} e^{\lambda_1 t} & & \\ & \ddots & \\ & & e^{\lambda_n t} \end{bmatrix} \begin{bmatrix} \underline{q}_1^T \\ \vdots \\ \underline{q}_n^T \end{bmatrix}$$

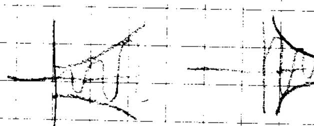
$$e^{At} = [e^{\lambda_1 t}\underline{p}_1 \;\; e^{\lambda_2 t}\underline{p}_2 \; \cdots \; e^{\lambda_n t}\underline{p}_n] \begin{bmatrix} \underline{q}_1^T \\ \vdots \\ \underline{q}_n^T \end{bmatrix}$$

→ Not Hermitian unless $\underline{p}_i \perp \underline{q}_j$

$$= e^{\lambda_1 t}\underset{n\times 1}{\underline{p}_1}\,\underset{1\times n}{\underline{q}_1^T} + e^{\lambda_2 t}\underline{p}_2\underline{q}_2^T + \cdots + e^{\lambda_n t}\underline{p}_n\underline{q}_n^T$$

$$= \sum_{k=1}^{n} e^{\lambda_k t}\underline{p}_k\underline{q}_k^T$$

$$\underline{x}(t) = e^{At}\underline{x}(0) = \sum_{k=1}^{n} e^{\lambda_k t}\underline{p}_k\underline{q}_k^T\underline{x}(0)$$



$$= \sum_{k=1}^{n} \langle \underline{x}(0), \underline{q}_k\rangle \underbrace{e^{\lambda_k t}\underline{p}_k}$$

> Eigenvalues of A should have negative real part = System is stable.

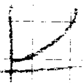
mode of the system – each mode has an associated exponent and direction.

$e^{(\sigma+j\omega)t}$

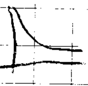
$e^{(\sigma-j\omega)t}$

$e^{\sigma t}\cos(\omega t+\theta)$

$\lambda_i \in \mathbb{R} > 0$



$\lambda_k \in \mathbb{R} < 0$



If $\langle\underline{x}(0), \underline{q}_k\rangle = 0$ the mode k will not be observed at $\underline{x}(t)$.

$\dot{\underline{x}}(t) = A\underline{x}(t) + B\underline{u}(t) \qquad t = t_0 \quad \underline{x}(t_0) \qquad \underline{u}(t),\ t \geq t_0$

$$\underline{x}(t) = \underbrace{e^{A(t-t_0)}\underline{x}(t_0)}_{\text{zero-input Response}} + \underbrace{\int_{t_0}^{t} e^{A(t-\tau)}B\underline{u}(\tau)\,d\tau}_{\text{zero-state Response}}$$

Time domain ←

Consider the scalar case

$x(t) = e^{a(t-t_0)}x(t_0) + \int_{t_0}^{t} b e^{a(t-\tau)}u(\tau)\,d\tau$

$\dot{\underline{x}}(t) = a x(t) + b u(t) \qquad h(t) = ?$

$sX(s) = aX(s) + bU(s)$

$(s-a)X(s) = bU(s)$

$X(s) = \dfrac{bU(s)}{(s-a)}$

$\dfrac{X(s)}{U(s)} = \dfrac{b}{s-a} \qquad h(t) = be^{at} \quad t \geq 0$

→ Try to take the derivative of this and see that $\underline{x}(t)' = \dot{\underline{x}}(t)$.

$$\underline{y}(t) = C\underline{x}(t) + D\underline{u}(t)$$

$$= Ce^{A(t-t_0)}\underline{x}(t_0) + \int_{t_0}^{t} Ce^{A(t-\tau)}B\underline{u}(\tau)\,d\tau + D\underline{u}(t)$$

$\int \delta(t-\tau)u(\tau)\,d\tau$

$$= Ce^{A(t-t_0)}\underline{x}(t_0) + \int_{t_0}^{t}\left(Ce^{A(t-\tau)}B + D\delta(t-\tau)\right)u(\tau)\,d\tau$$

$$= C e^{A(t-t_0)} \underset{\sim}{x}(t_0) + \int_{t_0}^{t} h(t-\tau) u(\tau)\, d\tau$$

$$\underbrace{\underset{\in \mathbb{R}^{p\times m}}{\underset{\sim}{h(t)}} = \overset{p\times n}{C}\, \underset{p\times m}{e^{\overset{n\times n}{A}t}\overset{n\times m}{B}} + D\delta(t)}_{\text{Impulse response}} \leftarrow (A,B,C,D)$$

Transform Domain Approach

C.T. Case

$$X(s) = \mathcal{L}_-\{\underset{\sim}{x}(t)\} = \begin{bmatrix} \mathcal{L}_-\{x_1(t)\} \\ \vdots \\ \mathcal{L}_-\{x_n(t)\} \end{bmatrix} \qquad t_0 = 0$$

$$\mathcal{L}_-\{f(t)\} = \int_{0^-}^{\infty} f(t)\, e^{-st}\, dt$$

Now given the state space eqs.

$$\mathcal{L}_-\{\dot{\underset{\sim}{x}}(t)\} = \mathcal{L}_-\{A\underset{\sim}{x}(t)\} + \mathcal{L}_-\{B\underset{\sim}{u}(t)\}$$

$$s\underset{\sim}{X}(s) - \underset{\sim}{x}(0) = A\underset{\sim}{X}(s) + B\underset{\sim}{U}(s)$$

$$sI\underset{\sim}{X}(s) - A\underset{\sim}{X}(s) = \underset{\sim}{x}(0) + BU(s)$$

$$(sI - A)\underset{\sim}{X}(s) = \underset{\sim}{x}(0) + BU(s)$$

↳ freq. domain → time domain

$$\underset{\sim}{X}(s) = (sI-A)^{-1}\underset{\sim}{x}(0) + (sI-A)^{-1}BU(s)$$

$$\underset{\sim}{Y}(s) = C\underset{\sim}{X}(s) + D\underset{\sim}{U}(s)$$

$$\underset{\sim}{Y}(s) = C(sI-A)^{-1}\underset{\sim}{x}(0) + \left(C(sI-A)^{-1}B + D\right)\underset{\sim}{U}(s)$$

$$y(t) = C e^{At}\underset{\sim}{x}(0) + \int_0^t \left(Ce^{A(t-\tau)}B + D\delta(t-\tau)\right) u(\tau)\, d\tau$$

$$\mathcal{L}_-\{e^{At}\} = (sI-A)^{-1}$$

$$H(s) = C(sI-A)^{-1}B + D = \mathcal{L}_-\{\overbrace{Ce^{At}B + D\delta(t)}^{h(t)}\}$$

↖ Transfer function

State Space

$$\begin{tikzcd} & \text{Impulse Response } h(t) = Ce^{At}B + D\delta(t) \\ (A,B,C,D) \arrow[ur] \arrow[dr] & \\ & \text{Transfer Function } H(s) = C(sI-A)^{-1}B + D \end{tikzcd}$$

Another way of calculating e^{At}.

$$= \frac{C\,\mathrm{Adj}(sI-A)B}{\det(sI-A)} + D$$

$$= \frac{C \,\mathrm{Adj}(sI-A)B + B \det(sI-A)}{\det(sI-A)}$$

Poles of $H(s)$ are when $\det(sI-A) \Rightarrow s$ is an eigenvalue of A.

$$\det(sI-A) = (s-\lambda_1)(s-\lambda_2) \cdots (s-\lambda_n)$$

$\Rightarrow$ If there is a factor in the numerator that would cancel $(s-\lambda_i)$,
λ_i will not be a pole of $H($
$\downarrow$
Pole zero cancellation.

Poles of $H(s) \subset$ eigenvalues of A

<u>MT Questions:</u>

$R(AA^T) \subseteq R(A) \rightarrow \boxed{R(AB) \subseteq R(A) \\ \subseteq R(B)}$

$AA^T\underset{\sim}{x} = A\underset{\sim}{y} \in R(A)$ $\quad \boxed{R(A) \neq R(A^T)}$

$\begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}$

$\underset{\sim}{x} \in N(AA^T)$ $AA^T\underset{\sim}{x} = \underset{\sim}{0}$

$\underset{\sim}{x}^T AA^T \underset{\sim}{x} = 0$

$\|A^T\underset{\sim}{x}\| = \underset{\sim}{0} \Rightarrow x \in N(A^T)$

$N(AA^T) \subseteq N(A^T)$

$R(A) \subseteq R(AA^T)$

$\Rightarrow \boxed{R(AA^T) = R(A)}$

<u>Matrix Transformations:</u>

Assume that we are transforming (multiplying) a given matrix $A \in \mathbb{C}^{m\times n}$ with another matrix (a transformation matrix) $T \in \mathbb{C}^{p\times m}$. Thus,

$$A_T = TA$$

Suppose we want the new matrix to have the same covariance as the previous one after being transformed.

Let $A = [\underset{\sim}{a}_1\ \underset{\sim}{a}_2 \ \cdots\ \underset{\sim}{a}_N]$, then the sample covariance matrix for the vector sequence is generated using the formula:

$$K_A = \frac{1}{N}\sum_{i=1}^{N} \underset{\sim}{a}_i \underset{\sim}{a}_i^* = \frac{1}{N} AA^*$$

$$K_{A_T} = \frac{1}{N} A_T A_T^* = \frac{1}{N} TAA^*T^* = TK_AT^*$$

$$\Rightarrow K_{A_T} - TK_AT^* = \underset{\sim}{0} \quad \Rightarrow K_{A_T} - TK_{A_T}T^* = \underset{\sim}{0}$$

Assume that K_{A_T} is P.D. Thus $K_{A_T} > 0$, then $K_{A_T} = R^*R = LL^*$ (via Cholesky factorization

$$K_{A_T} - \underbrace{TL}_{B}\underbrace{L^*T^*}_{B^*} = \underset{\sim}{0}$$

$$K_{A_T} = BB^*$$

$$B = LU \quad \text{where } U \text{ is unitary.}$$

$$TL = LU$$

$$T = LUL^{-1}$$

$$LUL^{-1}K_{A_T}L^{-*}U^*L^*$$

$$= LUL^{-1}LL^*L^{-*}U^*L^*$$

$$= LL^* = K_{A_T}$$

Without P.D. assumption:

$$K_{A_T} - TK_{A_T}T^* = \underset{\sim}{0}$$

$$B = TK_{A_T}^{1/2}$$

$$K_{A_T} - BB^* = \underset{\sim}{0}$$

$$B = K_{A_T}^{1/2}U$$

$$TK_{A_T}^{1/2} = K_{A_T}^{1/2}U$$

$$T = U$$

Therefore, to preserve the covariance matrix, T needs to be unitary. Furthermore, if K_{A_T} is P.D., then $T = LUL^{-1}$.

In some context, we may want to do the opposite of the coloring operation. Let's remember the coloring operation. If K_A is P.D., then we can use the Cholesky factorization such that $K_A = R^*R = LL^*$. Then for a given random matrix:

$$\underset{\sim}{z} = \text{randn}(m,n); \qquad E(\underset{\sim}{z}\underset{\sim}{z}^*) = I$$

$$\underset{\sim}{x} = L * \underset{\sim}{z} \qquad E(\underset{\sim}{x}\underset{\sim}{x}^*) = L\underset{\sim}{z}\underset{\sim}{z}^*L^* = LE(\underset{\sim}{z}\underset{\sim}{z}^*)L^* = LL^* = K_A.$$

Thus, we made the elements of $\underset{\sim}{x}$ correlated with K_A. $\underset{\sim}{x}$ is now considered as "colored", whereas $\underset{\sim}{z}$ can be considered as "white".

If we do the opposite operation, then this can be called decoloring, whitening, or sphereing:

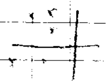
$$\underset{\sim}{z} = L^{-1} * \underset{\sim}{x}, \quad \text{where} \quad E(\underset{\sim}{x}\underset{\sim}{x}^*) = K_A = LL^*.$$

From a machine learning perspective, decoloring operation makes the within-class covariance isotropic (i.e. covariance matrix is proportional to I) and the distributions of classes close to spheres in the space. When this is the case, Fisher's linear discriminant (PRML, page 189, Christopher Bishop) gives us the projection direction along which we can separate the classes from each other.

all poles are on the left side

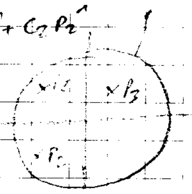
all poles are in the unit circ

LINEAR SYSTEM THEORY



$c_1 e^{p_1 t} + c_2 e^{p_2 t}$

$c_1 p_1^n + c_2 p_2^n$



C.T. | D.T.

$$\dot{\underset{\sim}{x}}(t) = A\underset{\sim}{x}(t) + B\underset{\sim}{u}(t) \qquad \underset{\sim}{x}_{k+1} = A\underset{\sim}{x}_k + B\underset{\sim}{u}_k$$

$$\underset{\sim}{y}(t) = C\underset{\sim}{x}(t) + D\underset{\sim}{u}(t) \qquad \underset{\sim}{y}_k = C\underset{\sim}{x}_k + D\underset{\sim}{u}_k$$

$$y(t) = \overset{p\times n}{C}\, \overset{n\times n}{e^{A(t-t_0)}}\, \overset{n\times 1}{\underset{\sim}{x}(t_0)} + \int_{t_0}^{t}\big(\underbrace{Ce^{A(t-\tau)}B + D\delta(t-\tau)}_{h(t-\tau)}\big)u(\tau)\,d\tau$$

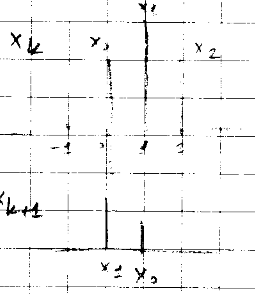
$$\underset{\sim}{y}_k = CA^{k-k_0}\underset{\sim}{x}_{k_0} + \sum_{i=k_0}^{k-1} CA^{k-i-1}B\underset{\sim}{u}_i +$$

$$h(t) = Ce^{At}B + D\delta(t)$$

$$Y(s) = C(sI-A)^{-1}\underset{\sim}{x}(t_0) + H(s)\underset{\sim}{U}(s)$$

$$\underset{p\times m}{H(s)} = C(sI-A)^{-1}B + D$$

Transform Domain Approach for D.T.



$$x_0 + x_1 z^{-1} + x_2 z^{-2} = \underset{\sim}{X}(z)$$

$$z\underset{\sim}{X}(z) - x_0 z + \underbrace{x_1 + x_2 z^{-1}}$$

$$x_1 + x_2 z^{-1} + x_3 z^{-2} \dots$$

$$Z\{x_{k+1}\} = Z\{A\underset{\sim}{x}_k + B\underset{\sim}{u}_k\}$$

$$z\underset{\sim}{X}(z) - z\underset{\sim}{x}_0 = A\underset{\sim}{X}(z) + B\underset{\sim}{U}(z)$$

$$z\underset{\sim}{X}(z) - A\underset{\sim}{X}(z) = z\underset{\sim}{x}_0 + B\underset{\sim}{U}(z)$$

$$(zI-A)\underset{\sim}{X}(z) = z\underset{\sim}{x}_0 + B\underset{\sim}{U}(z)$$

$$z\,\underset{\sim}{Z}\{\underset{\sim}{x}_k\} = zx_0 + x_1 + x_2 z^{-1} + \dots$$

$$\underset{\sim}{X}(z) = (zI-A)^{-1}z\underset{\sim}{x}_0 + (zI-A)^{-1}B\underset{\sim}{U}(z)$$

$$\underset{\sim}{Z}\{x_{k+1}\} = x_1 + x_2 z^{-1} +$$

$$\underset{\sim}{Y}(z) = C\underset{\sim}{X}(z) + D\underset{\sim}{U}(z)$$

$$\underset{\sim}{Y}(z) = \underbrace{C(zI-A)^{-1}z\,x_0}_{\text{zero-input response}} + \underbrace{(C(zI-A)^{-1}B+D)\,\underset{\sim}{U}(z)}_{\text{zero-state response}}$$

SISO

$$H(s) = \frac{n(s)}{d(s)}$$

$$\left.\begin{array}{l} H(s) = N(s)D^{-1}(s) \\ D^{-1}(s)\,N(s) \end{array}\right\} \text{Matrix fraction descriptions}$$

Normed and Inner Product Vector Spaces

Kolmogorov used these in estimation theory

Gives us a geometric approach to estimation theory.

