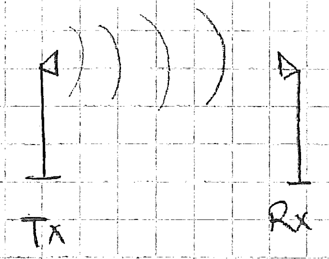
Gülin Tüfekci

L.1-2

EE503 Statistical Signal Processing and Modeling

12.10.2020

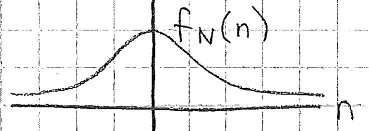


$r = s + n$

r: received signal under noise (observation)

s: symbol $s = \{\pm 1\}$ (binary symbols)

n: noise $N(0,1)$ (assumption)

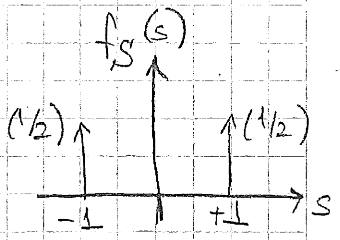


$$\frac{1}{\sqrt{2\pi\sigma_n^2}} \cdot e^{-\frac{(n-\mu_n)^2}{2\sigma_n^2}}$$

We are interested in s!

$\downarrow$

estimate s



- * How do we model noise?
- * What is the density (power spectral density) of the noise?

$$\left.\begin{aligned} r_1 &= s + n_1 \\ r_2 &= s + n_2 \\ r_3 &= s + n_3 \end{aligned}\right\} \quad \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} s + \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix}$$

$$\underline{r} = \underline{1}\, s + \underline{n}$$

Approach Define a loss function $\rightarrow \ell(s, \hat{s})$

① Square Error Loss (Quadratic) $\rightarrow (s-\hat{s})^2$

② Absolute Error Loss $\longrightarrow |s-\hat{s}|$

③ 0-1 Error Loss $\longrightarrow \mathbb{1}_{s\neq\hat{s}} = \begin{cases} 1, & s\neq\hat{s} \\ 0, & s=\hat{s} \end{cases}$

$\hat{s} \Rightarrow$ estimate of s $\Rightarrow f(r_1, r_2, r_3) = f(\underline{r})$

How do we find f?

Cost / Risk $\longrightarrow$ $J = \underset{s,\hat{s}}{E}\left\{\ell(s,\hat{s})\right\}$

$\downarrow$ expectation of errors

$= \underset{s,\underline{r}}{E}\left\{\ell(s,\hat{s})\right\}$

s & $\hat{s}$ $\rightarrow$ random

$\hat{s}$ $\downarrow$ observation (a function of observations ($\underline{r}$))

For case ① $\longrightarrow$ Square Error loss $\longrightarrow$ $J = \underbrace{\underset{s,\underline{r}}{E}\left\{(s-\hat{s})^2\right\}}_{\text{"MSE"}}$

mean square error

* there's a special f function that minimizes the cost function.

$$f_{OPT}^{\text{squareerror}}(\underline{r}) = \underbrace{\underset{s}{E}\left\{s|\underline{r}\right\}}$$ $\longleftarrow$ conditional expectation

$\downarrow$ optimal estimator minimizing MSE.

Let's restrict the estimator to a linear estimator.

$$\hat{s} = f(\underline{r}) = \left[w_1 \quad w_2 \quad w_3\right]\begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} = \sum_{k=1}^{3} w_k r_k$$ $\longrightarrow$ averaging over time (FIR filter gibi)

$$J(\underline{w}) = \underset{s,\underline{r}}{E}\left\{(s-\hat{s})^2\right\} = \underset{s,\underline{r}}{E}\left\{(s-\underline{w}^T\underline{r})^2\right\}$$

$\downarrow$

$$\hat{\underline{w}} = \underset{\underline{w}}{\operatorname{argmin}}\,(J(\underline{w}))$$

optimal weight

LMMSE

Linear Minimum Mean Square Error Problem

The solution of LMMSE problem is also called as **Wiener Filter**. Wiener Filter has several applications such as smoothing, prediction, filtering, etc.

Kalman Filter is a time varying case of Wiener Filter.

$$\text{cost} \rightarrow \underset{s,\hat{s}}{E}\left\{\left(s-\underbrace{\hat{s}}_{\underline{w}^T\underline{r}}\right)^2\right\}$$

* Expectation operation connects this calculation of optimal weights (mathematical problem) to the real world by Law of Large Numbers.

$$\underbrace{E\{X\}}_{x=\{1,2,3,4,5,6\}} = \lim_{\text{trials}\to\infty} \frac{1}{\text{trials}} \left[\sum_{k=1}^{6} k.(\#\text{ times we have } k \text{ as output})\right]$$

Mathematical Reasoning (Proofs, Necessary / Sufficient Conditions, etc)

In many maths problems, we need to process / generate statements including if, if and only if, necessary conditions, sufficient conditions, etc.

$P \rightarrow Q$

P implies Q	Q is implied by P
If P, then Q	P only if Q
P is a sufficient condition for Q	Q is a necessary condition for P

P, Q are logic variables that they are either "False" or "True". "True" can be thought as something taken for granted, correct and so on.

Truth Table

P	Q	$P \to Q$
0	0	1
0	1	1
1	0	0
1	1	1

Claim $\underbrace{(x=1)}_{P} \to \underbrace{(x^2=1)}_{Q}$

This $P \to Q$ claim tells us that whenever $x=1$ ($P=$ True), we have $x^2=1$ ($Q=$ True). When $x \neq 1$ ($P=$ False), Q can be either True or False.

$$x \neq 1 \begin{array}{l} \nearrow x^2 = 1 \\ \searrow x^2 \neq 1 \end{array}$$

The only case outlawed by $P \to Q$ example is $P=$ True and $Q=$ False, $(x=1)$ but $(x^2 \neq 1)$

How to prove $P \to Q$?

① **Direct Proof**

1) $x=1$
2) $x^2=x$

$\updownarrow$ multiply / divide by x

3) $x^2 = x \overset{①}{=} 1 \Rightarrow$ Q is True.

② **Proof by Contraposition**

$P \to Q$ is equivalent to $\overline{P} \vee Q \to$ truth table

$$(P \to Q) \equiv (\overline{P} \vee Q) \equiv (Q \vee \overline{P}) \equiv (\overline{Q} \to \overline{P})$$

* In contraposition proofs, you prove $\overline{Q} \to \overline{P}$ instead of $P \to Q$

$(x=1) \to (x^2=1)$	$(x^2 \neq 1) \to (x \neq 1)$
$P \to Q$	$\overline{Q} \to \overline{P}$

$$\overline{Q} \to x^2 \neq 1 \to |x| \neq 1 \to x \neq \{-1,+1\} \to \boxed{x \neq 1}\ \overline{P}$$

OR

$x \neq -1$

$P \to Q$ → P is a sufficient condition for Q.

$\Updownarrow$

$\overline{Q} \to \overline{P}$ → $\overline{Q}$ is a sufficient condition for $\overline{P}$.

OR

$\overline{P}$ is a necessary condition for $\overline{Q}$.

---

③ Proof by Contradiction

Proof by contradiction focuses on the outlawed case in $P \to Q$ truth table, that is $P=\text{True}$, $Q=\text{False}$.

The goal is to show $P=\text{True}$ and $Q=\text{False}$ is inconsistent (False).

$P \to Q$, show that $P \wedge \overline{Q}$ is inconsistent.

for the example given, $\underbrace{(x=1)}_{P} \to \underbrace{(x^2=1)}_{Q}$

$(P \wedge \overline{Q}) \to (x=1)$ and $(x^2 \neq 1)$

↓

$|x| \neq 1 \to x \neq \{-1,+1\} \to$ contradiction

↓

$(x=1)$ is violated!