Norm Spaces

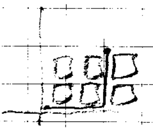
- Extension of the length concept in ~~3-D~~ 2-D Euclidian spaces to general vector spaces.
- Definition (Norm): Let S be a vector space with elements $\underset{\wedge}{x}$. A real valued function $\|\underset{\wedge}{x}\|$ is said to be a **norm** if $\|\underset{\wedge}{x}\|$ satisfies the following

1) $\|\underset{\wedge}{x}\| \geqslant 0$ for any $\underset{\wedge}{x} \in S$ (non-negative)
2) $\|\underset{\wedge}{x}\| = 0$ iff $\underset{\wedge}{x} = \underset{\wedge}{0}$ (definiteness)
3) $\|\alpha \underset{\wedge}{x}\| = |\alpha| \|\underset{\wedge}{x}\|$ (scaling)
4) $\|\underset{\wedge}{x} + \underset{\wedge}{y}\| \leq \|\underset{\wedge}{x}\| + \|\underset{\wedge}{y}\|$ (Triangle inequality)

Examples: Let $S = C^n$

$$\|\underset{\wedge}{x}\|_2 = \sqrt{|x_1|^2 + |x_2|^2 + \ldots + |x_n|^2} \quad \text{(Euclidian)}$$

$$\|\underset{\wedge}{x}\|_1 = |x_1| + |x_2| + \ldots + |x_n|$$



Manhattan distance

$$\|\underset{\wedge}{x}\|_p = \left(\sum_{k=1}^{n} |x_k|^p \right)^{1/p} \quad 1 \leq p < \infty$$

if $p < 1$, then triangle inequality is not satisfied

$$\|\underset{\wedge}{x}\|_\infty = \max_{k \in \{1, \ldots, n\}} |x_k|$$

Example: $\underset{\wedge}{x} = \begin{bmatrix} 3 \\ 4i \end{bmatrix}$

$$\|\underset{\wedge}{x}\|_1 = |3| + |4i| = 3 + 4 = 7$$

$$\|\underset{\wedge}{x}\|_2 = \sqrt{|3|^2 + |4i|^2} = 5$$

$$\|\underset{\wedge}{x}\|_\infty = 4$$

As $p \uparrow$ norm value $\downarrow$.

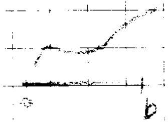
Example: Consider the vector space: Space Time: $\begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix}$

Is the following function a norm?

$$f(\underset{\wedge}{x}) = \sqrt{x^2 + y^2 + z^2 - \underset{\llcorner \text{ speed of light}}{c^2} t^2}$$

Doesn't satisfy definiteness property. Not a (definite) norm! But used in Einstein's formulations. This is an indefinite norm.

Example: For function spaces, such as continuous functions defined over $[a, b]$



$$\|\underline{x}(t)\|_1 = \int_a^b |x(t)|\, dt$$

$$\|\underline{x}(t)\|_2 = \left(\int_a^b |x(t)|^2 dt\right)^{1/2}$$

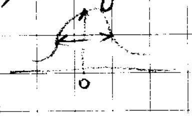
$$\|\underline{x}(t)\|_p = \left(\int_a^b |x(t)|^p dt\right)^{1/p}$$

$$d(\underline{x}(t), \underline{y}(t)) = \|x(t) - y(t)\| \quad \text{(Distance)}$$

Example: Vector Space of zero-mean random variables

$$\|\underline{x}\|^2 = Var(\underline{x}) = E(\underline{x}\,\underline{x}^*) \quad \text{(measures how big a r.v. is)}$$

If there is no variance, then your variable is not random. $Var(\underline{x})$ is a measure of uncertainty about $\underline{x}$.



Example: In C^n weighted norm

$$\|\underline{x}\|_w = (\underline{x}^* W \underline{x})^{1/2} \quad w \in C^{n \times n}$$

W needs to be Hermitian → real values
$W > 0$ s.t. $\|\underline{x}\|_w = 0$ iff $\underline{x} = \underline{0}$

Normed Vector Space: is a pair $(V, \|\cdot\|)$ where V is a vector space and $\|\cdot\|$ is a norm over V.

$$\lim_{k \to \infty} \underline{x}_k \to \underline{z} \quad \text{iff} \quad \lim_{k \to \infty} \|\underline{x}_k - \underline{z}\| \to 0$$

(use of norms allows such topological analysis tools)

Inner Products and Inner Product Spaces

Def: Let V be a vector space defined over a scalar field R. An inner product is a function $\langle \cdot, \cdot \rangle : V \times V \to R$ with the following properties

- $\langle \underline{x}, \underline{y} \rangle = \overline{\langle \underline{y}, \underline{x} \rangle}$ → complex conjugated
- $\langle \alpha \underline{x}, \underline{y} \rangle = \alpha \langle \underline{x}, \underline{y} \rangle$
- $\langle \underline{x} + \underline{y}, \underline{z} \rangle = \langle \underline{x}, \underline{z} \rangle + \langle \underline{y}, \underline{z} \rangle$
- $\langle \underline{x}, \underline{x} \rangle \geq 0$, $\langle \underline{x}, \underline{x} \rangle = 0$ iff $\underline{x} = \underline{0}$

$$\langle \underline{x}, \alpha \underline{y} \rangle = \overline{\langle \alpha \underline{y}, \underline{x} \rangle} = \overline{\alpha}\, \overline{\langle \underline{y}, \underline{x} \rangle}$$

Example: In C^n

$$\langle \underline{x}, \underline{y} \rangle = \underline{y}^* \underline{x} \quad \text{(Usual inner product)}$$
$$= \sum_{i=1}^{n} x_i y_i^*$$

Example: V = functions over [a, b]

$$\langle x(t), y(t)\rangle = \int_a^b x(t)\, y^*(t)\, dt$$

Example: V = m×n complex matrices

$$\langle A, B\rangle = \operatorname{Tr}(B^* A)$$

Example: $V = C^n$

$$\langle \underset{\sim}{x}, \underset{\sim}{y}\rangle_W = \underset{\sim}{y}^* W \underset{\sim}{x} \qquad W \in C^{n\times n}$$

$W > 0$ $\qquad \langle \underset{\sim}{x}, \underset{\sim}{x}\rangle \geq 0$, $\langle \underset{\sim}{x}, \underset{\sim}{x}\rangle = 0$ if $\underset{\sim}{x} = \underset{\sim}{0}$

$$\|A\|_2 = \max_{\|\underset{\sim}{x}\|_2 = 1} \|A\underset{\sim}{x}\|_2$$ (Induced 2-norm)

ΣV^*

$\sigma_1 \underset{\sim}{u}_1$

$$\|A\|_2 = \sigma_1$$

—Normed Spaces

—Inner Product Spaces

$(V, \langle\ \rangle)$ is inner product space where

- V: Vector space

- $\langle \cdot, \cdot \rangle$ is an inner product defined over that space

- Induced Norm: We can use inner product to produce a special norm called induced norm.

$$\|\underset{\sim}{x}\| = (\langle \underset{\sim}{x}, \underset{\sim}{x}\rangle)^{1/2}$$

Example: $V = C^{m\times n}$ $\quad A, B \in C^{m\times n}$

$$\langle A, B\rangle = \operatorname{Tr}(B^H A)$$

$$\|A\| = \langle A, A\rangle^{1/2}$$

$$\|A\|_F = (\operatorname{Tr}(A^H A))^{1/2} \leftarrow \text{Frobenius norm} = (\operatorname{Tr}(AA^H))^{1/2}$$

$$= \left(\sum_{i=1}^{n} \sum_{j=1}^{m} |A_{ji}|^2\right)^{1/2}$$

$$A = [\underset{\sim}{a}_1 \ \ \underset{\sim}{a}_2 \ \cdots \ \underset{\sim}{a}_n]$$

$$A^H A = \begin{bmatrix} \underset{\sim}{a}_1^* \\ \underset{\sim}{a}_2^* \\ \vdots \\ \underset{\sim}{a}_n^* \end{bmatrix} [\underset{\sim}{a}_1 \ \ \underset{\sim}{a}_2 \ \cdots \ \underset{\sim}{a}_n] \Bigg\} \operatorname{Tr}(A^H A) = \underset{\sim}{a}_1^* \underset{\sim}{a}_1 + \underset{\sim}{a}_2^* \underset{\sim}{a}_2 + \cdots + \underset{\sim}{a}_n^* \underset{\sim}{a}_n$$

$$= \|\underset{\sim}{a}_1\|_2^2 + \|\underset{\sim}{a}_2\|_2^2 + \cdots + \|\underset{\sim}{a}_n\|_2^2$$

We can define a vec operator given $A = [\underset{\sim}{a}_1 \ \underset{\sim}{a}_2 \ \dots \ \underset{\sim}{a}_n]$

$$\text{vec}(A) \triangleq \begin{bmatrix} \underset{\sim}{a}_1 \\ \underset{\sim}{a}_2 \\ \vdots \\ \underset{\sim}{a}_n \end{bmatrix} \in C^{mn \times 1}$$

$$\langle A, B \rangle = \text{vec}(B)^* \text{vec}(A)$$

$B = [\underset{\sim}{b}_1 \dots \underset{\sim}{b}_n]$

$A = [\underset{\sim}{a}_1 \dots \underset{\sim}{a}_n]$

$$\underset{\overset{\shortparallel}{B_{11}^* A_{11} + B_{12}^* A_{12} + \cdots}}{Tr(B^H A)} = Tr\left(\begin{bmatrix} \underset{\sim}{b}_1^* \\ \vdots \\ \underset{\sim}{b}_n^* \end{bmatrix} [\underset{\sim}{a}_1 \ \dots \ \underset{\sim}{a}_n] \right)$$

$$= \underset{\sim}{b}_1^* \underset{\sim}{a}_1 + \underset{\sim}{b}_2^* \underset{\sim}{a}_2 + \dots + \underset{\sim}{b}_n^* \underset{\sim}{a}_n$$

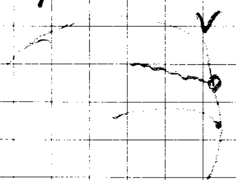
$$= \underbrace{[\underset{\sim}{b}_1^* \ \dots \ \underset{\sim}{b}_n^*]}_{\text{vec}(B)^*} \underbrace{\begin{bmatrix} \underset{\sim}{a}_1 \\ \vdots \\ \underset{\sim}{a}_n \end{bmatrix}}_{\text{vec}(A)}$$

$\|\underset{\sim}{x}\|_\infty$ does not have an associated inner product space.

Banach Space: Complete Normed Space

Hilbert Space: Complete Inner Product Space

Completeness: Every convergent sequence in V, converges to an element of V.



→ Not a complete vector space

[illegible] = volume [illegible]

Is matrix multiplication a complex inner product space?

Gram-Scmidt Applied to General Inner Product Spaces

example: V = Vector space of functions over $[-1, 1]$ for which

$$\int_{-1}^{1} |f(t)|^2 dt < \infty \quad \text{where } f(t) \in \mathbb{R}$$

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

$(V, \langle \cdot, \cdot \rangle) \rightarrow$ Inner product space

$$\| f(t) \| = \langle f(t), f(t) \rangle^{1/2} = \left(\int_{-1}^{1} |f(t)|^2 dt \right)^{1/2} \quad \text{(induced norm)}$$

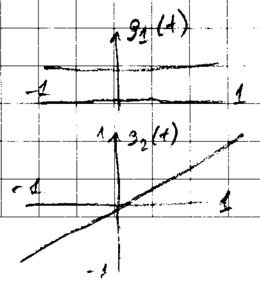
Let's have the set of functions

$g_1(t) = 1 \qquad t \in [-1, 1]$

$g_2(t) = t$

$\vdots$

$g_{n+1}(t) = t^n$



$G = \{g_1(t), \ldots, g_{n+1}(t)\}$ is linearly independent set.

Find $H = \{h_1(t), \ldots, h_{n+1}(t)\}$ such that $\mathrm{Span}(H) = \mathrm{Span}(G)$.

$$\langle h_i(t), h_j(t)\rangle = \delta_{i-j}$$

↑ polynomials ~~of~~ degree up to n

Let's obtain $h_i(t)$'s

$$h_1(t) = \frac{g_1(t)}{\|g_1(t)\|} \qquad \|g_1(t)\| = \left(\int_{-1}^{1} |g_1(t)|^2 dt\right)^{1/2} = \left(\int_{-1}^{1} 1\, dt\right)^{1/2} = \sqrt{2}$$

$$h_1(t) = \frac{1}{\sqrt{2}} \qquad t \in [-1, 1]$$

2. $$v_2(t) = g_2(t) - \overbrace{\underbrace{\langle g_2(t), h_1(t)\rangle}_{0} h_1(t)}^{\text{projection over } h_1}$$

$$\langle g_2(t), h_1(t)\rangle = \left(\int_{-1}^{1} t \frac{1}{\sqrt{2}}\, dt\right) = 0$$

$$V_2(t) = g_2(t)$$

$$h_2(t) = \frac{v_2(t)}{\|v_2(t)\|} \qquad \|v_2(t)\| = \|g_2(t)\| = \left(\int_{-1}^{1} t^2 dt\right)^{1/2} = \sqrt{\frac{2}{3}}$$

$$h_2(t) = \sqrt{\frac{3}{2}}\, t \qquad t \in [-1, 1]$$

3. $$v_3(t) = g_3(t) - \left(\langle g_3(t), h_1(t)\rangle h_1(t) + \langle g_3(t), h_2(t)\rangle h_2(t)\right)$$

$$\langle g_3(t), h_1(t)\rangle = \int_{-1}^{1} t^2 \frac{1}{\sqrt{2}}\, dt = \frac{1}{\sqrt{2}} \frac{t^3}{3}\Big|_{-1}^{1} = \frac{\sqrt{2}}{3}$$

$$\langle g_3, h_2\rangle = \int_{-1}^{1} t^2 \frac{\sqrt{3}}{\sqrt{2}} t\, dt = 0$$

$$v_3(t) = g_3(t) - \frac{\sqrt{2}}{3} \cdot \frac{1}{\sqrt{2}} = t^2 - \frac{1}{3}$$

$$h_3(t) = \frac{v_3(t)}{\|v_3(t)\|} = \frac{t^2 - \frac{1}{3}}{\left(\int_{-1}^{1} (t^4 - \frac{2}{3} t^2 + \frac{1}{9})\, dt\right)^{1/2}} =$$

$$\frac{2}{5} - \frac{4}{9} + \frac{2}{9} = \frac{8}{45}$$

$$\mathrm{Span}\overbrace{\{\underset{1}{v_1(t)}, \underset{t}{v_2(t)}, \underset{t^2 - \frac{1}{3}}{v_3(t)}, \ldots, v_{n+1}(t)\}}^{\text{orthogonal set}}$$

← Legendre polynomials.

$\|$

polynomials of degree up to n.

If we change the inner product definition to

$$\langle f(t), g(t)\rangle = \int_{-1}^{1} \frac{1}{\sqrt{1 - t^2}} f(t)\, g(t)\, dt$$

If you apply G.S. procedure to G you obtain another orthogonal set

$$T_n(t) = \cos((n-1)\cos^{-1}(t))$$

$T_1(t) = 1 \qquad T_3(t) = 2t^2 - 1$

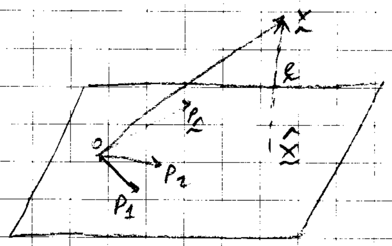
$T_2(t) = t \qquad \vdots$

} Chebyshev Polynomials

Whole goal of inner product spaces is to convert algebraic problems into geometric problems and projection is very important in this interpretati

Projection Theorem: Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space, and $\| \ \|$ be the induced norm. Let $p_1, p_2, \dots, p_n$ and $x \in V$. We call $\hat{x}$ as the projection of x on $\text{Span}\{p_1, \dots, p_n\}$ if it is the solution of

$$(*) \qquad \hat{x} = \underset{\hat{x} \in \text{Span}\{p_1, \dots, p_n\}}{\text{argmin}} \|x - \hat{x}\|$$



Projection theorem states that if

$$e = x - \hat{x} \Rightarrow \langle e, p_i \rangle = 0 \quad i = 1, \dots, n$$

Your original norm problem can be posed as an inner product problem when the norm used is an induced norm.

How to use projection thm. to solve our problem (*)

$$\hat{x} \in \text{Span}\{p_1, \dots, p_n\} \Rightarrow \hat{x} = \alpha_1 p_1 + \alpha_2 p_2 + \dots + \alpha_n p_n$$

Finding $\hat{x}$ = Finding $(\alpha_1 \dots \alpha_n)$ (n unknowns ⇒ we need atleast n equations)

Use $\langle e, p_i \rangle = 0$ $i = 1, \dots, n$ to find α_i's.

$$\langle x - \hat{x}, p_i \rangle = 0$$

$$\langle x, p_i \rangle = \langle \hat{x}, p_i \rangle = \langle \alpha_1 p_1 + \alpha_2 p_2 + \dots + \alpha_n p_n, p_i \rangle$$

$$= \alpha_1 \langle p_1, p_i \rangle + \alpha_2 \langle p_2, p_i \rangle + \dots + \alpha_n \langle p_n, p_i \rangle$$

$$i=1 \quad \alpha_1 \langle p_1, p_1 \rangle + \alpha_2 \langle p_2, p_1 \rangle + \dots + \alpha_n \langle p_n, p_1 \rangle = \langle x, p_1 \rangle$$

$$i=2 \quad \alpha_1 \langle p_1, p_2 \rangle + \alpha_2 \langle p_2, p_2 \rangle + \dots + \alpha_n \langle p_n, p_2 \rangle = \langle x, p_2 \rangle$$

$$\vdots$$

$$i=n \quad \alpha_1 \langle p_1, p_n \rangle + \alpha_2 \langle p_2, p_n \rangle + \dots + \alpha_n \langle p_n, p_n \rangle = \langle x, p_n \rangle$$

$$\begin{bmatrix} \langle p_1, p_1 \rangle & \langle p_2, p_1 \rangle & \cdots & \langle p_n, p_1 \rangle \\ & \langle p_2, p_2 \rangle & & \\ & & \ddots & \\ & & & \langle p_n, p_n \rangle \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} \langle x, p_1 \rangle \\ \langle x, p_2 \rangle \\ \vdots \\ \langle x, p_n \rangle \end{bmatrix}$$

For $A > 0 \Rightarrow \det(A) > 0 \Rightarrow \|A\|_{New} = \log(\det(A))$ can be defined.

Matrix Transformations (continued)

$$\overset{n \times N}{X = \{ x_1 \quad \tilde{x}_N \}} \qquad \tilde{x}_i \in C^n$$

$$R_X = \frac{1}{N} \sum_{k=1}^{N} \tilde{x}_k \tilde{x}_k^* = \frac{1}{N} X X^*$$

$$Y = T X$$

1. $R_Y = R_X \;,\; Y = T_1 X \;,\; T_1$ is unitary $\Rightarrow \text{Cond}(Y) = \text{Cond}(X)$

$$X = U \Sigma V^*$$

$$X X^* = U \Sigma \underbrace{V^* V}_{1} \Sigma^* U^* = U \Sigma^2 U^*$$

$$\Rightarrow \text{eig}(X X^*) = \text{sing}(X)^2$$

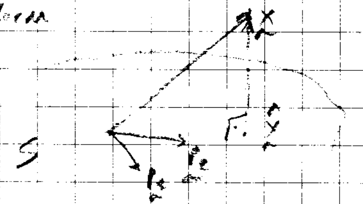
2. $R_Y = I \;,\; Y = T_2 X \;,\; \text{cond}(Y) = 1.$

Inner Product Spaces Projection Theorem

Inner product space

$(V, \langle \cdot , \cdot \rangle)$

Induced Norm $\| \; \| = \langle \cdot , \cdot \rangle^{1/2}$



Projection Theorem

$$\hat{x} \text{ minimizes } \|\tilde{x} - \hat{x}\| = \overset{\tilde{e}}{(\tilde{x} - \hat{x})} \perp^{\langle \cdot , \cdot \rangle} \text{Span}\{\tilde{p}_1, \ldots, \tilde{p}_n\}$$

$$\Downarrow$$

Core of the estimation theory $\Longrightarrow$

$$\begin{bmatrix} \langle \tilde{p}_1, \tilde{p}_1 \rangle & \langle \tilde{p}_2, \tilde{p}_1 \rangle & \cdots & \langle \tilde{p}_n, \tilde{p}_1 \rangle \\ \vdots & & & \\ \langle \tilde{p}_1, \tilde{p}_n \rangle & \cdots & & \langle \tilde{p}_n, \tilde{p}_n \rangle \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \langle \tilde{x}, \tilde{p}_1 \rangle \\ \vdots \\ \langle \tilde{x}, \tilde{p}_n \rangle \end{bmatrix}$$

$$S \subseteq V$$

$$\hat{x} = a_1 \tilde{p}_1 + \cdots + a_n \tilde{p}_n$$

Goal = Choose a_i's to minimize $\|\tilde{x} - \hat{x}\|$

If $\{\tilde{p}_1, \ldots, \tilde{p}_n\}$ is an orthonormal set then

$$\begin{bmatrix} \langle \tilde{p}_1, \tilde{p}_1 \rangle & \cdots & \langle \tilde{p}_n, \tilde{p}_1 \rangle \\ & & \vdots \\ \langle \tilde{p}_1, \tilde{p}_n \rangle & \cdots & \langle \tilde{p}_n, \tilde{p}_n \rangle \end{bmatrix} = I$$

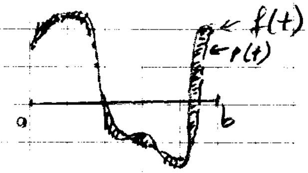
$\longrightarrow$ Grammian Matrix $= P^* P \longrightarrow$ PSD, Hermitian

$$\Rightarrow a_i = \langle \tilde{x}, \tilde{p}_i \rangle$$

Example: Find best n^{th} degree polynomial approximation of a given $f(t)$, $t \in [a,b]$ such that

$$\int_a^b (f(t) - \underset{\uparrow}{p(t)})^2 \, dt \text{ is minimized}$$

polynomial approximation. $p(t) = c_0 + c_1 t + \dots + c_n t^n$



One way: Write down the Taylor series expansion for $p(t)$ and cut it down.

Define $V \rightarrow$ vector space of functions $\{g(t) : t \in [a,b]\}$

when $\int_a^b (g(t))^2 dt < \infty$ → square integrable in the domain t

Define $<g,h> = \int_a^b g(t)\, h(t)\, dt$

Induced norm $\|g\| = <g, g>^{1/2} = \left(\int_a^b (g(t))^2 dt \right)^{1/2}$

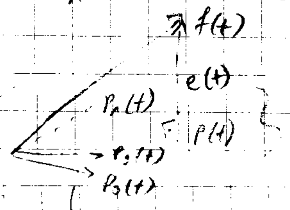
Our problem: $\underset{c_0, c_1, \dots, c_n}{\text{minimize}} \, \|\underbrace{f(t) - p(t)}_{e(t)}\|^2$

$$p(t) = c_0 + c_1 t + \dots + c_n t^n = c_0 P_0(t) + c_1 P_1(t) + \quad + c_n P_n(t)$$

$P_0(t) = 1$

$P_1(t) = t$

$P_n(t) = t^n$ } Basis for n^{th} degree polynomials



S = Space of n^{th} degree polynomials

Projection Theorem: $<e(t), P_i(t)> = 0 \quad i = 0, \dots, n$

$$\begin{bmatrix} <P_0(t), P_0(t)> & <P_1(t), P_0(t)> & \cdots & <P_n(t), P_0(t)> \\ <P_0(t), P_1(t)> & <P_1(t), P_1(t)> & \cdots & <P_n(t), P_1(t)> \\ \vdots & & & \\ <P_0(t), P_n(t)> & <P_1(t), P_n(t)> & \cdots & <P_n(t), P_n(t)> \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} <f(t), P_0(t)> \\ <f(t), P_1(t)> \\ \vdots \\ <f(t), P_n(t)> \end{bmatrix}$$

$$<P_i(t), P_j(t)> = \int_a^b t^{i+j} \, dt$$

Let $[a,b] = [0,1]$

$$<P_i(t), P_j(t)> = \int_0^1 t^{i+j} \, dt = \frac{1}{i+j+1}$$

$$\underset{\substack{\downarrow \\ \text{Hilbert matrix}}}{\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n+1} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & & \vdots \\ \frac{1}{3} & & \vdots & & \vdots \\ \vdots & & & & \\ \frac{1}{n+1} & \frac{1}{n+2} & \frac{1}{n+3} & \cdots & \frac{1}{2n+1} \end{bmatrix}} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} = \underset{\substack{\uparrow \\ \text{Moments of the function}}}{\begin{bmatrix} \int_0^1 f(t)\, dt \\ \int_0^1 t\, f(t)\, dt \\ \vdots \\ \int_0^1 t^n f(t)\, dt \end{bmatrix}}$$

$\downarrow$

ill conditioned for $n \longrightarrow \infty$

> Harder to find the inverse

If we use Legendre polynomials instead $\{l_0(t), l_1(t), \dots, l_n(t)\}$

$$p(t) = b_0 l_0(t) + b_1 l_1(t) + \cdots + b_n l_n(t)$$

$$\langle l_i(t), l_j(t) \rangle = 0 \iff i \neq j$$

$$\begin{bmatrix} \langle l_0(t), l_0(t)\rangle & & \bigcirc \\ & \langle l_1(t), l_1(t)\rangle & \\ \bigcirc & & \ddots \quad \langle l_n(t), l_n(t)\rangle \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} \int_0^1 f(t)\, l_0(t)\, dt \\ \vdots \\ \int_0^1 f(t)\, l_n(t)\, dt \end{bmatrix}$$

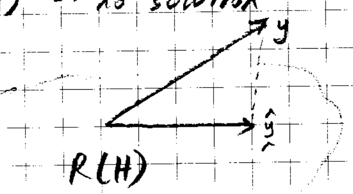
$$b_i = \frac{\langle f(t), l_i(t) \rangle}{\langle l_i(t), l_i(t) \rangle}$$ → we can easily scale this

DETERMINISTIC LEAST SQUARES (Ch 2 of Kailath) → start reading this chapter

Suppose $H\underset{\sim}{x} = \underset{\sim}{y}$ $\quad H \in \mathbb{C}^{N \times n} \quad N \gg n$

if $\underset{\sim}{y} \in R(H)$ → solution exists

if $\underset{\sim}{y} \notin R(H)$ → no solution → There is some noise. I still want to pursue this.



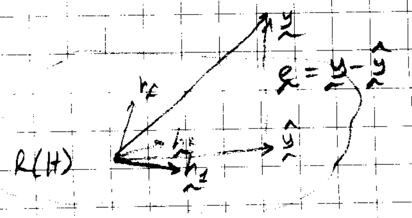
In $V = \mathbb{C}^n$, we are trying to minimize

$$\|\underset{\sim}{y} - H\underset{\sim}{x}\|_2 \quad \text{where} \quad \|\underset{\sim}{a}\| = (\underset{\sim}{a}^* \underset{\sim}{a})^{1/2}$$

induced by

$$\langle \underset{\sim}{a}, \underset{\sim}{b} \rangle = \underset{\sim}{b}^* \underset{\sim}{a}$$

$$H = [\underset{\sim}{h_1} \ \dots \ \underset{\sim}{h_n}] \Rightarrow H\underset{\sim}{x} = h_1 \underset{\sim}{x_1} + h_2 \underset{\sim}{x_2} + \cdots + h_n \underset{\sim}{x_n}$$



Goal: $\underset{x_1, \dots, x_n}{\text{minimize}} \|\underset{\sim}{e}\|_2 \overset{\text{P.T.}}{\Longrightarrow} \underset{\sim}{e} \perp R(H)$

$$\Longrightarrow \underset{\sim}{e} \perp \underset{\sim}{h_i} \quad i = 1, \dots, n$$

$$\Longrightarrow \langle \underset{\sim}{e}, \underset{\sim}{h_i} \rangle = 0$$

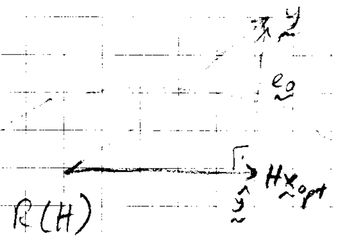
$\langle \underset{\sim}{e_0}, \underset{\sim}{h_i} \rangle = 0$ $\qquad \underset{\sim}{e_0}$ = optimal $\underset{\sim}{e}$

$\underset{\sim}{h_i}^* \underset{\sim}{e_0} = 0 \quad i = 1, \dots, n$ $\qquad \underset{\sim}{e_0} = \underset{\sim}{y} - H\underset{\sim}{x}_{opt}$

$$\left.\begin{matrix} \underset{\sim}{h_1}^* \underset{\sim}{e_0} = 0 \\ \underset{\sim}{h_2}^* \underset{\sim}{e_0} = 0 \\ \vdots \\ \underset{\sim}{h_n}^* \underset{\sim}{e_0} = 0 \end{matrix}\right\} \quad \underbrace{\begin{bmatrix} \underset{\sim}{h_1}^* \\ \underset{\sim}{h_2}^* \\ \vdots \\ \underset{\sim}{h_n}^* \end{bmatrix}}_{H^*} \underset{\sim}{e_0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \rightarrow H^* \underset{\sim}{e_0} = \underset{\sim}{0} \Rightarrow H^*(\underset{\sim}{y} - H\underset{\sim}{x}_{opt}) = 0$$

$$\Rightarrow \underset{n\times N}{H^*}\underset{N\times n}{y} - \underset{n\times N}{H^*}\underset{N\times 1}{H}\,\hat{x}_{opt} = 0$$

$$\Rightarrow \boxed{\underset{n\times n}{H^*H}\,\underset{n\times 1}{x_{opt}} = \underset{n\times 1}{H^*y}}$$ Normal equations



$H^*\hat{y} \in R(H^*H)$

$\Rightarrow$ always consistent (There is always a solution)

Why? $\underline{R(H^*H)} = R(H^*)$

prove this!