If and Only If $(P \leftrightarrow Q)$

P	Q	$P \leftrightarrow Q$
0	0	1
0	1	0
1	0	0
1	1	1

Example $(x=1) \leftrightarrow (2x=2)$

Clearly, $P \leftrightarrow Q$ is equivalent to $P \to Q$ and $P \leftarrow Q$

If and only if is an equivalency statement, so we can use P instead of Q or vice versa with no harm.

$$\begin{tikzcd} (P \to Q) \arrow[d] & \text{and} & (Q \to P) \arrow[d] \arrow[r] & P \leftrightarrow Q \\ (\bar{P} \lor Q) & \text{and} & (\bar{Q} \lor P) \arrow[r] & (\bar{P} \lor Q) \land (\bar{Q} \lor P) \end{tikzcd}$$

same truth table

Proof methods for $P \leftrightarrow Q$

① Prove $P \to Q$ and $Q \to P$

② Prove $P \to Q$ and $\bar{P} \to \bar{Q}$

equivalent to each other

Comments In $P \to Q$, the statement can be considered as an indicator P, that is Q is a necessary condition for P.

In some problems, we may have several necessary conditions and combination of many necessary conditions can result in a necessary and sufficient condition.

Example

① $(x=1) \rightarrow (x^2=1)$

② $(x=1) \rightarrow (x>0)$

$(x=1) \rightarrow \boxed{(x^2=1) \text{ and } (x>0)}$

↓

$x=1$

Two necessary conditions   Combined necessary conditions

The logic is strict, since it does not allow almost correct or almost incorrect logic states.

Let's have an example of human reasoning in court.

Example

Let's assume there's a murder in Kızılay at 24:00 by a male and male wears pink, yellow, white, black mixed colored t-shirt.

Murderer is X →
- X is in Kızılay at 24:00.
- X is male
- X wears mixed colored t-shirt.

} probabilistic reasoning NOT LOGICAL

↓ necessary conditions, not logic

↓ assign probabilities to those events.

Example $\sqrt{2}$ is an irrational number.

(by contradiction)

$(1=1) \longrightarrow \left(\sqrt{2} = \frac{a}{b}\right)$ $(P \rightarrow Q) \Rightarrow P \wedge \bar{Q}$?

↓ any statement $a, b \in I$ $(1=1)$ and $\left(\sqrt{2} = \frac{a}{b}\right)$ = False

a, b : integers

Linear Algebra Review

$$\underline{\underline{A}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & & & \vdots \\ \vdots & & & \vdots \\ a_{m1} & \cdots & \cdots & a_{mn} \end{bmatrix}_{M \times N} = \left[\underline{a_1} \;\; \underline{a_2} \;\; \cdots \;\; \underline{a_N} \right],$$

$\underline{a_i} \rightarrow M \times 1$

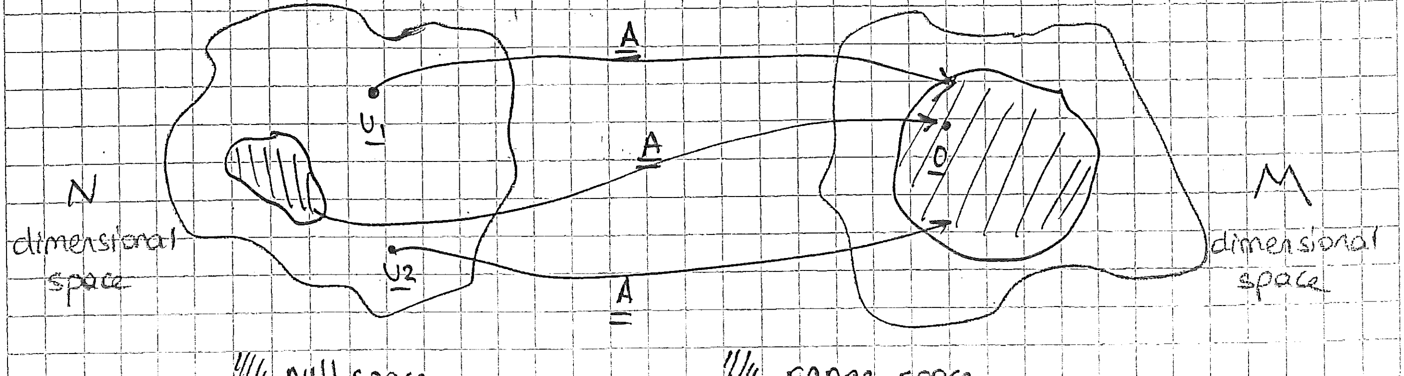
$$\underline{\underline{A}}\,\underline{x} = \underbrace{\left[\underline{a_1} \;\; \underline{a_2} \;\; \underline{a_3} \right]}_{\downarrow \; \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \underbrace{x_1 \underline{a_1} + x_2 \underline{a_2} + x_3 \underline{a_3}}_{\text{a linear combination of columns of } \underline{\underline{A}}.}$$

1) Range$(\underline{\underline{A}})$ $\quad \{ \underline{v} : \underline{v} = \underset{M \times N}{\underline{\underline{A}}}\, \underline{x} \;,\; x \in R^N \}$

- range space of $\underline{\underline{A}}$
- column space of $\underline{\underline{A}}$

2) Null $(\underline{\underline{A}})$ $\quad \{ \underline{v} : \underset{M \times N}{\underline{\underline{A}}}\, \underline{v} = \underline{0} \}$

- null space of $\underline{\underline{A}}$



▨ null space ▨ range space

$\underline{\underline{A}}\,\underline{x} = \underline{b}$ → for $\underline{x}$, we are searching for $\underline{b}$ in range space of $\underline{\underline{A}}$

↓

Assume that $\underline{u_x}$ is in the null space of $\underline{\underline{A}}$.

$\underline{\underline{A}}\,\underline{u_x} = \underline{0}$

Let's assume $\underline{x_*}$ satisfies $\underline{\underline{A}}\,\underline{x} = \underline{b}$ equality.

$$\left[\underline{\underline{A}}\,\underline{x_*}\right] = \left[\underline{b}\right] = \left[\underline{b} + \underline{0}\right] = \left[\underline{b} + \underline{\underline{A}}\,\underline{u_x}\right] = \left[\underline{\underline{A}}\,\underline{x_*} + \underline{\underline{A}}\,\underline{u_x}\right] = \left[\underline{\underline{A}}\left(\underline{x_*} + \underline{u_x}\right)\right]$$

a vector in the null space

a solution

* We can generate other solutions.

↓

infinite number of solutions.

How do we understand that null$(\underline{\underline{A}}) = \underline{0}$ that is null space of $\underline{\underline{A}}$ is just $\underline{0}$?

This is important since this case shows that if there's a solution, that solution is a <u>unique solution</u> of $\underline{\underline{A}}\,\underline{x} = \underline{b}$ equation system.

Nullspace check

$$\left[\underline{a_1} \;\; \underline{a_2} \;\; \cdots \;\; \underline{a_N}\right] \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

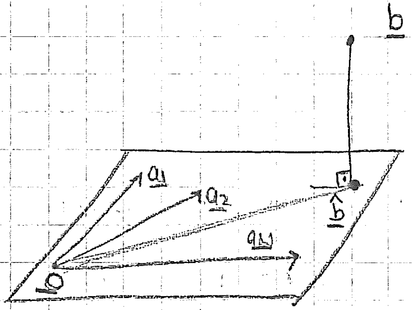
} the linear combination of columns of $\underline{\underline{A}}$ contains $\underline{0}$ that is columns of $\underline{\underline{A}}$ are NOT linearly independent.

$$\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} \longrightarrow \text{column rank}(\underline{\underline{A}}) = \#\text{ independent columns.}$$

→ finding the biggest non-zero determinant.