* Normal equations always have a solution

* Uniqueness?

if $N(H^*H) = \{0\}$, solution is unique.

$\Updownarrow$

H is full rank $(N(H) = \{0\})$

Tall; columns are indep.

$x_{opt} = (H^*H)^{-1}H^*y$

$$\hat{y} = Hx_{opt} = \underbrace{H(H^*H)^{-1}H^*}_{P_H \Rightarrow \text{orthogonal projection matrix}}y$$

$P_H^2 \overset{?}{=} P_H$ & $P_H = P_H^*$

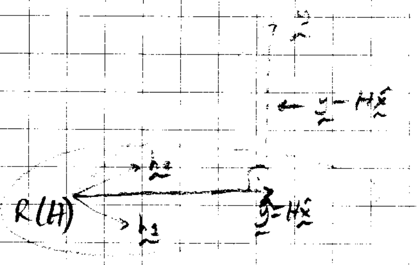
if H is not full rank

$\Rightarrow$ infinitely many solutions

Least Squares

$Hx \cong y$

$$\underset{x \in C^n}{\text{minimize}} \; \|y - \underbrace{Hx}_{h_1x_1 + h_2x_2 + \cdots + h_nx_n}\|_2$$



P.T. $y - Hx \perp h_i \quad i=1,\dots,n$

$$\begin{bmatrix} h_1^* \\ \vdots \\ h_n^* \end{bmatrix}(y - Hx) = 0$$

$\boxed{H^*H\hat{x} = H^*y}$ ← Normal equations

pseudoinverse of H

$H \rightarrow$ full rank $\Rightarrow (H^*H) \rightarrow$ full rank $\Rightarrow \hat{x} = (H^*H)^{-1}H^*y$

$H \rightarrow$ not full rank $\Rightarrow \hat{y}$ is still unique but $\hat{x}$ is not unique $\quad \hat{y} = \underbrace{H(H^*H)^{-1}H^*}y$

orthogonal projection matrix

If H is not full rank, then

1. $H = \hat{Q}\hat{R}$ $\qquad \hat{Q}x' = y$

full rank $\quad \hat{y} = \hat{Q}(\hat{Q}^* \hat{Q})^{-1}\hat{Q}^* y$

2. Suppose you know the rank of H. Then you can search for the columns that are not independent from the rest.

3. $\hat{y} = H \underbrace{(H^* H)^{-1}}_{} H^* y$

Take the pseudo-inverse by SVD's pseudo-inverse, which is a better option

Weighted Least Squares

$$e = y - Hx = \begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix}$$

$$\|y - Hx\|_W^2 = (y - Hx)^* W (y - Hx)$$

$$W > 0, \quad W = U \Lambda U^*$$

$$\underset{\text{eigenvectors}}{[u_1 \cdots u_n]} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} U^*$$

$$e^* W e$$

$$e^* [u_1 \cdots u_n] \Lambda \underbrace{\begin{bmatrix} u_1^* \\ \vdots \\ u_n^* \end{bmatrix} e}_{}$$

$$[\langle e, u_1\rangle \ldots \langle e, u_n\rangle] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} \langle e, u_1\rangle \\ \vdots \\ \langle e, u_n\rangle \end{bmatrix} \rightarrow \text{errors in each direction}$$

$$= \sum_{k=1}^{N} \lambda_k |\underbrace{\langle e, u_k\rangle}_{\text{If } U = I,\ e_i}|^2$$

$$\|a\|_W^2 = a^* W a = \langle a, a\rangle \qquad a \in \mathbb{C}^N$$

What should be the new inner product?

$$\langle a, b\rangle_W = b^* W a \quad \leftarrow \text{New inner product}$$

$$\hat{x} \text{ is minimizing } \|y - Hx\|_W^2 \xRightarrow{\text{P.T.}} (y - H\hat{x}) \perp R(H)$$

$$h_i \quad i = 1, \ldots, n$$

$$\Downarrow$$

$$\langle y - H\hat{x}, h_i\rangle_W \quad i = 1, \ldots, n$$

$$h_i^* W (y - H\hat{x}) = 0$$

$$\left.\begin{array}{l} h_1^* W (y - H\hat{x}) = 0 \\ h_2^* W (y - H\hat{x}) = 0 \\ \vdots \end{array}\right\} \quad \overbrace{\begin{bmatrix} h_1^* \\ \vdots \\ h_n^* \end{bmatrix}}^{H^*} W (y - H\hat{x}) = 0$$

$W > 0 \Leftrightarrow W = LL^*$ (cholesky)

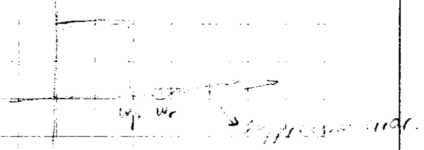
$\det(H^*WH) = $ [illegible] $\ne 0$ [illegible] $\ne 0$

$$\boxed{H^*WH\hat{x} = H^*Wy}$$ Normal Equations

If H is full rank $\Rightarrow$ H^*WH is invertible (show this!)

$$\hat{x} = (H^*WH)^{-1}H^*Wy$$

HW: Filter design



$W = I$ or W is general

Suppose H^*WH is not invertible, then:

$$\underbrace{(\epsilon I + H^*WH)}_{\text{Regularization}}$$

A^*A is always PSD. If H is not full rank H^*WH is PSD.

ϵI is PD $\Rightarrow$ $\epsilon I + H^*WH > 0$

$$\underset{A>0\ \ B\geq 0}{x^*(A+B)x} \qquad x \ne 0$$

$$\underbrace{\underbrace{x^*Ax}_{>0} + \underbrace{x^*Bx}_{\geq 0}}_{>0}$$

Regularized Least Squares

The new cost function

$$J(x) = \|y - Hx\|_W^2 + \underbrace{(x - x_0)^* \overset{\text{a weighting matrix}}{\Pi_0^{-1}} (x - x_0)}_{\text{Measure of distance of } x \text{ from } x_0} \qquad \Pi_0 > 0$$

Trying to keep x as close to x_0 as possible.

$\Pi_0 = 0.0001 \times I$ vs. $10{,}000 \times I$.

★ If we don't have enough information to fix x to a single point ($H^*WH \geq 0$), then keep x as close to x_0 as possible. (x_0 = a priori information)

Π_0^{-1} reflects confidence about apriori information. (Maybe prob. of multiple translations) ★

Π_0 is not necessarily diagonal. Then, in some directions you have more confidence.

Π_0 can be a certain covariance matrix.

$x_0 \approx$ apriori guess about the solution

$$J(x) = \underset{\substack{\text{Linear model} \\ \text{(weather } \underset{\text{last 2 days}}{\text{today}})}}{L} + \underset{\substack{\text{statistics from the past} \\ \text{(weather statistics)}}}{R}$$

$$\underset{\sim}{x}' = \underset{\sim}{x} - \underset{\sim}{x}_0$$

$$\underset{\sim}{y} - H\underset{\sim}{x} = \underset{\sim}{y} - H(\underset{\sim}{x}' + \underset{\sim}{x}_0)$$

$$= \underbrace{(\underset{\sim}{y} - H\underset{\sim}{x}_0)}_{\underset{\sim}{y}'} - H\underset{\sim}{x}'$$

$$= \underset{\sim}{y}' - H\underset{\sim}{x}' \qquad X X^*$$

→ There can be ~~many~~ such factorizations. One is Cholesky.

$$J(\underset{\sim}{x}') = \|\underset{\sim}{y}' - H\underset{\sim}{x}\|_W^2 + \underset{\sim}{x}'^* \overset{n\times n}{\Pi_0^{-1}} \underset{\sim}{x}' \qquad \Pi_0 = \Pi_0^{1/2}\underbrace{(\Pi_0^{1/2})^*}_{\Pi_0^{*/2}}$$

$\underset{\sim}{y} \in C^N$

$$= (\underset{\sim}{y}' - H\underset{\sim}{x}')^* W (\underset{\sim}{y}' - H\underset{\sim}{x}') + \underset{\sim}{x}'^* \Pi_0^{-*/2} \Pi_0^{-1/2} \underset{\sim}{x}' \qquad \Pi_0^{-1} = \Pi_0^{-*/2} \Pi_0^{-1/2}$$

$\underset{\sim}{x} \in C^n$

$$= (\underset{\sim}{y}' - H\underset{\sim}{x}')^* W (\underset{\sim}{y}' - H\underset{\sim}{x}') + (\underset{\sim}{0} - \Pi_0^{-1/2}\underset{\sim}{x}')^* I (\underset{\sim}{0} - \Pi_0^{-1/2}\underset{\sim}{x}')$$

$$\underset{\sim}{a}^* P_1 \underset{\sim}{a} + \underset{\sim}{b}^* P_2 \underset{\sim}{b} = \begin{bmatrix} \underset{\sim}{a} \\ \underset{\sim}{b} \end{bmatrix}^* \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix} \begin{bmatrix} \underset{\sim}{a} \\ \underset{\sim}{b} \end{bmatrix}$$

$$= \left(\begin{bmatrix} \underset{\sim}{y}' \\ \underset{\sim}{0} \end{bmatrix} - \begin{bmatrix} H \\ \Pi_0^{-1/2} \end{bmatrix} \underset{\sim}{x}' \right)^* \begin{pmatrix} W & 0 \\ 0 & I \end{pmatrix} \left(\begin{bmatrix} \underset{\sim}{y}' \\ \underset{\sim}{0} \end{bmatrix} - \begin{bmatrix} H \\ \Pi_0^{-1/2} \end{bmatrix} \underset{\sim}{x}' \right)$$

$$= \left\| \begin{bmatrix} \underset{\sim}{y}' \\ \underset{\sim}{0} \end{bmatrix} - \begin{bmatrix} \overset{N\times n}{H} \\ \overset{n\times n}{\Pi_0^{-1/2}} \end{bmatrix} \underset{\sim}{x}' \right\|^2_{\begin{bmatrix} W & 0 \\ 0 & I \end{bmatrix}}$$

Let's define an inner product space

$$(V, \langle , \rangle) : \quad V = C^{N+n \times 1}$$

$$\underset{\sim}{a}, \underset{\sim}{b} \in V \Rightarrow \langle \underset{\sim}{a}, \underset{\sim}{b} \rangle_{\begin{bmatrix} W & 0 \\ 0 & I \end{bmatrix}} \triangleq \underset{\sim}{b}^* \underbrace{\begin{bmatrix} \overset{N\times N}{W} & 0 \\ 0 & \overset{n\times n}{I} \end{bmatrix}}_{(N+n)\times(N+n)} \underset{\sim}{a}$$

$$\text{D.T.} \quad \left(\begin{bmatrix} \underset{\sim}{y}' \\ \underset{\sim}{0} \end{bmatrix} - \begin{bmatrix} H \\ \Pi_0^{-1/2} \end{bmatrix} \underset{\sim}{x}' \right) \perp R\left(\begin{bmatrix} H \\ \Pi_0^{-1/2} \end{bmatrix} \right)$$

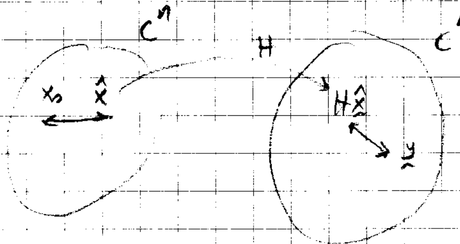
$$\begin{bmatrix} H^* & \Pi_0^{-*/2} \end{bmatrix} \begin{bmatrix} W & 0 \\ 0 & I \end{bmatrix} \left(\begin{bmatrix} \underset{\sim}{y}' \\ \underset{\sim}{0} \end{bmatrix} - \begin{bmatrix} H \\ \Pi_0^{-1/2} \end{bmatrix} \hat{\underset{\sim}{x}}' \right) = \underset{\sim}{0}$$

$$\begin{bmatrix} H^* W & \Pi_0^{-*/2} \end{bmatrix} \left(\begin{bmatrix} \underset{\sim}{y}' \\ \underset{\sim}{0} \end{bmatrix} - \begin{bmatrix} H \\ \Pi_0^{-1/2} \end{bmatrix} \hat{\underset{\sim}{x}}' \right) = \underset{\sim}{0}$$

$$\boxed{(H^* W H + \Pi_0^{-1}) \hat{\underset{\sim}{x}}' = H^* W \underset{\sim}{y}'}$$

$$\underset{\sim}{\hat{x}}' = (H^* W H + \Pi_0^{-1})^{-1} H^* W \underset{\sim}{y}'$$

$$\underset{\sim}{\hat{x}} = (H^* W H + \Pi_0^{-1})^{-1} H^* W (\underset{\sim}{y} - H \underset{\sim}{x_0}) + \underset{\sim}{x_0}$$



If multiple $\underset{\sim}{x_0}$ initial guesses, then $\underset{\sim}{a}^* P_1 \underset{\sim}{a} + \underset{\sim}{b}^* P_2 \underset{\sim}{b} + \underset{\sim}{c}^* P_3 \underset{\sim}{c} + \ldots$ Σ Bol

→ [illegible] olur mu?

Indefinite [illegible] $\Pi_0 \geq 0$ olabilir $\Pi_0 \to$ [illegible]

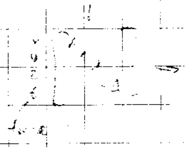
→ Birden fazla [illegible] olabilir

Bir Hilbert Space belirleyor

Krene space → PD değil

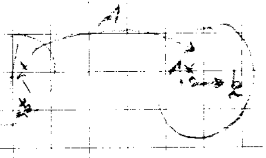
↑ Indefinite metric olan bir space

Einstein da kullanıyor → Minkowski distance



→ Işık hızı sabit [illegible] gözlemci [illegible]

DETERMINISTIC LEAST SQUARES



$$A \underset{\sim}{x} \overset{?}{=} \underset{\sim}{b}$$

Ordinary L.S.	Weighted L.S.	Regularized L.S.
$\min_{\underset{\sim}{x}} \lVert A\underset{\sim}{x} - \underset{\sim}{b} \rVert_2$	$\min_{\underset{\sim}{x}} \lVert A\underset{\sim}{x} - \underset{\sim}{b} \rVert_W$	$J(\underset{\sim}{x}) = \lVert A\underset{\sim}{x} - \underset{\sim}{b} \rVert_W^2 + (\underset{\sim}{x} - \underset{\sim}{x_0})^* \Pi_0^{-1} (\underset{\sim}{x} - \underset{\sim}{x_0})$
$\underset{\sim}{x} = \underbrace{(A^*A)^{-1} A^*}_{\text{pseudo-inverse of } A} \underset{\sim}{b}$	$\underset{\sim}{\hat{x}} = (A^* W A)^{-1} A^* W \underset{\sim}{b}$	$\underset{\sim}{\hat{x}} = \underset{\sim}{x_0} + (\Pi_0^{-1} + A^* W A) A^* W (\underset{\sim}{b} - A \underset{\sim}{x_0})$

UPDATING LEAST SQUARES SOLUTIONS: RLS ALGORITHM

$$H \underset{\sim}{x} \cong \underset{\sim}{y} \qquad H_{i-1} = \begin{bmatrix} \underset{\sim}{h_0}^* \\ \vdots \\ \underset{\sim}{h_{i-1}}^* \end{bmatrix} \qquad \underset{\sim}{y_{i-1}} = \begin{bmatrix} y_0 \\ \vdots \\ y_{i-1} \end{bmatrix}$$

$\underset{\sim}{\hat{x}_{i-1}} = \operatorname{argmin} \lVert H_{i-1} \underset{\sim}{x} - \underset{\sim}{y_{i-1}} \rVert_2$ or $\underset{\sim}{\hat{x}_{i-1}}$ is the solution of regularized LS for $(H_{i-1}, \underset{\sim}{y_{i-1}})$.

$$H_i = \begin{bmatrix} H_{i-1} \\ \underset{\sim}{h_i}^* \end{bmatrix} \qquad \underset{\sim}{y_i} = \begin{bmatrix} \underset{\sim}{y_{i-1}} \\ y_i \end{bmatrix}$$

$$\underset{\sim}{\hat{x}_i} = ?$$

Suppose $\underset{\sim}{\hat{x}}_{i-1}$ is the solution of

$$\min_{\underset{\sim}{x}} \left(\underset{\sim}{x}^* \Pi_0^{-1} \underset{\sim}{x} + \| \underset{\sim}{y}_{i-1} - H_{i-1} \underset{\sim}{x} \|_2^2 \right)$$

→ I instead of W. In the more general case, W needs to extend to incorporate the new dimensions

We want to obtain the solution $\underset{\sim}{\hat{x}}_i$ for

$$\min_{\underset{\sim}{x}} \left(\underset{\sim}{x}^* \Pi_0^{-1} \underset{\sim}{x} + \| \underset{\sim}{y}_i - H_i \underset{\sim}{x} \|_2^2 \right)$$

$$\underset{\sim}{x}_{i-1} = (\Pi_0^{-1} + H_{i-1}^* H_{i-1})^{-1} H_{i-1}^* \underset{\sim}{y}_{i-1}$$

$$H_i = \begin{bmatrix} H_{i-1} \\ \underset{\sim}{h}_i^* \end{bmatrix} \qquad \underset{\sim}{y}_i = \begin{bmatrix} \underset{\sim}{y}_{i-1} \\ y_i \end{bmatrix}$$

$$\underset{\sim}{x}_i = (\Pi_0^{-1} + H_i^* H_i)^{-1} H_i^* \underset{\sim}{y}_i$$

$$\underset{\sim}{\hat{x}}_i = \left(\Pi_0^{-1} + [H_{i-1}^* \ \underset{\sim}{h}_i] \begin{bmatrix} H_{i-1} \\ \underset{\sim}{h}_i^* \end{bmatrix} \right)^{-1} [H_{i-1}^* \ \underset{\sim}{h}_i] \begin{bmatrix} \underset{\sim}{y}_{i-1} \\ y_i \end{bmatrix}$$

$$= (\Pi_0^{-1} + H_{i-1}^* H_{i-1} + \underset{\sim}{h}_i \underset{\sim}{h}_i^*)^{-1} (H_{i-1}^* \underset{\sim}{y}_{i-1} + \underset{\sim}{h}_i y_i)$$

$$P_i = (\Pi_0^{-1} + H_i^* H_i)^{-1} \qquad P_{i-1} = (\Pi_0^{-1} + H_{i-1}^* H_{i-1})^{-1}$$

$$= (P_{i-1}^{-1} + \underset{\sim}{h}_i \underset{\sim}{h}_i^*)^{-1}$$

$$= P_{i-1} - P_{i-1} \underset{\sim}{h}_i (\overset{1\times 1}{I} + \overset{1\times n}{\underset{\sim}{h}_i^*} \overset{n\times n}{P_{i-1}} \overset{n\times 1}{\underset{\sim}{h}_i})^{-1} \underset{\sim}{h}_i^* P_{i-1}$$

$$(A + BCD)^{-1} = A^{-1} - A^{-1} B (C^{-1} + D A^{-1} B)^{-1} D A^{-1}$$

$$P_i = P_{i-1} - \frac{P_{i-1} \underset{\sim}{h}_i \underset{\sim}{h}_i^* P_{i-1}}{(1 + \underset{\sim}{h}_i^* P_{i-1} \underset{\sim}{h}_i)}$$

Note that

$$\underset{\sim}{\hat{x}}_{i-1} = P_{i-1} H_{i-1}^* \underset{\sim}{y}_{i-1}$$

$$\underset{\sim}{\hat{x}}_i = P_i H_i^* \underset{\sim}{y}_i = P_i (H_{i-1}^* \underset{\sim}{y}_{i-1} + \underset{\sim}{h}_i y_i)$$

$$\underset{\sim}{\hat{x}}_i = \left(P_{i-1} - \frac{P_{i-1} \underset{\sim}{h}_i \underset{\sim}{h}_i^* P_{i-1}}{(1 + \underset{\sim}{h}_i^* P_{i-1} \underset{\sim}{h}_i)} \right) (H_{i-1}^* \underset{\sim}{y}_{i-1} + \underset{\sim}{h}_i y_i)$$

$$= P_{i-1} H_{i-1}^* \underset{\sim}{y}_{i-1} + P_{i-1} \underset{\sim}{h}_i y_i - \frac{P_{i-1} \underset{\sim}{h}_i \underset{\sim}{h}_i^* \overbrace{P_{i-1} H_{i-1}^* \underset{\sim}{y}_{i-1}}^{\underset{\sim}{\hat{x}}_{i-1}}}{1 + \underset{\sim}{h}_i^* P_{i-1} \underset{\sim}{h}_i} - \frac{P_{i-1} \underset{\sim}{h}_i \underset{\sim}{h}_i^* P_{i-1}}{1 + \underset{\sim}{h}_i^* P_{i-1} \underset{\sim}{h}_i} \underset{\sim}{h}_i y_i$$

$$= \underset{\sim}{\hat{x}}_{i-1} + P_{i-1} \underset{\sim}{h}_i \underbrace{\left(1 - \frac{\underset{\sim}{h}_i^* P_{i-1} \underset{\sim}{h}_i}{1 + \underset{\sim}{h}_i^* P_{i-1} \underset{\sim}{h}_i} \right)}_{\frac{1}{1 + \underset{\sim}{h}_i^* P_{i-1} \underset{\sim}{h}_i}} y_i - \frac{P_{i-1} \underset{\sim}{h}_i \underset{\sim}{h}_i^* \underset{\sim}{\hat{x}}_{i-1}}{(1 + \underset{\sim}{h}_i^* P_{i-1} \underset{\sim}{h}_i)}$$

$$\boxed{\begin{array}{l} \hat{\underset{\sim}{x}}_i = \hat{\underset{\sim}{x}}_{i-1} + \dfrac{P_{i-1} \underset{\sim}{h_i}}{(1+\underset{\sim}{h_i}^* P_{i-1} \underset{\sim}{h_i})} \left(y_i - \underset{\sim}{h_i}^* \hat{\underset{\sim}{x}}_{i-1} \right) \\ O(m^3) \quad P_i = P_{i-1} - \dfrac{P_{i-1} \underset{\sim}{h_i} \underset{\sim}{h_i}^* P_{i-1}}{(1+\underset{\sim}{h_i}^* P_{i-1} \underset{\sim}{h_i})} \qquad P_0 = \Pi_0 \end{array}}$$

A special case of Kalman filtering.

$H_N = \begin{bmatrix} H_{N-1} \\ h_N \end{bmatrix}$ Amount of computation to solve

$$\|\underset{\sim}{y} - H\underset{\sim}{x}\|^2 + \underset{\sim}{x}\,\Pi_0^{-1}\underset{\sim}{x} \qquad O(\underset{\substack{\uparrow \\ 1000-10{,}000}}{N^2}\overset{\nwarrow}{m}{}_{10-50})$$

$$O(m^3) < O(N^2 m) \text{ in general.}$$

If H_N has a special structure, $O(m^3)$ can be reduced to $O(m)$.

Fast algorithms exists. If H_N is a convolution matrix, it has a shift structure.
(presented in the book in a section)

$$H = \begin{bmatrix} h_0 & \cdot & h_L & & \\ h_{-1} & h_0 & \cdot & h_L & \\ h_{-2} & h_{-1} & h_0 & \cdots & h_L \\ & & \cdots & & \end{bmatrix} \quad \text{Toeplitz matrix} \qquad \text{FFT ?}$$

STOCHASTIC ESTIMATION

$\nearrow$ (models the ambiguity)
Probabilistic Info.
\+
$\searrow$ Observations
(using the structure)
(current measurements)

Die throwing
$p(\text{side}) = 1/6$
outcome is even
$P(\cdot) = \frac{1}{3} = P(\cdot) = P(\cdot)$

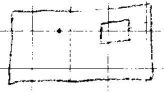
Preliminary Info:

Random Vectors

$$\text{Probability Space } (\underset{\substack{\uparrow \\ \text{sample} \\ \text{space}}}{\mathbb{R}^n}, \underset{\substack{\uparrow \\ \text{event} \\ \text{space}}}{B(\mathbb{R}^n)}, \underset{\substack{\uparrow \\ \text{probability} \\ \text{measure}}}{P})$$

(a set of intervals & their intersections)

$f_{\underset{\sim}{x}}(x_1, \ldots, x_n)$


$P = 0$

has a prob. $= \int f_{\underset{\sim}{x}}(x_1 \ldots x_n)$

$$\text{Mean } E[\underset{\sim}{x}] = \begin{bmatrix} E[x_1] \\ \vdots \\ E[x_n] \end{bmatrix} = \int \cdots \int \underset{\sim}{x}\, f_{\underset{\sim}{x}}(\ldots)\, dx_1 \ldots dx_n$$

$\underset{\sim}{x} \to$ bin number
dependence structure
PSD

$$\text{Correlation Matrix: } E[\underset{\sim}{x}\underset{\sim}{x}^T] \geq 0 \quad (\text{when is it } \neq 0 ?)$$

$$\text{Covariance Matrix: } E[\underset{\sim}{y}\underset{\sim}{y}^T] = E[(\underset{\sim}{x} - E[\underset{\sim}{x}])(\underset{\sim}{x} - E[\underset{\sim}{x}])^T]$$

$$\underset{\sim}{y} = \underset{\sim}{x} - E[\underset{\sim}{x}] \qquad = E[\underset{\sim}{x}(\underset{\sim}{x} - E[\underset{\sim}{x}])^T] - E[E[\underset{\sim}{x}](\underset{\sim}{x} - E[\underset{\sim}{x}])^T]$$

$$= E[\underset{\sim}{x}\underset{\sim}{x}^T] - E[\underset{\sim}{x}](E[\underset{\sim}{x}])^T$$

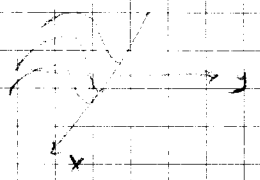
Complex Vectors

Correlation Matrix : $E[\underset{\sim}{x}\underset{\sim}{x}^*]$

Covariance Matrix : $E[\underset{\sim}{x}\underset{\sim}{x}^*] - E[\underset{\sim}{x}]E[\underset{\sim}{x}]^*$

PROBLEM OF STOCHASTIC ESTIMATION

→ Estimating random variable x (x=3) from random variable y (even/odd)



Knowledge of y reduces the ambiguity in x.

$E\left[(\hat{\underset{\sim}{x}} - \underset{\sim}{x})^2\right]$ → mean square error.

y: even $\quad 0 \cdot \frac{1}{3} + 2^2 \cdot \frac{1}{3} + 4^2 \cdot \frac{1}{3} = \frac{20}{3}$ → choosing 2 → P of error $\frac{2}{3}$

$2^2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 2^2 \cdot \frac{1}{3} = \frac{8}{3}$ → choosing 4 → $\frac{2}{3}$

If no y : P of error $\frac{5}{6}$ mean square estimate : 3.5 → P of error = 1.

STOCHASTIC LEAST SQUARES ESTIMATION

1. SCALAR CASE :

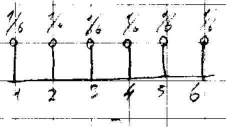
(f_x, f_y are not enough)

Given r.v.s x, y, we know f_{xy}. We observe y. We would like to estimate x using observation y.

$\hat{\underset{\sim}{x}} = h(y)$ (We can use different criteria such as
- Probability of error $P(\hat{\underset{\sim}{x}} \neq x)$ → For continuous variables, it generally becomes 1.
- MSE $E(|x-\hat{x}|)^2$

We choose $E(|x-\hat{x}|^2)$

- Suppose no y value is given



$\hat{x} = a$ $\qquad E[(x-1)^2] = 0 \cdot \frac{1}{6} + 1 \cdot \frac{1}{6} \cdots$

a that minimizes $E(|x-a|^2)$ is $a = E(\underset{\sim}{x})$

$$E((x-a)^2) = E\left((x-E(x)+E(x)-a)^2\right) = E\left((x-E(x))^2\right) + 2E\left(\underbrace{(x-E(x))(E(x)-a)}_{0}\right) + (E(x)-a)^2$$

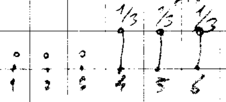
$$= E\left((x-E(x))^2\right) + (E(x)-a)^2$$

$$= \sigma_x^2 + (E(x)-a)^2 \quad \left(\text{minimized at } a = E(x)\right)$$

Var – uncertainty
[illegible] skewness?