The largest dimensional non-zero determinant gives you the column rank of $\underline{\underline{A}}$.

Projection Matrices



$\underline{b} \notin \text{Range}(\underline{\underline{A}})$

$3\underline{a_1} + 5\underline{a_2} + \dots + 10\underline{a_N}$

$$\text{Span}(\underline{a_1}, \underline{a_2}, \dots, \underline{a_N}) = \text{Range}([\underline{a_1}\ \underline{a_2} \dots \underline{a_N}])$$

$\underline{\underline{A}}\underline{x} = \hat{\underline{b}}$, goal is minimizing $\|\underline{b} - \hat{\underline{b}}\| = \|\underline{b} - \underline{\underline{A}}\underline{x}\|$

$$\underset{\substack{\downarrow \\ N\times 1}}{\|\underline{x}\|} = \sqrt{\sum_{k=1}^{N} x_k^2} \quad \Big\} \text{ Euclidian norm}$$

$\|\underline{x}\|$ is a mapping from $\underline{x} \in R^N$ to a real number $\|\underline{x}\| : R^N \to R$

Find nearest vector on the plane to vector b.
↓
distance between points.

$\underline{v} = \underline{\underline{A}}\underline{x} = \underline{a_1}x_1 + \underline{a_2}x_2 + \underline{a_3}x_3 + \dots + \underline{a_N}x_N$

$\underline{x}$ ↓ linear combination vector

$\underline{v}$ ↓ element of Range($\underline{\underline{A}}$)

$\hat{\underline{b}} = \underline{\underline{A}}x$

Distance metric → $d(\underline{b}, \hat{\underline{b}}) : R^N \times R^N \to [0, \infty)$

Axiom ① $d(\underline{x},\underline{y}) = d(\underline{y},\underline{x})$ ——— (symmetric)

Axiom ② $d(\underline{x},\underline{y}) \leqslant d(\underline{x},\underline{z}) + d(\underline{z},\underline{y})$ ——— (triangle inequality)

Axiom ③ $d(\underline{x},\underline{y}) = 0 \overset{\text{iff}}{\longleftrightarrow} \underline{x} = \underline{y}$

Norm Function $\longrightarrow$ $\|\underline{x}\| \,;\, R^N \longrightarrow [0,\infty)$

Axiom ① $\|\underline{x}\| = 0 \xleftrightarrow{\text{iff}} \underline{x} = \underline{0}$

Axiom ② $\|\underline{x}+\underline{y}\| \leq \|\underline{x}\| + \|\underline{y}\|$ ______ (triangle inequality)

Axiom ③ $\|\alpha \underline{x}\| = |\alpha| \, \|\underline{x}\|$ __________ (scaling property)

$\alpha \rightarrow$ real number

Assume, we have a norm function and define a metric as

$$d(\underline{x},\underline{y}) \triangleq \|\underline{x}-\underline{y}\|$$

$d(\underline{x},\underline{y}) \rightarrow$ metric induced by norm

<u>Question</u> Is $d(\underline{x},\underline{y})$ a valid metric function?

Yes.

$$\underbrace{\|\underline{x}-\underline{y}\|} = \|\underline{x}-\underline{z}+\underline{z}-\underline{y}\| \leqslant \underbrace{\|\underline{x}-\underline{z}\|} + \underbrace{\|\underline{z}-\underline{y}\|}$$

$$d(\underline{x},\underline{y}) \leqslant d(\underline{x},\underline{z}) + d(\underline{z},\underline{y}) \quad \checkmark \quad ②$$

$$\underbrace{\|\underline{x}-\underline{y}\|} = |-1| \, \|\underline{x}-\underline{y}\| = \underbrace{\|\underline{y}-\underline{x}\|} \quad \checkmark \quad ①$$

$$d(\underline{x},\underline{y}) = d(\underline{y},\underline{x})$$

$$\underbrace{\|\underline{x}-\underline{y}\|} = 0 \longleftrightarrow \underbrace{\underline{x}-\underline{y} = \underline{0}} \quad \checkmark \quad ③$$

$$d(\underline{x},\underline{y}) = 0 \longleftrightarrow \underline{x} = \underline{y}$$

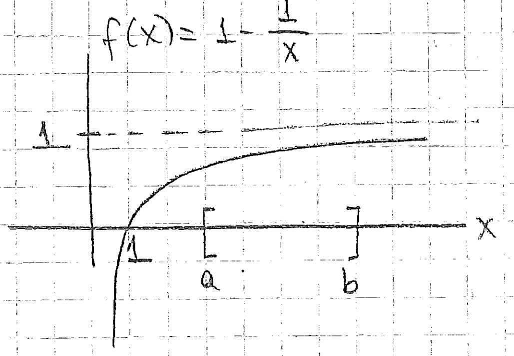
Problem $\underline{b}_* = \operatorname{argmin} \|\underline{b} - \hat{\underline{b}}\|$

$\hat{\underline{b}} \in \text{Range}(\underline{\underline{A}})$

$\boxed{Q_1}$ $\underline{b}_*$ exists or not?

$\boxed{Q_2}$ If exists, is it unique?

$\boxed{Q_3}$ Is there a method (feasible method) to calculate $\underline{b}_*$?



* is there a maximum in [a,b]?

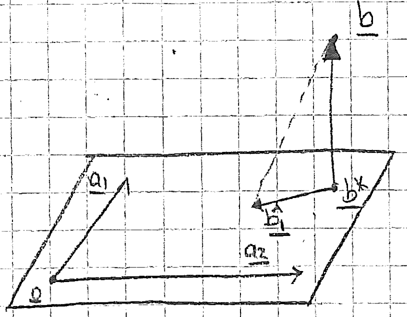
Yes.

* is there a maximum in [0,∞)?

No, keeps increasing

! Optimum solution for a problem MAY or MAY NOT exist.

3D-Case



$\underline{b} - \underline{b}_*$ } error (orthogonal to the plane)

$\|\underline{x}\|_2 = \sqrt{x_1^2 + x_2^2 + x_3^2}$ → Euclidian norm

For given $\hat{\underline{b}}_1 \Rightarrow \|\underline{b} - \hat{\underline{b}}_1\|^2 = \|\underline{b} - \underline{b}_*\|^2 + \|\hat{\underline{b}}_1 - \underline{b}_*\|^2$ → Phythagorean's Theorem

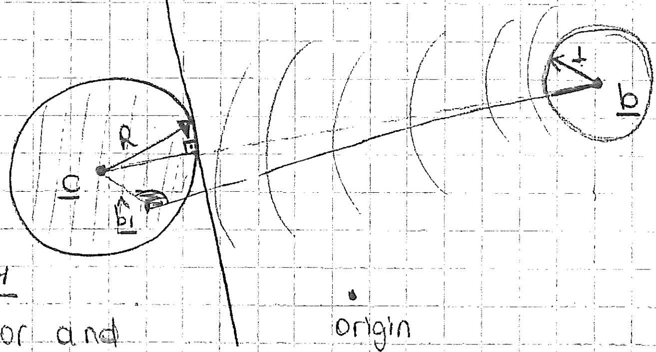
The distance between any other candidate becomes larger.