- Suppose you know the value of y

Die experiment : $y = \begin{cases} 0 & x \leq 3 \\ 1 & x > 3 \end{cases}$ $\qquad y = 1 \quad P_{x|y}(x \mid y=1) = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$



Calculate new pmf of x
pdf

$$P_{x|y}(x \mid y) = \frac{P_{xy}(x,y)}{P_y(y)}$$

$$f_{x|y}(x \mid y) = \frac{f_{xy}(x,y)}{f_y(y)}$$

The best MSE estimate = mean calculated based on new pdf (conditional) pmf

$E(x \mid y)$ ← in general a nonlinear function of y
- you need to know $f_{x|y}$ in general

Constrain estimator to be linear function of y.

$\hat{x} = ay + b$ $\qquad E(|x-\hat{x}|^2)$

(Best $\underline{\text{linear}}$ MSE estimator)

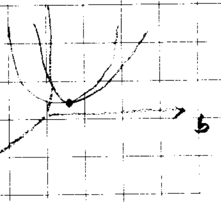
1. Approach

$$E(|x-(ay+b)|^2) = f(a,b)$$

Convex function (Why?)

$$\frac{\partial f}{\partial a} = 0$$

$$\frac{\partial f}{\partial b} = 0$$



2. Approach

Assume first x,y zero mean

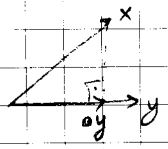
$\hat{x} = ay$ $\qquad E(\hat{x}) = E(ay+b) = aE(\overset{0}{y}) + b \to 0$

Define a Hilbert (Inner product) Space zero mean random variables with inner product

$$\langle a,b\rangle = E(ab^*) \xrightarrow{\text{Induced Norm}} \|a\| = \langle a,a\rangle^{1/2}$$

$$= (E(aa^*))^{1/2} = \sqrt{E(|a|^2)}$$

$$E(|x-ay|^2) = \|x-ay\|^2$$



P.T. $(x-ay) \perp y$ $\qquad$ Estimation error should be orthogonal to observations

$\|$

Estimation error and y should be uncorrelated.

$$\langle \underbrace{x-ay}_{e}, y\rangle = 0$$

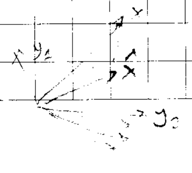
$$E((\underbrace{x-ay}_{e})y^*) = 0$$

$$E(xy^*) - aE(yy^*) = 0$$

$$a = \frac{E(xy^*)}{E(yy^*)} = \frac{\langle x,y\rangle}{\langle y,y\rangle}$$

$$\hat{x} = \frac{\langle x,y\rangle}{\langle y,y\rangle} y$$

If multiple observations, then this is equivalent to projection into the span of Y.



If x, y were not zero mean

$$E(x)=\mu_x \quad E(y)=\mu_y$$

Define $x' = x-\mu_x \quad E(x')=0$

$$y' = y-\mu_y \quad E(y')=0$$

$$\hat{x}' = \frac{\langle x', y'\rangle}{\langle y', y'\rangle} y$$

$$\hat{x} - E(x) = \hat{x}' = \frac{E((x-\mu_x)(y-\mu_y)^*)}{E((y-\mu_y)(y-\mu_y)^*)}(y-\mu_y)$$

$$\hat{x} = \mu_x + \frac{\langle (x-E(x)), (y-E(y))\rangle}{\langle (y-\mu_y)(y-\mu_y)\rangle}(y-\mu_y)$$

(without loss of generality)

For the rest of our classes we'll assume zero mean r.v.s WLOG.

2. VECTOR CASE:

Suppose $\underset{\sim}{x} \in C^m$ ← Desired

$\underset{\sim}{y_i} \in C^p$ $i=1,\dots,N$ ← Observations

What is the best linear minimum MSE estimate of $\underset{\sim}{x}$ from y_i's ?

$$\underset{\sim}{\hat{x}} = \overset{m\times p}{K_1}\underset{\sim}{y_1} + K_2\underset{\sim}{y_2} + \dots + K_N y_N$$

$$= [K_1\ K_2 \ \dots\ K_N]\begin{bmatrix} \underset{\sim}{y_1} \\ \vdots \\ \underset{\sim}{y_N} \end{bmatrix} = \overset{m\times pN}{\underset{\sim}{K}}\ \overset{pN\times 1}{\underset{\sim}{y}}$$

$$\underset{\sim}{\hat{x}} = K\underset{\sim}{y}$$

Estimation error : $\underset{\sim}{e} = \underset{\sim}{x} - Ky \in C^m$ → PSD matrix → You can compare covariance matrices.

Criterion : minimize $E(\underset{\sim}{e}\underset{\sim}{e}^*) = P(K)$

⇒ Find K_0 such that

$$P(K) \geq P(K_0) \text{ for all } K.$$

Approach : Geometric approach

$e_k \perp y_l \qquad k=1,\dots,m$

$\qquad\qquad\quad l=1,\dots,Np$

$$E(e_k y_\ell^*) = 0$$

$$E\left(\left| \overline{} \right. \right)$$

$$E(\underset{\sim}{e}\, \underset{\sim}{y}^*) = 0 \Leftrightarrow \langle \underset{\sim}{e}, \underset{\sim}{y} \rangle = 0 \qquad E\left(\begin{bmatrix} e_1 \\ \vdots \\ e_r \end{bmatrix} \begin{bmatrix} y_1^* & \cdots & y_{Np}^* \end{bmatrix} \right) = \begin{bmatrix} E(e_1 y_1^*) & \cdots \\ \vdots & \end{bmatrix}$$

Given vectors $\underset{\sim}{a}, \underset{\sim}{b}$ we can define a generalized inner product

$$\langle \underset{\sim}{a}, \underset{\sim}{b} \rangle = E(\underset{\sim}{a}\, \underset{\sim}{b}^*)$$

$$E(\underset{\sim}{e}\, \underset{\sim}{y}^*) = 0$$

$$E((\underset{\sim}{x} - K_0 \underset{\sim}{y})\, \underset{\sim}{y}^*) = 0$$

$$\boxed{E(\underset{\sim}{x}\underset{\sim}{y}^*) = K_0 E(\underset{\sim}{y}\underset{\sim}{y}^*)} \leftarrow \text{May have infinitely many solutions}$$

If $R_{\underset{\sim}{y}}$ is singular

$$\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle = K_0 \langle \underset{\sim}{y}, \underset{\sim}{y} \rangle$$

$$\boxed{R_{\underset{\sim}{x}\underset{\sim}{y}} = K_0 R_{\underset{\sim}{y}}}$$

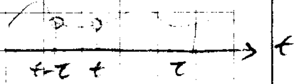
No longer a scalar

If R_y is nonsingular $\boxed{K_0 = R_{\underset{\sim}{x}\underset{\sim}{y}} R_{\underset{\sim}{y}}^{-1}} = \langle \underset{\sim}{x}, \underset{\sim}{y} \rangle \langle \underset{\sim}{y}, \underset{\sim}{y} \rangle^{-1}$

Examples

1. Consider zero-mean real valued $y(t)$ with autocovariance function

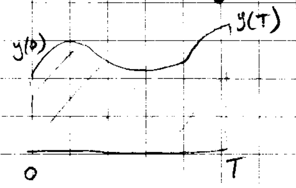
$$R_y(\tau) = \langle y(t), y(t-\tau) \rangle$$



Second-order stationarity

Find the best linear MMSE estimator ~~of~~ its integral

$$z = \int_0^T y(t)\, dt$$ from its end points $y(0), y(T)$



Solution: Find estimate of z from $y(0), y(T)$

$$\hat{z} = a\, y(0) + b\, y(T)$$

$$= \underbrace{\begin{bmatrix} a & b \end{bmatrix}}_{K} \underbrace{\begin{bmatrix} y(0) \\ y(T) \end{bmatrix}}_{\underset{\sim}{y}} \qquad K = \langle z, y \rangle \langle y, y \rangle^{-1} \leftarrow \text{Suggests that}$$

$\mathrm{Tr}\, E(e\, e^*) =$ [illegible]

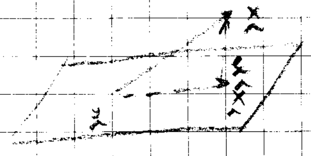
LINEAR MMSE ESTIMATION THEORY

- Estimating a zero-mean r.v. x from y

$$\hat{x} = \frac{\langle x, y\rangle}{\langle y, y\rangle}\, y \qquad \langle x, y\rangle = E(x y^*)$$

- Estimate of a random vector $\underset{\sim}{x}$ from random vector $\underset{\sim}{y}$

$$\hat{\underset{\sim}{x}} = \underbrace{\langle \underset{\sim}{x}, \underset{\sim}{y}\rangle \langle \underset{\sim}{y}, \underset{\sim}{y}\rangle^{-1}}_{K}\, \underset{\sim}{y}$$



$$\underset{\sim}{e} \perp \underset{\sim}{y}$$

$$\langle \underset{\sim}{e}, \underset{\sim}{y}\rangle \underset{\sim}{0}$$

$$\underset{\sim}{e} = \underset{\sim}{x} - K\underset{\sim}{y}$$

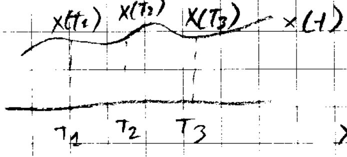
$$\langle \underset{\sim}{x} - K\underset{\sim}{y}, \underset{\sim}{y}\rangle = 0$$

$$\langle \underset{\sim}{x}, \underset{\sim}{y}\rangle = K\langle \underset{\sim}{y}, \underset{\sim}{y}\rangle$$

$$K = \langle \underset{\sim}{x}, \underset{\sim}{y}\rangle \langle \underset{\sim}{y}, \underset{\sim}{y}\rangle^{-1}$$

Example 2: Consider a zero mean scalar process $\{x(t)\}$ with

$$\langle x(t), x(t-\tau)\rangle = R_x(\tau) = e^{-\alpha|\tau|}$$



Find the best linear MMSE estimate of $x(T_3)$ from $x(T_1)$ and $x(T_2)$. $(T_1 < T_2 < T_3)$

Desired: $z = X(T_3)$

Observations: $\underset{\sim}{y} = \begin{bmatrix} X(T_1) \\ X(T_2) \end{bmatrix}$

$$\hat{z} = \langle z, \underset{\sim}{y}\rangle \langle \underset{\sim}{y}, \underset{\sim}{y}\rangle^{-1} \underset{\sim}{y}$$

$$\langle z, \underset{\sim}{y}\rangle = \left\langle x(T_3), \begin{bmatrix} x(T_1) \\ x(T_2) \end{bmatrix} \right\rangle = E\left(x(T_3)\,[x^*(T_1) \quad x^*(T_2)]\right)$$

$$= \left[E(x(T_3)x^*(T_1)) \;\; E(x(T_3)x^*(T_2))\right]$$

$$= e^{-\alpha(T_3-T_1)} . e^{-\alpha(T_3-T_2)}$$

$$\langle \underset{\sim}{y}, \underset{\sim}{y}\rangle = E(\underset{\sim}{y}\,\underset{\sim}{y}^*) = E\left(\begin{bmatrix} x(T_1) \\ x(T_2) \end{bmatrix} [x^*(T_1) \quad x^*(T_2)]\right)$$

$$= \begin{bmatrix} E(x(T_1)x^*(T_1)) & E(x(T_1)x^*(T_2)) \\ E(x(T_2)x^*(T_1)) & E(x(T_2)x^*(T_2)) \end{bmatrix}$$

$$\langle \underset{\sim}{y}, \underset{\sim}{y} \rangle = \begin{bmatrix} 1 & e^{-\alpha(T_2-T_1)} \\ e^{-\alpha(T_2-T_1)} & 1 \end{bmatrix}$$

$$\langle \underset{\sim}{y}, \underset{\sim}{y} \rangle^{-1} = \begin{bmatrix} 1 & -e^{-\alpha(T_2-T_1)} \\ -e^{-\alpha(T_2-T_1)} & 1 \end{bmatrix} \frac{1}{1-e^{-2\alpha(T_2-T_1)}}$$

$$\langle z, \underset{\sim}{y} \rangle \langle \underset{\sim}{y}, \underset{\sim}{y} \rangle^{-1} = \begin{bmatrix} e^{-\alpha(T_3-T_1)} & e^{-\alpha(T_3-T_2)} \end{bmatrix} \begin{bmatrix} 1 & -e^{-\alpha(T_2-T_1)} \\ -e^{-\alpha(T_2-T_1)} & 1 \end{bmatrix} \frac{1}{1-e^{-2\alpha(T_2-T_1)}}$$

$$= \Big[\overset{0}{\big(e^{-\alpha(T_3-T_1)} - e^{-\alpha(T_3-T_1)}\big)} \quad \frac{-e^{-\alpha(T_3+T_2-2T_1)} + e^{-\alpha(T_3-T_2)}}{1-e^{-2\alpha(T_2-T_1)}} \Big]$$

$$\frac{e^{-\alpha(T_3-T_2)}\left(1-e^{-\alpha(2T_2-2T_1)}\right)}{\left(1-e^{-2\alpha(T_2-T_1)}\right)}$$

$$\langle z, \underset{\sim}{y} \rangle \langle \underset{\sim}{y}, \underset{\sim}{y} \rangle^{-1} = \begin{bmatrix} 0 & e^{-\alpha(T_3-T_2)} \end{bmatrix}$$

$$\hat{x}(T_3) = \hat{z} = \begin{bmatrix} 0 & e^{-\alpha(T_3-T_2)} \end{bmatrix} \begin{bmatrix} x(T_1) \\ x(T_2) \end{bmatrix}$$

$$= e^{-\alpha(T_3-T_2)} x(T_2)$$

→ Do not use $x(T_1)$. This is a special case for the exponential r.v. we have used
↳ memoryless property of the autocorrelation function.

LINEAR MODELS

$\underset{\sim}{y} \rightarrow$ Observation vector $(p \times 1)$

$\underset{\sim}{x} \rightarrow$ Desired vector $(N \times 1)$

In many applications

$$\underset{\sim}{y} = H\underset{\sim}{x} + \underset{\sim}{v} \qquad \underset{\sim}{x}, \underset{\sim}{v} \text{ zero mean} \qquad \left.\begin{matrix} H \\ R_{\underset{\sim}{x}} = \langle \underset{\sim}{x}, \underset{\sim}{x} \rangle \\ R_{\underset{\sim}{v}} = \langle \underset{\sim}{v}, \underset{\sim}{v} \rangle \end{matrix}\right\} \text{are known.}$$

observation — desired — noise

$$\langle \underset{\sim}{x}, \underset{\sim}{v} \rangle = 0$$

$$H \in \mathbb{C}^{p \times N}$$

Goal: $\hat{\underset{\sim}{x}} = K\underset{\sim}{y}$

$$\hat{\underset{\sim}{x}} = \langle \underset{\sim}{x}, \underset{\sim}{y} \rangle \langle \underset{\sim}{y}, \underset{\sim}{y} \rangle^{-1} \qquad \langle H\underset{\sim}{x}, \underset{\sim}{x} \rangle = H\langle \underset{\sim}{x}, \underset{\sim}{x} \rangle$$

$$\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle = \langle \underset{\sim}{x}, H\underset{\sim}{x} + \underset{\sim}{v} \rangle = \langle \underset{\sim}{x}, H\underset{\sim}{x} \rangle + \underbrace{\langle \underset{\sim}{x}, \underset{\sim}{v} \rangle}_{0}$$

(Linearity of the inner product)

$$= E(x x^*) H^*$$

$$\langle \underset{\sim}{x}, \underset{\sim}{x} \rangle H^*$$

$$= R_x H^*$$

$\overset{0}{H\langle \underset{\sim}{x}, \underset{\sim}{v}\rangle}$ $\quad 0$

$$\langle \underset{\sim}{y}, \underset{\sim}{y}\rangle = \langle H\underset{\sim}{x}+\underset{\sim}{v}, H\underset{\sim}{x}+\underset{\sim}{v}\rangle = \langle H\underset{\sim}{x}, H\underset{\sim}{x}\rangle + \langle H\underset{\sim}{x}, \underset{\sim}{v}\rangle + \langle \underset{\sim}{v}, H\underset{\sim}{x}\rangle + \langle \underset{\sim}{v}, \underset{\sim}{v}\rangle$$

$$= H\langle \underset{\sim}{x}, \underset{\sim}{x}\rangle H^* + \langle \underset{\sim}{v}, \underset{\sim}{v}\rangle$$

$$= H R_{\underset{\sim}{x}} H^* + R_{\underset{\sim}{v}}$$

$$\hat{\underset{\sim}{x}} = R_{\underset{\sim}{x}} H^* (\overset{B}{H} \overset{C}{R_{\underset{\sim}{x}}} \overset{D}{H^*} + \overset{A}{R_{\underset{\sim}{v}}})^{-1} \underset{\sim}{y}$$

N×N p×p

If we assume $R_{\underset{\sim}{x}}, R_{\underset{\sim}{v}}$ invertible

Apply MIL $\hat{\underset{\sim}{x}} = R_{\underset{\sim}{x}} H^* (R_{\underset{\sim}{v}}^{-1} - R_{v}^{-1} H (R_{\underset{\sim}{x}}^{-1} + H^* R_{\underset{\sim}{v}}^{-1} H)^{-1} H^* R_v^{-1}) \underset{\sim}{y}$

$$= R_{\underset{\sim}{x}} (I - \underbrace{H^* R_{\underset{\sim}{v}}^{-1} H}_{(-R_{\underset{\sim}{x}}^{-1} + (R_{\underset{\sim}{x}}^{-1} + H^* R_{\underset{\sim}{v}}^{-1} H))} (R_{\underset{\sim}{x}}^{-1} + H^* R_{\underset{\sim}{v}}^{-1} H)^{-1}) H^* R_{\underset{\sim}{v}}^{-1} \underset{\sim}{y}$$

$$= R_{\underset{\sim}{x}} (R_{\underset{\sim}{x}}^{-1} (R_{\underset{\sim}{x}}^{-1} + H^* R_{\underset{\sim}{v}}^{-1} H)^{-1}) H^* R_{\underset{\sim}{v}}^{-1} \underset{\sim}{y}$$

$$\rightarrow \boxed{\begin{aligned} \hat{\underset{\sim}{x}} &= (R_{\underset{\sim}{x}}^{-1} + H^* R_{\underset{\sim}{v}} H)^{-1} H^* R_{\underset{\sim}{v}}^{-1} \underset{\sim}{y} \\ \hat{\underset{\sim}{x}} &= R_{\underset{\sim}{x}} H^* \underbrace{(R_{\underset{\sim}{v}} + H R_{\underset{\sim}{x}} H^*)^{-1}}_{p\times p} \underset{\sim}{y} \end{aligned}}$$

$R_{\underset{\sim}{v}}^{-1} = \sigma_v^{-2} I$ (in general)

We can also choose $R_{\underset{\sim}{x}}$ to be white.

$p \rightarrow 10.000$

$N \rightarrow 10$

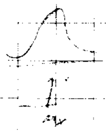
Resemblance with Regularized Deterministic Least Squares

Regularized Least Squares	Stochastic Least Squares with Non-zero Means
$(\underset{\sim}{x} - \underset{\sim}{x}_0)^* \Pi_0^{-1} (\underset{\sim}{x} - \underset{\sim}{x}_0) + \lVert \underset{\sim}{y} - H\underset{\sim}{x} - \underset{\sim}{v}_0 \rVert_W^2$	$\hat{\underset{\sim}{x}} = \mu_{\underset{\sim}{x}} + (R_{\underset{\sim}{x}}^{-1} + H^* R_{\underset{\sim}{v}}^{-1} H)^{-1} H^* R_{\underset{\sim}{v}}^{-1} (\underset{\sim}{y} - H\mu_{\underset{\sim}{x}} - \mu_{\underset{\sim}{v}})$
$\hat{\underset{\sim}{x}} = \underset{\sim}{x}_0 + (\Pi_0^{-1} + H^* W H)^{-1} H^* W (\underset{\sim}{y} - H\underset{\sim}{x}_0 - \underset{\sim}{v}_0)$	

$\underset{\sim}{v}_0 \rightarrow \mu_{\underset{\sim}{v}}$

$\underset{\sim}{x}_0 \rightarrow \mu_{\underset{\sim}{x}}$

$\Pi_0 \rightarrow R_{\underset{\sim}{x}}$ → when (ambiguity) covariance, $R_{\underset{\sim}{x}}$, is small, Π_0^{-1} is large.



$W \rightarrow R_{\underset{\sim}{v}}^{-1}$

$$R_v = \begin{bmatrix} \sigma_{v_1}^2 & \\ & \sigma_{v_2}^2 \end{bmatrix}$$

$$W = \begin{bmatrix} \sigma_{v_1}^{-2} & 0 \\ 0 & \sigma_{v_2}^{-2} \end{bmatrix}$$

$\lVert y_1 + H_{1,1} \underset{\sim}{x} \rVert^2 \sigma_{v_1}^{-2} + \lVert y_2 + H_{2,2} \underset{\sim}{x} \rVert^2 \underline{\sigma_{v_2}^{-2}}$

$\sigma_{v_1} > \sigma_{v_2} \Rightarrow \sigma_{v_1}^{-2} < \sigma_{v_2}^{-2}$ Weighted more

Gives more weight to the observations with smaller noise.

If noise is very large, R will diminish.

Signal to noise ratio is important. $(\overset{\text{power of the signal}}{H R_{x} H^*}$ vs. $R_v)$ → Is there a condition number of an estimation problem?

$\underset{\sim}{x}$ from $\underset{\sim}{y}$

$$\hat{\underset{\sim}{x}} = K \underset{\sim}{y}$$

$$x - \hat{x} \perp \underset{\sim}{y} \Rightarrow E((\underset{\sim}{x} - \hat{\underset{\sim}{x}}) \underset{\sim}{y}^*) = 0$$

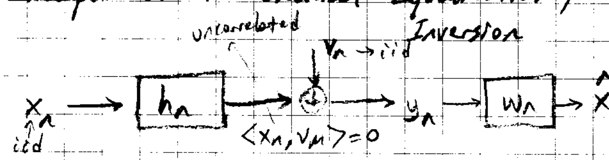
$$E((\underset{\sim}{x} - K\underset{\sim}{y}) \underset{\sim}{y}^*) = 0$$

$$R_{xy} = K R_{y}$$

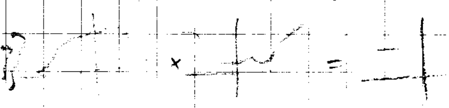
$$R_{\underset{\sim}{y}} > 0 \Rightarrow K = R_{xy} R_{y}^{-1}$$

Example (FIR Channel Equalization) Inversion

$$h_n * \overset{\text{Inverse of } h_n}{w_n} = \delta_n$$



$$H(e^{jw}) \, W(e^{jw}) = 1 \quad \forall w$$



Equal at all frequencies

Determine w_n : FIR filter $\{w_n : n \in \{0, \ldots, M-1\}\}$

such that $E((x_{n-d} - \hat{x}_{n-d})^2)$ is minimized.

$$\hat{x}_{n-d} = w_0 y_n + w_1 y_{n-1} + \ldots + w_{M-1} y_{n-M+1}$$

$$= [w_0 \; \ldots \; w_{M-1}] \begin{bmatrix} y_n \\ \vdots \\ y_{n-M+1} \end{bmatrix}$$

$$\hat{x}_{n-d} = W \underset{\sim}{y_n}$$

$$W = R_{x_{n-d} \underset{\sim}{y_n}} R_{\underset{\sim}{y_n}}^{-1}$$

Assuming h_n is FIR also with N taps

$$\underset{\sim}{y_n} = \begin{bmatrix} y_n \\ \vdots \\ y_{n-M+1} \end{bmatrix}$$

$$y_n = h_0 x_n + h_1 x_{n-1} + \ldots + h_{N-1} x_{n-N+1} + v_n$$

$$y_{n-1} = h_0 x_{n-1} + h_1 x_{n-2} + \ldots + h_{N-1} x_{n-N} + v_{n-1}$$

$$\underset{\sim}{y_n} = \underbrace{\begin{bmatrix} h_0 & h_1 & \cdots & h_{N-1} & 0 & & 0 \\ 0 & h_0 & h_1 & & h_{N-1} & 0 & 0 \\ & & & & & & \\ & & & h_0 & \cdots & & h_{N-1} \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_{n-N+1} \\ \vdots \\ x_{n-N-M+2} \end{bmatrix}}_{\underset{\sim}{x_n}} + \underbrace{\begin{bmatrix} v_n \\ \vdots \\ \vdots \\ v_{n-M+1} \end{bmatrix}}_{\underset{\sim}{v_n}} \qquad 0 \leq d \leq N-M+2$$

Toeplitz Matrix

$$\underset{\sim}{y_n} = H \underset{\sim}{x_n} + \underset{\sim}{v_n}$$

$$R_{x_{n-d} \underset{\sim}{y_n}} = E(x_{n-d} \underset{\sim}{y_n}^*) = E(x_{n-d}(\underset{\sim}{x_n}^* H^* + \underset{\sim}{v_n}^*))$$

$$= E(x_{n-d} \underset{\sim}{x_n}^*) H^* + \overset{0}{E(x_{n-d} \underset{\sim}{v_n}^*)}$$

$$= E\left(x_{n-d}\left[x_n^* \; x_{n-1}^* \; \dots \; x_{n-M-N+2}^*\right]\right) H^*$$

$$= \left[0 \quad 0 \quad \dots \quad \underset{\substack{\uparrow \\ d+1}}{\sigma_x^2} \quad \dots \quad 0\right] H^*$$

$$R_{\underset{\sim}{x}_{n-d}\underset{\sim}{y}_n} = \sigma_x^2 \, \underset{\sim}{e}_{d+1}^T H^* \quad \leftarrow \text{standard basis vector}$$

$$R_{\underset{\sim}{y}_n} = E(\underset{\sim}{y}_n \underset{\sim}{y}_n^*) = E\left((H\underset{\sim}{x}_n + \underset{\sim}{v}_n)(H\underset{\sim}{x}_n + \underset{\sim}{v}_n)^*\right)$$

$$= E(H\underset{\sim}{x}_n \underset{\sim}{x}_n^* H^*) + \overset{0}{\cancel{E(H\underset{\sim}{x}_n \underset{\sim}{v}_n^*)}} + \overset{0}{\cancel{E(\underset{\sim}{v}_n \underset{\sim}{x}_n H^*)}} + E(\underset{\sim}{v}_n \underset{\sim}{v}_n^*)$$

$$= H \overset{\sigma^2 I}{R_{\underset{\sim}{x}_n}} H^* + R_{\underset{\sim}{v}_n}$$

$$= \sigma_x^2 H H^* + \sigma_v^2 I$$

$$W = R_{x_{n-d}\underset{\sim}{y}_n} R_{\underset{\sim}{y}_n}^{-1} = \sigma_x^2 \, \underset{\sim}{e}_{d+1}^T H^* (\sigma_x^2 HH^* + \sigma_v^2 I)^{-1}$$

$$= \underset{\sim}{e}_{d+1}^T H^* \left(HH^* + \frac{\sigma_v^2}{\sigma_x^2} I\right)^{-1}$$

ESTIMATION OF RANDOM PROCESSES

$\{s_i\}$ → not observable, desired signal to be estimated

$\{y_i\}$ → observation signal

Known statistics $R_{sy}(i,l) = \langle s_i, y_l \rangle$

$$R_y(i,l) = \langle y_i, y_l \rangle$$

We'll first look at finite window problem: i.e., estimating $\{s_1, \dots, s_N\}$ from $\{y_1, \dots, y_N\}$. (This is still vector estimation problem)

Three types of problems:

a) Smoothing: to estimate s_i, use $\{y_1, \dots, y_N\}$

$$\begin{bmatrix} \hat{s}_1 \\ \vdots \\ \hat{s}_N \end{bmatrix} = \begin{bmatrix} k_{11}^{(s)} & \dots & k_{1N}^{(s)} \\ \vdots & & \\ k_{N1}^{(s)} & \dots & k_{NN}^{(s)} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

Wiener

$\hat{\underset{\sim}{s}} = K^{(s)} \underset{\sim}{y}$ $\quad K^{(s)}$ → full matrix $\quad K^{(s)} = R_{\hat{s}\underset{\sim}{y}} R_{\underset{\sim}{y}}^{-1}$

b) Filtering Problem: to estimate s_i, use $\{y_1, \dots, y_i\}$ → Causal

$$\begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \vdots \\ \hat{s}_N \end{bmatrix} = \begin{bmatrix} K_{11}^{(f)} & 0 & \dots & 0 \\ K_{21}^{(f)} & K_{22}^{(f)} & 0 \dots & 0 \\ \vdots & & & \\ K_{N1}^{(f)} & \dots & & K_{NN}^{(f)} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

$\hat{\underset{\sim}{s}} = K^{(f)} \underset{\sim}{y}$ $\quad$ Hopf

$K^{(f)}$ → Lower triangular
⇒ Causal

minimize $E\left((\underset{\sim}{s}-K^{(f)}\underset{\sim}{y})(\underset{\sim}{s}-K^{(f)}\underset{\sim}{y})^*\right)$ s.t. $K^{(f)} \to$ Lower Triangular.

$$s_i - \hat{s}_i \perp y_1, \dots, y_i$$

$$\left\{E\left((\underset{\sim}{s}-\hat{\underset{\sim}{s}})\,\underset{\sim}{y}^*\right)\right\}_{\text{lower}} = 0 \qquad \left\{\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}\right\}_{\text{lower}} = \begin{bmatrix} 1 & 0 & 0 \\ 4 & 5 & 0 \\ 7 & 8 & 3 \end{bmatrix}$$

$$\left\{E\left((\underset{\sim}{s}-K^{(f)}\underset{\sim}{y})\,\underset{\sim}{y}^*\right)\right\}_{\text{lower}} = 0 \qquad \underset{\sim}{e} = \underset{\sim}{s} - \hat{\underset{\sim}{s}} \quad E(\underset{\sim}{e}\,\underset{\sim}{y}^*)$$

$$= \begin{bmatrix} E(e_1 y_1^*) & E(e_1 y_2^*) & \dots & E(e_1 y_N^*) \\ E(e_2 y_1^*) & E(e_2 y_2^*) & & \\ & & & \end{bmatrix}$$

$$\left\{E\left(\underset{\sim}{s}\,\underset{\sim}{y}^* - K^{(f)}\underset{\sim}{y}\,\underset{\sim}{y}^*\right)\right\}_{\text{lower}} = 0$$

$$\left\{R_{\underset{\sim}{s}\underset{\sim}{y}} - K^{(f)}R_{\underset{\sim}{y}}\right\}_{\text{lower}} = 0 \qquad K^{(f)} \to \text{Lower Triangular} \qquad \downarrow = 0$$

$$R_{\underset{\sim}{s}\underset{\sim}{y}} - K^{(f)}R_{\underset{\sim}{y}} = U \to \text{strictly upper triangular}$$