(norms are always greater than zero)

Projection operation is the mapping of $\underline{b}$ to $\underline{b}_*$ → optimal

(closest distance vector)

2^{nd} Projection Case



start with unit circle, infinitely extend till the tangent.
(enlarge the radius efficiently enough, it will be touching at single point (probably))

set C

interior and boundaries are included.

closed, convex set of points

$$\underline{b_*} = \underset{\hat{\underline{b}} \in C}{\operatorname{argmin}} (\| \underline{b} - \hat{\underline{b}} \|)$$

↓

at the tangency point
↓
90 degrees

$\boxed{\underline{b_*} \text{ is just at the boundary}}$

another candidate $\hat{\underline{b_1}}$ → forms an obtuse angle triangle (wide angle)

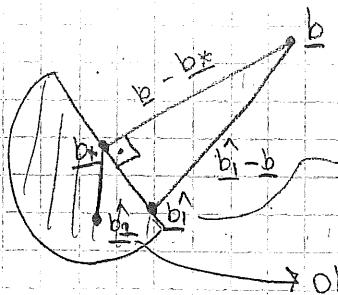
$$\underline{b_*} = \underline{c} + \boxed{\frac{(\underline{b} - \underline{c})}{\| \underline{b} - \underline{c} \|}} \cdot R$$

unit norm (from $\underline{c}$ towards $\underline{b}$)

3^{rd} Projection Case

$$\underline{b_*} = \underset{\hat{\underline{b}} \in C}{\operatorname{argmin}} \| \underline{b} - \hat{\underline{b}} \|$$

↓



→ Phy. Thm. → larger distance

→ obtuse triangle

set C

closed, convex set of points
↓
important for existence of solution.

optimal point $\underline{b_*}$ satisfies

$$(\underline{b} - \underline{b_*})^T (\hat{\underline{b_2}} - \underline{b_*}) \leq 0$$

$$\underline{v_1}^T \underline{v_2} \leq 0$$

angle between $\underline{v_1}$ and $\underline{v_2}$ is obtuse wide angle.

In 2D/3D examples given, the concept of angle turned out to be very useful for the decision of optimality.

To introduce angles, we need to define inner products.

Inner Product

$$(\underline{x},\underline{y}) : R_n \times R_n \rightarrow R$$
$$(C_n \times C_n \rightarrow C)$$

Axiom ① $(\underline{x},\underline{y}) = (\underline{y},\underline{x})^*$ ———— (conjugate symmetry)

Axiom ② $(\underline{x}+\underline{y},\underline{z}) = (\underline{x},\underline{z}) + (\underline{y},\underline{z})$

Axiom ③ $(\lambda \underline{x},\underline{y}) = \lambda(\underline{x},\underline{y})$

} ———— (linearity conditions in first variable)

Axiom ④ $(\underline{x},\underline{x}) \geqslant 0$ and $(\underline{x},\underline{x}) = 0 \Leftrightarrow \underline{x} = \underline{0}$

Remember, usual inner product for Euclidian Geometry is

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad \underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \qquad (\underline{x},\underline{y}) = \underline{y}^T \underline{x} = x_1 y_1 + x_2 y_2 + x_3 y_3$$

Claim Given a valid inner product $(\underline{x},\underline{y})$; we can define a norm $\|\underline{x}\|^2 \triangleq (\underline{x},\underline{x})$

$\hookrightarrow$ induced norm by inner product

Norm Axiom ① $\longrightarrow$ $\|\underline{x}\|^2 = 0 = (\underline{x},\underline{x}) \Leftrightarrow \underline{x} = 0$ ✓

Norm Axiom ③ $\longrightarrow$ $\|\alpha \underline{x}\|^2 = (\alpha \underline{x}, \alpha \underline{x}) = \alpha^2 (\underline{x},\underline{x}) = \alpha^2 \|\underline{x}\|^2$ ✓

Norm Axiom ② $\longrightarrow$ So, we need to only verify Δ Inequality axiom for the norm, to prove that the induced norm by inner product is indeed a norm, i.e. $\|\underline{x}\| \triangleq \sqrt{(\underline{x},\underline{x})}$

To show Δ inequality, we will first prove another important result, called <u>Cauchy - Schwarz</u> inequality.

Cauchy - Schwarz Inequality

$$|(\underline{x},\underline{y})| \leqslant \|\underline{x}\| . \|\underline{y}\|$$

19.10.2020

To verify norm axiom #2 $\left(\|\underline{x}+\underline{y}\| \leqslant \|\underline{x}\| + \|\underline{y}\| \right)$,

prove $|(\underline{x},\underline{y})| \leqslant \|\underline{x}\| . \|\underline{y}\| \quad \forall\ \underline{x}, \underline{y}$

Proof of Cauchy-Schwarz →

$$(\underline{x}+\overset{\text{scalar}}{\lambda}\underline{y}, \underline{x}+\lambda\underline{y}) \overset{\text{Linearity}}{=} \lambda^2\underbrace{(\underline{y},\underline{y})}_{a} + 2\lambda\underbrace{(\underline{x},\underline{y})}_{b} + \underbrace{(\underline{x},\underline{x})}_{c} \overset{\text{Inner Product Axiom \#4}}{\geqslant} 0$$

$$= \underbrace{a\lambda^2 + b\lambda + c}_{\text{2nd degree polynomial in } \lambda}$$

$p(\lambda) \geqslant 0 \longrightarrow$ no real roots or repeated roots

Since $p(\lambda) \geqslant 0 \quad \forall \lambda$, $\underset{\text{discriminant}}{\Delta} = b^2 - 4ac \leqslant 0$

$$4(\underline{x},\underline{y})^2 - 4\,(\underline{y},\underline{y})\,(\underline{x},\underline{x}) \leqslant 0$$

$$(\underline{x},\underline{y})^2 - \|\underline{y}\|^2\,\|\underline{x}\|^2 \leqslant 0$$

$$|(\underline{x},\underline{y})| \leqslant \|\underline{x}\|\,\|\underline{y}\|$$

Case of Equality for Cauchy-Schwarz:

$|(\underline{x},\underline{y})| \leq \|\underline{x}\|\cdot\|\underline{y}\| \longrightarrow \Delta = 0 \longrightarrow \exists\ \lambda_x$ s.t. $p(\lambda_x)=0$

↳ root of $p(\lambda)$

Then, $p(\lambda_x)=0 \longrightarrow (\underline{x}+\lambda_x\underline{y},\ \underline{x}+\lambda_x\underline{y})=0$

$\longrightarrow \|\underline{x}+\lambda_x\underline{y}\|^2=0$

$\longrightarrow \underline{x}=-\lambda_x\underline{y} \longrightarrow$ Equality condition for Cauchy-Schwarz

$\underline{x}$ and $\underline{y}$ are along the same direction

Cauchy-Schwarz states that

$$\frac{|(\underline{x},\underline{y})|}{\|\underline{x}\|\cdot\|\underline{y}\|} \leq 1 \longrightarrow -1 \leq \boxed{\frac{(\underline{x},\underline{y})}{\|\underline{x}\|\cdot\|\underline{y}\|}} \leq +1$$

↓

$\cos(\theta)$

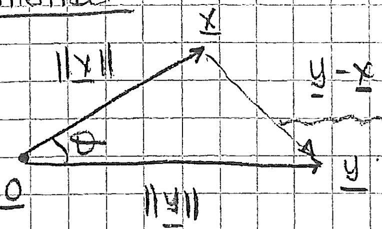
↳ angle between x-vector and y-vector

$$\cos(\theta) \triangleq \frac{(\underline{x},\underline{y})}{\|\underline{x}\|\cdot\|\underline{y}\|}$$

Usual 2/3 Dimensional Geometry → $(\underline{x},\underline{y}) = x_1y_1+x_2y_2$

$\swarrow\ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \searrow\ \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$

Remember



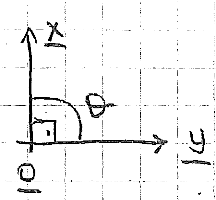
→ length: $\|\underline{y}-\underline{x}\|$

$$\|\underline{y}-\underline{x}\|^2 \triangleq (\underline{y}-\underline{x},\ \underline{y}-\underline{x}) = \|\underline{y}\|^2 - 2\overbrace{(\underline{x},\underline{y})}^{\|\underline{x}\|\cdot\|\underline{y}\|\cdot\cos\theta} + \|\underline{x}\|^2$$