$R_{\underset{\sim}{y}} = L R_e L^*$ $\quad L \to$ Lower Triangular (Diagonal elements $= 1$)

$R_e \to$ Diagonal $\qquad$ Cholesky Factorization

$$R_{sy} - K^{(f)} L R_e L^* = U$$

$$\underbrace{R_{sy} L^{-*} R_e^{-1}}_{\text{Full in general}} - \underbrace{K^{(f)} L}_{\substack{\text{lower}\\ \text{triangular}}} = \underbrace{U L^{-*} R_e^{-1}}_{\text{strictly upper triangular}}$$

$$K^{(f)} L = \left\{R_{sy} L^{-*} R_e^{-1}\right\}_{\text{lower}}$$

$$K^{(f)} = \left\{R_{sy} L^{-*} R_e^{-1}\right\}_{\text{lower}} L^{-1} \quad \text{vs.} \quad \left\{R_{sy} \underbrace{L^{-*} R_e^{-1} L^{-1}}_{R_y^{-1}}\right\}_{\text{lower}}$$

ad-hoc solution

$\underset{\sim}{s}$ from $\underset{\sim}{y}$ $\qquad \hat{\underset{\sim}{s}} = K\underset{\sim}{y}$

No constraints on K: $\quad K = R_{\underset{\sim}{s}\underset{\sim}{y}} R_{\underset{\sim}{y}}^{-1}$

$K \to$ Lower triangular: $\quad K = \left\{R_{sy} L^{-*} R_{\underset{\sim}{e}}^{-1}\right\}_{\text{lower}} L^{-1} \qquad R_y = L R_e L^*$

Example: Signals in additive white noise

$$\underset{\sim}{y} = \underset{\sim}{s} + \underset{\sim}{v} \qquad \left\langle \begin{bmatrix} \underset{\sim}{s} \\ \underset{\sim}{v} \end{bmatrix}, \begin{bmatrix} \underset{\sim}{s} \\ \underset{\sim}{v} \\ 1 \end{bmatrix} \right\rangle = \begin{bmatrix} R_s & 0 & 0 \\ 0 & R_v & 0 \end{bmatrix}$$

$\to E(\underset{\sim}{s}\,1) = E(\underset{\sim}{s}) = \underset{\sim}{0}$ $\Rightarrow$ means are zero.

Causal llms estimator $K^{(f)} \to$ Lower triangular

$$K^{(f)} = \left\{R_{\underset{\sim}{s}\underset{\sim}{y}} L^{-*} R_e^{-1}\right\}_{\text{lower}} L^{-1}$$

$$R_{\underset{\sim}{s}\underset{\sim}{y}} = E(\underset{\sim}{s}\,\underset{\sim}{y}) = E\left(\underset{\sim}{s}(\underset{\sim}{s}+\underset{\sim}{v})^*\right) = E(\underset{\sim}{s}\,\underset{\sim}{s}^*) + E(\underset{\sim}{s}\,\underset{\sim}{v}^*)^{\overset{\wedge}{\,}0} = R_{\underset{\sim}{s}}$$

$$R_{\underset{\sim}{y}} = E(\underset{\sim}{y}\,\underset{\sim}{y}^*) = E\left((\underset{\sim}{s}+\underset{\sim}{v})(\underset{\sim}{s}+\underset{\sim}{v})^*\right) = R_{\underset{\sim}{s}} + R_{\underset{\sim}{v}}$$

$$R_{\underset{\sim}{s}} = R_{\underset{\sim}{y}} - R_{\underset{\sim}{v}}$$

$$K^{(f)} = \{ \underbrace{(R_{\underset{\sim}{y}} - R_{\underset{\sim}{v}})}_{R_{\underset{\sim}{s}\underset{\sim}{y}} = R_{\underset{\sim}{s}}} L^{-*} Re^{-1} \}_{lower} L^{-1}$$

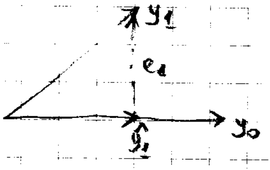
$$= \{ (\underbrace{L Re L^* L^{-*} Re^{-1}}_{I} - R_{\underset{\sim}{v}} L^{-*} Re^{-1} \}_{lower} L^{-1}$$

$$\boxed{K^{(f)} = I - \{ R_{\underset{\sim}{v}} L^{-*} Re^{-1} \}_{lower} L^{-1}}$$

INNOVATIONS (NEW INFORMATION) PROCESS

Observations $y_0, y_1, \dots, y_N$ → Their inner product is the correlation. I can us

either y_0 or y_1 if they are fully correlated.



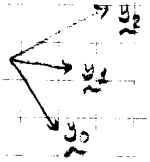
Linear MMSE

$\hat{y}_1 = \langle y_1, y_0 \rangle \langle y_0, y_0 \rangle^{-1} y_0$ → Prediction of y_1 from y_0.

$e_1 = y_1 - \hat{y}_1$ ← Prediction error

If $e_1 = 0$, then y_1 does not have any new information than y_0.

e_1 is the new information embedded inside $\underset{\sim}{y_1}$.



$$\text{Span}\{y_0, \dots, y_i\} = \text{Span}\{e_0, \dots, e_i\}$$

$e_0 = y_0$ $\quad$ $e_2 = y_2 - \hat{y}_2$

$\hat{y}_2$ = the prediction of y_2 from e_0, e_1

Projection of y_2 over $\text{span}\{e_0, e_1\} = \text{Span}\{y_0, y_1\}$

$$\hat{y}_2 = \sum_{k=0}^{1} \langle y_2, e_k \rangle \langle e_k, e_k \rangle^{-1} e_k$$

$$\{y_0, \dots, y_N\} \xrightarrow[\text{without normalization}]{\text{Gram-Schmidt}} \{e_0, \dots, e_N\}$$

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_N \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ l_{21} & 1 & & \\ \times & \ddots & \ddots & \\ \times & \times & \times & 1 \end{bmatrix}}_{L} \begin{bmatrix} e_0 \\ e_1 \\ \vdots \\ e_N \end{bmatrix}$$

$$e_1 = y_1 - \underset{l_{21} e_0}{\hat{y}_1}$$

$$y_1 = e_1 + l_{21} e_0$$

$\underset{\sim}{y} = L \underset{\sim}{e}$ → $\underset{\sim}{e}$ is related to $\underset{\sim}{y}$ by a causal mapping

→ $\underset{\sim}{e}$ is uncorrelated

$R_e = R_e^{1/2} I R_e^{*/2}$

[illegible] $R_e^{1/2}$

$$R_{\underset{\sim}{y}} = \langle \underset{\sim}{y}, \underset{\sim}{y} \rangle = \langle L\underset{\sim}{e}, L\underset{\sim}{e} \rangle = L \langle \underset{\wedge}{e}, \underset{\wedge}{e} \rangle L^* = L R_e L^*$$

↳ since we are not doing normalizat
R_e is a diagonal matrix.

Suppose you want to estimate s_i from $\{y_0, \dots, y_N\}$.

$$\hat{s}_i = \text{Projection of } s_i \text{ over } \underset{\displaystyle \text{Span}\{e_0, \dots, e_N\}}{\overset{\|}{\text{Span}\{y_0, \dots, y_N\}}}$$

$$\hat{s}_i = \sum_{k=0}^{N} \langle s_i, e_k \rangle \langle e_k, e_k \rangle^{-1} e_k$$ → since e_k are uncorrelated, we can just sum the projections.

$\underset{\wedge}{e} = L^{-1} \underset{\wedge}{y}$

↳ whitening since the Fourier Transform turns out to be constant at all frequencies

White light has all the frequencies.

$\underset{\wedge}{y} = L \underset{\wedge}{e}$ → white since uncorrelated (not necessarily because covariances are equal).

↳ coloring

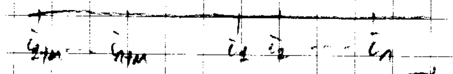
↳ colored

ESTIMATION WIDE SENSE STATIONARY RANDOM PROCESSES

Desired: $s_i \quad -\infty < i < \infty$
Observation: $y_i \quad -\infty < i < \infty$
} jointly WSS (zero mean)

Stationary process $\{x_i\} \Longleftrightarrow f_{x_{i_1} x_{i_2} \dots x_{i_n}} = f_{x_{i_{1+m}} x_{i_{2+m}} \dots x_{i_{n+m}}}$

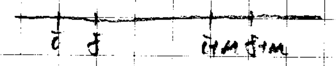
x_3 x_5



→ joint pdf is a function of the difference between the points

The process's behavior does not change.

WSS $\quad E(x_i x_j^*) = E(x_{i+m} x_{j+m}^*) = R_x(i-j) \quad \forall m$

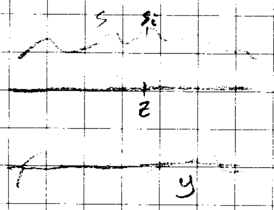


$E(x_i) = m_x \overset{\text{for us}}{=} 0$

Jointly WSS ⇒ s_i is WSS
⇒ y_i is WSS

⇒ $E(s_i y_j^*) = R_{sy}(i-j)$

A. Smoothing



$$\hat{s}_i = \sum_{k=-\infty}^{\infty} w_{ik} y_k$$

$s_i - \hat{s}_i \perp y_\ell \qquad -\infty < \ell < \infty$

$\langle s_i - \hat{s}_i, y_\ell \rangle = 0$

$$\langle s_i - \sum_{k=-\infty}^{\infty} w_{ik} y_k , y_\ell \rangle = 0 \qquad -\infty < \ell < \infty$$

$$\langle s_i , y_\ell \rangle - \sum_{k=-\infty}^{\infty} w_{ik} \langle y_k , y_\ell \rangle = 0$$

$$R_{sy}(i-\ell) - \sum_{k=-\infty}^{\infty} w_{ik} R_y(k-\ell) = 0 \qquad -\infty < \ell < \infty$$

$$k' = k-\ell \Rightarrow k = k'+\ell$$

$$i' = i-\ell \Rightarrow i = i'+\ell$$

$$R_{sy}(i') - \sum_{k'=-\infty}^{\infty} w_{i+\ell,k+\ell} R_y(k') = 0 \qquad -\infty < \ell < \infty$$

$$w_{i',k'} = w_{i'+1,k'+1} = w_{i'+\ell,k'+\ell} = V_{i'-k'}$$

$$R_{sy}(i') = \sum_{k'=-\infty}^{\infty} V_{i'-k'} R_y(k') \rightarrow \text{convolution.}$$

↗ If your estimator is time-invariant, WSS, then you obtain a convolution.

$$R_{sy} = V * R_y$$

$$S_{sy}(e^{jw}) = \mathcal{F}\{R_{sy}\}$$

$$S_y(e^{jw}) = \mathcal{F}\{R_y\}$$

$$S_{sy}(e^{jw}) = V(e^{jw}) \cdot S_y(e^{jw})$$

$$\boxed{V(e^{jw}) = \frac{S_{sy}(e^{jw})}{S_y(e^{jw})}} \qquad \frac{\langle s,y \rangle}{\langle y,y \rangle}$$

Wiener Filter

$$\mathcal{F}\{\hat{s}\} = \frac{S_{sy}(e^{jw})}{S_y(e^{jw})} Y(e^{jw})$$

→ cross-spectrum of observations

→ spectrum of observations

B. Filtering

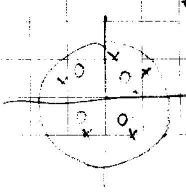
$$\hat{s}_i = \sum_{k=-\infty}^{i} V_{i-k}\, y_k = \sum_{k=0}^{\infty} V_k\, y_{i-k}$$

The solution is obtained through Wiener-Hopf procedure.

$$S_y(z) = L(z)\, r_e\, L^*(z^{-*})$$

↳ causal and causally invertible (minimum phase)

$L(z) \rightarrow$ poles, zeros } inside the unit circle



$$V(z) = \{ S_{sy}(z) L^{-*}(z^{-*}) r_e^{-1} \}_{\text{causal}} L^{-1}(z)$$

Derivative w.r.t. a matrix:

$A = E[(x-\hat{x})(x-\hat{x})^*]$ $\hat{x} = Ky + l$

1) $\frac{\partial a^* A a}{\partial a^* K}$ and $\frac{\partial a^* A a}{\partial a^* l}$ for any $a \in C^{n\times 1}$.

2) $\frac{\partial \, vec(A)}{\partial \, vec(K)}$

Question: Given that A is a square matrix whose diagonal elements are all 1s. Compute $det(e^A)$ given that A is diagonalizable.

$$A = TDT^{-1}$$

$$e^A = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \dots = I + TDT^{-1} + \frac{TD^2T^{-1}}{2!} + \frac{TD^3T^{-1}}{3!} + \dots$$

$$= T\left(I + D + \frac{D^2}{2!} + \frac{D^3}{3!} + \dots\right)T^{-1}$$

$$= T e^D T^{-1}$$

$$det(e^A) = det(\cancel{T}) . det(e^D) . det(\cancel{T^{-1}})$$

$$= det(e^D) \qquad e^D = \begin{bmatrix} e^{\lambda_1} & & 0 \\ & e^{\lambda_2} & \\ 0 & & e^{\lambda_n} \end{bmatrix}$$

$$= e^{\lambda_1} . e^{\lambda_2} . \dots . e^{\lambda_n}$$

$$= e^{\sum_{i=1}^{n} \lambda_i}$$

$$= e^{tr(D)}$$

Fixed-point iteration

$$\underset{\sim}{Ax} = \underset{\sim}{b} \Rightarrow I\underset{\sim}{x} + (A-I)\underset{\sim}{x} = \underset{\sim}{b}$$

$$I\underset{\sim}{x} = (I-A)\underset{\sim}{x} + \underset{\sim}{b}$$

(Richardson iteration) $x_{k+1} = (I-A)x_k + \underset{\sim}{b}$

For convergence we require $\|\underline{I-A}\| < 1$

> Iteration matrix

If $\|I-A\| \geq 1$, then $BA\underset{\sim}{x} = I\underset{\sim}{x} + (BA-I)\underset{\sim}{x} = B\underset{\sim}{b}$

$$x_{k+1} = (I-BA)x_k + Bb$$

If B is simple (i.e. diagonal) this is easy to compute. $I-BA$ is the iteration matrix

Conditional expectation is "better" than other orthogonal projections, because it preserves positivity ($f \geq 0 \Rightarrow E(f|Y) \geq 0$) and enjoys the comparison principle ($|f| \leq g \Rightarrow |E(f|Y)| \leq E(g|Y)$)