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

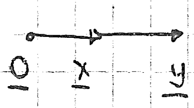
} Cosine Theorem for 2D Geometry

Case #1 $\theta = 90^\circ$



, $\underline{x} \perp \underline{y}$, $(\underline{x}, \underline{y}) = 0$

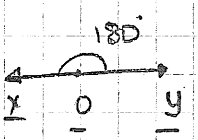
Case # 2 $\theta = 0^\circ$



, $\underline{x} \parallel \underline{y}$, $\underline{x} = \underset{\alpha > 0}{\alpha}\, \underline{y}$

Equality Cases for Cauchy-Schwarz

Case # 3 $\theta = 180^\circ$



, $\underline{x} = \underset{\beta < 0}{\beta}\, \underline{y}$

Let's finally prove that Δ inequality is indeed satisfied by

$$\|\underline{x}\| \triangleq \sqrt{(\underline{x}, \underline{x})}$$

Proof →
$$(\underline{x}+\underline{y}, \underline{x}+\underline{y}) = \|\underline{x}\|^2 + 2(\underline{x}, \underline{y}) + \|\underline{y}\|^2$$

→ max. value can be $\|\underline{x}\| . \|\underline{y}\|$

$$\leq \|\underline{x}\|^2 + \|\underline{y}\|^2 + 2\|\underline{x}\| . \|\underline{y}\|$$

Cauchy Schwarz

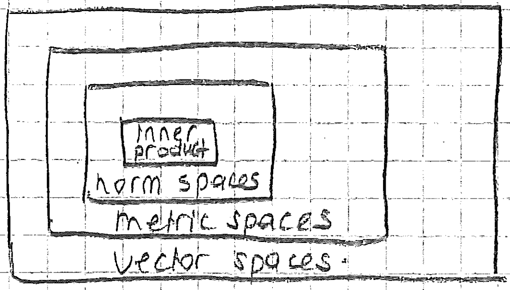
$$\|\underline{x}+\underline{y}\|^2 \leq (\|\underline{x}\| + \|\underline{y}\|)^2$$

$$\|\underline{x}+\underline{y}\| \leq \|\underline{x}\| + \|\underline{y}\|$$

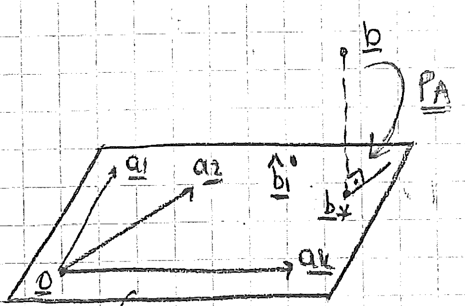
all positive, no need for absolute.

□

Induced norm from an inner product is indeed a norm.



Projection Matrices



$$\underbrace{\text{span}(\underline{a_1}, \underline{a_2}, \cdots, \underline{a_k})}_{\text{vectors in } \mathbb{R}^N}$$

$\underline{b} \in \mathbb{R}^N$

Projection operation maps $\underline{b}$ to the closest point on the constraint set $\{\text{span}(\underline{a_1}, \underline{a_2}, \dots, \underline{a_k})\}$

$$\hat{\underline{b}}_1 = \underset{N\times k}{\underline{A}}\ \underset{k\times 1}{\underline{x}}$$

$$\hat{\underline{b}}_1 = \begin{bmatrix} \underline{a_1} & \underline{a_2} & \dots & \underline{a_k} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix} = x_1\underline{a_1} + x_2\underline{a_2} + \dots + x_k\underline{a_k}$$

If b_* is the closest, then $\underbrace{\underline{b} - \underline{b_*}}_{\text{error vector}}$ should be orthogonal to the $\underline{a_i}$, the combinations of them, the plane.

$$\left.\begin{array}{lll} 1)\ (\underline{b}-\underline{b_*}) \perp \underline{a_1} & \rightarrow & (\underline{a_1}, \underline{b}-\underline{b_*})=0 \\ 2)\ (\underline{b}-\underline{b_*}) \perp \underline{a_2} & \rightarrow & (\underline{a_2}, \underline{b}-\underline{b_*})=0 \\ \quad \vdots & & \quad \vdots \\ k)\ (\underline{b}-\underline{b_*}) \perp \underline{a_k} & \rightarrow & (\underline{a_k}, \underline{b}-\underline{b_*})=0 \end{array}\right\}$$

by linearity, we can multiply each relation with x_i, sum them up

$\underline{b}-\underline{b_*}$ is orthogonal to the span of $\underline{a_i}$'s.

1) $\underline{a}_1^T(\underline{b}-\underline{b}_*)=0$

2) $\underline{a}_2^T(\underline{b}-\underline{b}_*)=0$

$\vdots$

k) $\underline{a}_k^T(\underline{b}-\underline{b}_*)=0$

$$\begin{bmatrix} \underline{a}_1^T \\ \underline{a}_2^T \\ \vdots \\ \underline{a}_k^T \end{bmatrix}_{k\times N} \begin{bmatrix} \underline{b}-\underline{b}_* \end{bmatrix}_{N\times 1} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$\underline{b}_*$ └ element of span($\underline{\underline{A}}$)

$\downarrow$

$$\underline{\underline{A}}\,\underline{x}_* = \underline{b}_*$$

a special combination of $\underline{a}_i$'s → $\underline{x}_*$ gives $\underline{b}_*$

$$\underline{\underline{A}}^T(\underline{b}-\underline{b}_*)=\underline{0}$$

$$\underline{\underline{A}}^T(\underline{b}-\underline{\underline{A}}\,\underline{x}_*)=\underline{0}$$

$$\boxed{\underline{\underline{A}}^T\underline{\underline{A}}\,\underline{x}_* = \underline{\underline{A}}^T\underline{b}}$$

searching for $\underline{x}_*$

Case 1: (A^TA) is invertible | Case 2: (A^TA) is not invertible

Case 1:

$$\underline{x}_* = (\underline{\underline{A}}^T\underline{\underline{A}})^{-1}\underline{\underline{A}}^T\underline{b}$$

$\downarrow$

$$\underline{b}_* = \underline{\underline{A}}\,\underline{x}_* = \boxed{\underline{\underline{A}}(\underline{\underline{A}}^T\underline{\underline{A}})^{-1}\underline{\underline{A}}^T}\,\underline{b}$$

$\underline{\underline{P_A}}$: projection matrix to Range($\underline{\underline{A}}$)

Case 2:

Equation system has a solution since (**)

Range($\underline{\underline{A}}^T\underline{\underline{A}}$) = Range($\underline{\underline{A}}^T$)

$\boxed{\underline{\underline{A}}^T\underline{\underline{A}}}$ is called as Gram matrix and is invertible if $\underline{\underline{A}}$ matrix if full column rank, i.e. $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_k$ are linearly independent.

(there's a unique way of expressing any vector in the span)

(**) Showing Range($\underline{\underline{A}}^T\underline{\underline{A}}$) = Range($\underline{\underline{A}}^T$) can be trivial via SVD
(singular value decomposition)

$\underline{\underline{A}} = \underline{\underline{V}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}^T$ → unitary, orthogonal (range space is not affected by them)

$\underline{\underline{A}}^T\underline{\underline{A}} = \underline{\underline{V}}\,\underline{\underline{\Sigma}}^2\,\underline{\underline{V}}^T$

→ (range space is affected by them)

$\underline{\underline{A}}^T = \underline{\underline{V}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}^T$

$\underline{\underline{\Sigma}}$ → singular values of $\underline{\underline{A}}$ (diagonal)

if they are not independent, find a linearly independent subspace
(eliminate dependent ones)
find a set (basis) spanning the subspace.

$$\underline{\underline{P_A}} = \underline{\underline{A}} \, (\underline{\underline{A}}^T \underline{\underline{A}})^{-1} \, \underline{\underline{A}}^T$$

Orthogonal Projectors

A matrix $\underline{\underline{P}}$ is an orthogonal projector if it satisfies

① $\underline{\underline{P}}^2 = \underline{\underline{P}}$ → projection condition → 2nd projection is meaningless.

② $\underline{\underline{P}}^T = \underline{\underline{P}}$ → orthogonality condition

$$\underline{\underline{P}}_A^2 \, \underline{b} = \underline{\underline{P_A}} \, \underbrace{(\underline{\underline{P_A}} \, \underline{b})}_{\underline{b_{\parallel}}} = \underline{b_{\parallel}}$$

↓

∈ Range($\underline{\underline{A}}$)
(already projected)
(already at 0 distance)

A matrix is called symmetric (Hermitian / symmetric)

if $\underline{\underline{A}}^T = \underline{\underline{A}}$ $\quad$ $(\underline{\underline{A}}^H = \underline{\underline{A}})$ $\quad$ [for complex → $\underline{\underline{A}}^H = (\underline{\underline{A}}^T)^*$]

Question Is $\underline{\underline{P_A}} = \underline{\underline{A}}(\underline{\underline{A}}^T\underline{\underline{A}})^{-1}\underline{\underline{A}}^T$ an orthogonal projector?

① Is $\underline{\underline{P_A}}^2 = \underline{\underline{P_A}}$?

$$\underline{\underline{P_A}}^2 = (\underline{\underline{A}}(\underline{\underline{A}}^T\underline{\underline{A}})^{-1}\underline{\underline{A}}^T)(\underline{\underline{A}}(\underline{\underline{A}}^T\underline{\underline{A}})^{-1}\underline{\underline{A}}^T) = \underline{\underline{A}}(\underline{\underline{A}}^T\underline{\underline{A}})^{-1}\underline{\underline{A}}^T = \underline{\underline{P_A}} \quad \checkmark$$

$$\underbrace{\underline{\underline{A}}^T \, \underline{\underline{A}} \, (\underline{\underline{A}}^T\underline{\underline{A}})^{-1}}_{\underline{\underline{I}}}$$

② Is $\underline{\underline{P_A}}^T = \underline{\underline{P_A}}$?

$$\underline{\underline{P_A}}^T = (\underline{\underline{A}}(\underline{\underline{A}}^T\underline{\underline{A}})^{-1}\underline{\underline{A}}^T)^T = \underline{\underline{A}}\,((\underline{\underline{A}}^T\underline{\underline{A}})^{-1})^T\,\underline{\underline{A}}^T$$

$$= \underline{\underline{A}}\,((\underline{\underline{A}}^T\underline{\underline{A}})^T)^{-1}\,\underline{\underline{A}}^T$$

$$= \underline{\underline{A}}\,(\underline{\underline{A}}^T\underline{\underline{A}})^{-1}\,\underline{\underline{A}}^T$$

$$= \underline{\underline{P_A}} \quad \checkmark$$

Remember

If $\underline{\underline{A}}^T = \underline{\underline{A}}$, then $\underline{\underline{A}}$ has orthogonal eigenvectors.

A more general theorem says that if $\underline{\underline{M}}\,\underline{\underline{M}}^T = \underline{\underline{M}}^T\underline{\underline{M}}$, then $\underline{\underline{M}}$ is called a normal matrix and it has orthogonal eigenvectors.

Orthogonal matrices: $\underline{\underline{A}}^T\underline{\underline{A}} = I_{k\times k}$

$\downarrow$ orthogonal matrix

$$\underline{\underline{A}}^T\underline{\underline{A}} = \begin{bmatrix} \underline{a_1}^T \\ \underline{a_2}^T \\ \vdots \\ \underline{a_k}^T \end{bmatrix} \begin{bmatrix} \underline{a_1} & \underline{a_2} & \cdots & \underline{a_k} \end{bmatrix} = \begin{bmatrix} \underline{a_1}^T\underline{a_1} & \underline{a_1}^T\underline{a_2} & \cdots & \underline{a_1}^T\underline{a_k} \\ \underline{a_2}^T\underline{a_1} & \cdots & & \\ \vdots & & & \\ \underline{a_k}^T\underline{a_1} & \cdots & \cdots & \underline{a_k}^T\underline{a_k} \end{bmatrix}$$

For the complex case, we use unitary matrices instead of orthogonal matrices.

So, since $\underline{\underline{P_A}}^T = \underline{\underline{P_A}}$, $\underline{\underline{P_A}}$ has orthogonal eigenvectors.

Let's study eigendecomposition of $\underline{\underline{P_A}}$.

① Eigenvalues

eigenvector

$$\underline{\underline{P_A}}\,\underline{e_k} = \lambda_k \underline{e_k}$$

eigenvalue

$$\underline{\underline{P_A}}^2 = \underline{\underline{P_A}} \rightarrow \underline{\underline{P_A}}^2 - \underline{\underline{P_A}} = \underline{\underline{0}}$$

$$\downarrow$$

$$(\underline{\underline{P_A}}^2 - \underline{\underline{P_A}})\underline{e_k} = \underline{\underline{0}}\,\underline{e_k}$$

non-trivial eigenvector

$$\underbrace{\underline{\underline{P_A}}\underbrace{\underline{\underline{P_A}}\,\underline{e_k}}_{\lambda_k \underline{e_k}}}_{\lambda_k^2 \underline{e_k}} - \underbrace{\underline{\underline{P_A}}\,\underline{e_k}}_{\lambda_k \underline{e_k}} = \underline{0} \longrightarrow (\lambda_k^2 - \lambda_k)\underline{e_k} = \underline{0}$$

$$\downarrow$$

$$\boxed{\lambda_k = \{0, +1\}}$$

eigenvalues of the projection matrix $(\underline{\underline{P_A}})$

② Eigenvectors $\underline{e_k}$ $k = 1, \dots, N$

$$\underline{\underline{P_A}}_{N\times N} \, \underline{e_k}_{N\times 1} = \lambda_k \, \underline{e_k}_{N\times 1}$$

$$\underline{\underline{P_A}} \underbrace{\left[\underline{e_1} \;\; \underline{e_2} \; \cdots \; \underline{e_N} \right]}_{\underline{\underline{E}}} = \left[\underline{e_1} \;\; \underline{e_2} \; \cdots \; \underline{e_N} \right] \underbrace{\begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_N \end{bmatrix}}_{\underline{\underline{\Omega}} \rightarrow \text{diagonal}}$$

$$\underline{\underline{P_A}} \, \underline{\underline{E}} = \underline{\underline{E}} \, \underline{\underline{\Omega}}$$

ice $\underline{\underline{E}}$ inearly dependent →

$$\underline{\underline{P_A}} = \underline{\underline{E}} \, \underline{\underline{\Omega}} \, \underline{\underline{E}}^{-1} = \underline{\underline{E}} \, \underline{\underline{\Omega}} \, \underline{\underline{E}}^{T}$$

since $\underline{\underline{E}}$ is orthogonal, $E^T E = I \rightarrow E^T = E^{-1}$

$$\underline{\underline{P_A}} = \left[\underline{e_1} \;\; \underline{e_2} \; \cdots \; \underline{e_N} \right] \underbrace{\begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_N \end{bmatrix} \begin{bmatrix} \underline{e_1}^T \\ \underline{e_2}^T \\ \vdots \\ \underline{e_N}^T \end{bmatrix}}_{\begin{bmatrix} \lambda_1 \underline{e_1}^T \\ \lambda_2 \underline{e_2}^T \\ \vdots \\ \lambda_N \underline{e_N}^T \end{bmatrix}}$$

$$\underline{\underline{P_A}} = \sum_{n=1}^{N} \boxed{\lambda_n \underline{e_n} \, \underline{e_n}^T}$$ rank-1 matrix

rank-1 matrix

$\rightarrow \lambda_k \in \{0, +1\}$

If $\underline{\underline{P_A}}$ projects to a space of dimension K, that is the projection space is Range $(\underline{\underline{A}}) \rightarrow \underbrace{\left[\underline{a_1} \;\; \underline{a_2} \; \cdots \; \underline{a_K} \right]}_{\text{linearly independent vectors.}}$

So, $\underline{\underline{P_A}} = \sum_{k=1}^{K} \underline{e_k}\,\underline{e_k}^T$

$$\lambda_1 = \lambda_2 = \dots = \lambda_K = 1$$

$$\lambda_{K+1} = \lambda_{K+2} = \dots = \lambda_N = 0$$

Question

If the span of $\{\underline{a_1}, \underline{a_2}, \dots, \underline{a_N}\}$ is represented by a different basis (such as $\underline{e_1}, \underline{e_2}, \dots, \underline{e_N}$), is there a change in $\underline{\underline{P_A}}$?

$$\underline{\underline{P_A}} = \underline{\underline{A}}\,(\underline{\underline{A}}^T \underline{\underline{A}})^{-1} \underline{\underline{A}}^T \longrightarrow$$ not a property of $\underline{\underline{A}}$, but a property of space

Answer

e is another representation of space.

$$\underset{N\times K}{\underline{\underline{A}}} = \underset{N\times K}{\underline{\underline{B}}}\ \underset{K\times K}{\underline{\underline{X}}}$$

$$\underline{\underline{A}} \downarrow [\underline{a_1}\ \underline{a_2} \dots \underline{a_K}] \qquad \underline{\underline{B}} \downarrow [\underline{b_1}\ \underline{b_2} \dots \underline{b_K}]$$

$$[\underline{a_1}\ \underline{a_2}\ \dots\ \underline{a_K}] = [\underline{b_1}\ \underline{b_2}\ \dots\ \underline{b_K}] \begin{array}{c} \begin{array}{ccc} \underline{x_1} & & \underline{x_K} \end{array} \\ \left[\begin{array}{ccc} x_{11} & \dots & x_{1K} \\ \vdots & & \\ x_{K1} & & x_{KK} \end{array}\right] \end{array}$$

$\underline{a_1}$ is represented by $\overset{\text{basis}}{\underline{\underline{B}}}\ \underline{x_1} \leftarrow$ coefficients

$$\underline{\underline{P_A}} = (\underline{\underline{B}}\,\underline{\underline{X}})\,(\underline{\underline{X}}^T \underline{\underline{B}}^T \underline{\underline{B}}\,\underline{\underline{X}})^{-1} (\underline{\underline{X}}^T \underline{\underline{B}}^T)$$

$$= \underline{\underline{B}}\,\underbrace{\underline{\underline{X}}\,\underline{\underline{X}}^{-1}}_{I}\,(\underline{\underline{B}}^T \underline{\underline{B}})^{-1} \underbrace{(\underline{\underline{X}}^T)^{-1}\,\underline{\underline{X}}^T}_{I}\,\underline{\underline{B}}^T$$

$$= \underline{\underline{B}}\,(\underline{\underline{B}}^T \underline{\underline{B}})^{-1} \underline{\underline{B}}^T \longrightarrow$$ independent of $\underline{\underline{X}}$

Complementary Projector

$$P_A^{\perp} = I - P_A$$

$\hookrightarrow$ Projector to the orthogonal space of Range(A)

Question Is $P_A^{\perp}$ a projector? (orthogonal projector)

① $(P_A^{\perp})^2 \overset{?}{=} (P_A^{\perp})$

$$(P_A^{\perp})^2 = (I - P_A)^2 = I - 2P_A + \underset{\underset{P_A}{\downarrow}}{P_A^2} = I - P_A = P_A^{\perp} \quad ✓$$

② $(P_A^{\perp})^T \overset{?}{=} (P_A^{\perp})$

$$(P_A^{\perp})^T = (I - P_A)^T = I - \underset{\underset{P_A}{\downarrow}}{P_A^T} = I - P_A = P_A^{\perp} \quad ✓$$

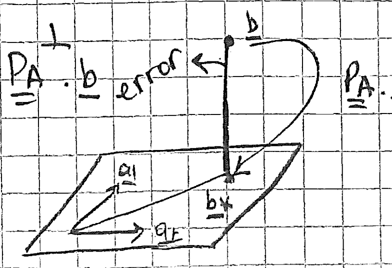
Observe that $\boxed{P_A P_A^{\perp} = P_A^{\perp} P_A = 0}$

$$P_A(I - P_A) = P_A - P_A^2 = P_A - P_A = 0$$

The projection spaces of P_A and $P_A^{\perp}$ are orthogonal to each other.

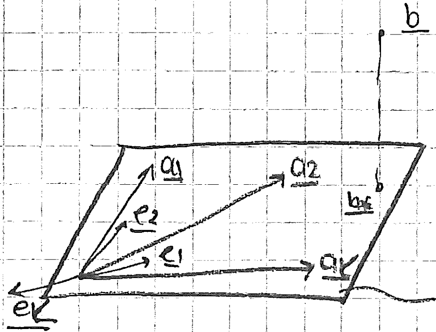
Note that Eigenvectors of $P_A^{\perp}$ are also e_k and eigenvalues of $P_A^{\perp}$ are $(1-\lambda_k)$

Since $P_A^{\perp} e_k = (I - P_A) e_k = e_k - \lambda_k e_k = \underbrace{(1-\lambda_k)}_{\text{eigenvalues of } P_A^{\perp}} e_k$



$$b = b_* + \text{error} = b_* + b - b_* = (P_A \cdot b) + (P_A^{\perp} \cdot b)$$
$$= \underbrace{(P_A + P_A^{\perp})}_{I} \cdot b$$

Orthogonal Basis / Representation with Orthogonal Bases (orthonormal)



subsets of N-dim space.

$$\underbrace{\text{span}(\underline{a_1}, \underline{a_2}, \dots, \underline{a_K}) = \text{span}(\underline{e_1}, \underline{e_2}, \dots, \underline{e_K})}$$

spanning the same space.
linear combinations are always in the same space for each.

Let's assume that $\underline{e_K}$ basis is an <u>orthonormal</u> basis,

$$\underline{e_k}^T \underline{e_l} = \underbrace{\delta[k-l]}_{\text{Kronecker Delta}} = \begin{cases} 1 & k=l \\ 0 & k \neq l \end{cases}$$

Previously, we have seen that Projection Operation / mapping is independent of representation basis.

$$\underset{\text{Range}(\underline{\underline{A}})}{\underline{\underline{P}}_{A}} = \underline{\underline{A}}(\underline{\underline{A}}^T\underline{\underline{A}})^{-1}\underline{\underline{A}}^T = \underline{\underline{E}}\,\underbrace{(\underline{\underline{E}}^T\underline{\underline{E}})}^{-1}\underline{\underline{E}}^T = \underline{\underline{E}}\,\underline{\underline{E}}^T$$

$$\underset{N \times K}{\underline{\underline{E}} = [\underline{e_1}\ \underline{e_2}\ \dots\ \underline{e_K}]}$$

$$[\underline{\underline{E}}^T\underline{\underline{E}}]_{k,l} = \underline{e_k}^T \underline{e_l} = \delta[k-l]$$

k is k^{th} row, l is l^{th} column

$$(\underline{\underline{E}}^T\underline{\underline{E}}) = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & \dots & \dots & 1 \end{bmatrix} = \underline{\underline{I}}$$

matrix of inner products

(orthogonal, $k \neq l \to 0$)

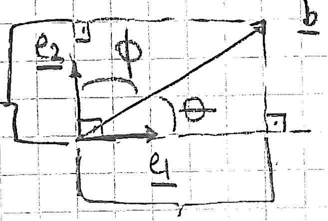
Then,

$$\underline{b_*} = \underline{\underline{P}}_A\,\underline{b} = \underline{\underline{E}}\,\underline{\underline{E}}^T\,\underline{b}$$

$$= \begin{bmatrix} \underline{e_1} & \underline{e_2} & \cdots & \underline{e_K} \end{bmatrix} \begin{bmatrix} \underline{e_1}^T \\ \underline{e_2}^T \\ \vdots \\ \underline{e_K}^T \end{bmatrix} \underline{b} = \begin{bmatrix} \underline{e_1} & \underline{e_2} & \cdots & \underline{e_K} \end{bmatrix} \begin{bmatrix} \underline{e_1}^T\underline{b} \\ \underline{e_2}^T\underline{b} \\ \vdots \\ \underline{e_K}^T\underline{b} \end{bmatrix}$$

$$= (\underline{e_1}^T, \underline{b})\,\underline{e_1} + (\underline{e_2}^T, \underline{b})\,\underline{e_2} + \dots + (\underline{e_K}^T\,\underline{b})\,\underline{e_K}$$ scalar

$$= \sum_{k=1}^{K} \underbrace{(\underline{e_k}^T \cdot \underline{b})}_{\text{expansion coefficient of } \underline{b_*} \text{ in Range}(\underline{\underline{A}})}\,\underline{e_k}$$



length: $\|\underline{b}\| \cdot \cos\theta = \underbrace{\|\underline{e_1}\|}_{\text{unit length}}\,\|\underline{b}\|\cos\theta = (\underline{b}, \underline{e_1})$

direction: $\underline{e_1}$

vector: $(\underline{b}, \underline{e_1}) \cdot \underline{e_1}$

length: $\|\underline{b}\| \cdot \cos\phi = \|\underline{e_2}\| \cdot \|\underline{b}\| \cdot \cos\phi = (\underline{b}, e_2)$

direction: $\underline{e_2}$

vector: $(\underline{b}, \underline{e_2}) \cdot \underline{e_2}$

Examples for Orthonormal Bases

① $\underline{\underline{E}} = \underline{\underline{I}}$ $\quad \underline{e_1} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \; \underline{e_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \; \dots, \; \underline{e_N} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$

$\boxed{\text{canonical basis}}$

② $\boxed{\text{DFT basis}}$

Let's select first K columns of N×N $\frac{1}{\sqrt{N}}\overset{\underline{\underline{F}}}{\text{DFT}}$ matrix as $\underline{\underline{E}}$.

$$\frac{1}{\sqrt{N}}\underline{\underline{F}} = \frac{1}{\sqrt{N}} \underbrace{\begin{array}{c} \begin{pmatrix} 0 & \cdots & & N-1 \end{pmatrix} \\ \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & W & W^2 & & W^{N-1} \\ 1 & W^2 & W^4 & & \\ \vdots & & & & \\ 1 & W^{N-1} & W^{2(N-1)} & & W^{(N-1)^2} \end{bmatrix} \end{array}}$$

entries of $\underline{\underline{F}}$ are $\overset{}{\underset{\downarrow\; e^{-j\frac{2\pi}{N}}}{\boxed{W}}}{}^{kl} = e^{-j\frac{2\pi}{N}\cdot k \cdot l}$

(*) Is $\underline{e_k} \perp \underline{e_l}$?

$\underline{e_k}^T \underline{e_l} = 0 \longrightarrow$ valid for real valued vector

$\Big\downarrow$ becomes

$\underline{e_k}^H \underline{e_l} = 0 \longrightarrow$ inner product definition for complex valued vectors

$$\underline{e_k}^H \underline{e_l} = \begin{bmatrix} 1 & W^{-k} & W^{-2k} & \cdots & W^{-k(N-1)} \end{bmatrix} \begin{bmatrix} 1 \\ W^{l} \\ W^{2l} \\ \vdots \\ W^{l(N-1)} \end{bmatrix} \cdot \frac{1}{N} =$$

$$\frac{1}{N}\sum_{k'=0}^{N-1} W^{-kk'} W^{lk'} = \frac{1}{N}\sum_{k'=0}^{N-1} \underbrace{\boxed{e^{j\frac{2\pi}{N}(k-l)}}}_{r}{}^{k'}$$

$$\sum_{k'=0}^{N-1} r^{k'} = \frac{1-r^N}{1-r}$$

$$= \frac{1}{N}\,\frac{1-e^{j\frac{2\pi}{\cancel{N}}(k-l)\cancel{N}}}{1-e^{j\frac{2\pi}{N}(k-l)}}$$

$$= \begin{cases} 0 & k \neq l \\ 1 & k = l \end{cases}$$

Indeed DFT basis is an orthonormal basis.

If $\underline{\underline{E}}\,\underline{\underline{E}}^H = \underline{\underline{I}}$, we have very similar forward mapping and inverse.

$$\underline{\underline{E}}^H \xrightarrow{} \underline{\underline{E}}^{-1}$$

→ Hermitian becomes the inverse, there's only a sign change.

Question

Given an arbitrary basis, how can I find an orthonormal matrix spanning the same space?

Answer: Gram-Schmidt Operation

Given $\underline{a_1}, \dots, \underline{a_K}$; we need $\underline{e_1}, \dots, \underline{e_K}$ s.t. $\underline{e_k}^T \underline{e_l} = \delta[k-l]$

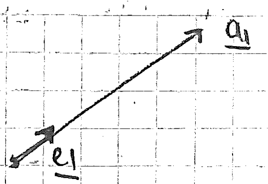
Step 1 $\underline{e_1} = \frac{\underline{a_1}}{\|\underline{a_1}\|}$

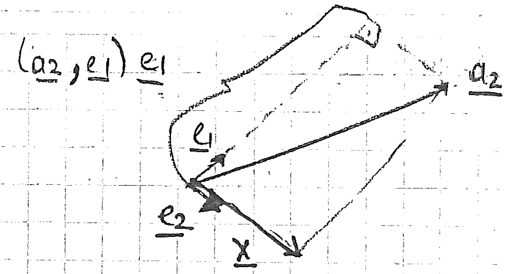
Step 2 $\underline{x} = \underline{a_2} - (\underline{a_2}, \underline{e_1})\underline{e_1}$; $\underline{e_2} = \frac{\underline{x}}{\|\underline{x}\|}$

Step 3 $\underline{x} = \underline{a_3} - (\underline{a_3}, \underline{e_1})\underline{e_1} - (\underline{a_3}, \underline{e_2})\underline{e_2}$; $\underline{e_3} = \frac{\underline{x}}{\|\underline{x}\|}$

⋮

Step K $\underline{x} = \underline{a_K} - \sum_{k=1}^{K-1} \underline{e_k}(\underline{a_K}, \underline{e_k})$; $\underline{e_K} = \frac{\underline{x}}{\|\underline{x}\|}$





orthogonality → $(\underline{x},\underline{e_1}) = (\underline{a_2},\underline{e_1}) - \overbrace{(\underline{a_2},\underline{e_1})}^{\text{scalar}}\underbrace{(\underline{e_1},\underline{e_1})}_{1} = 0$

$\underline{x}$ is orthogonal to $\underline{e_1}$.

Note

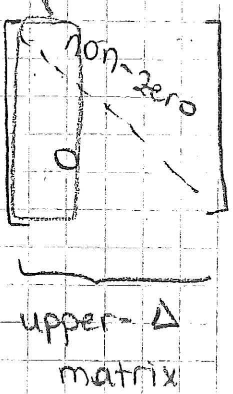
$\underline{e_k}$ is a linear combination of $\underbrace{\underline{a_1}, \underline{a_2}, \dots, \underline{a_k}}_{\text{k-vectors}}$

(vice-versa)

$\underline{\underline{A}} = [\underline{a_1} \;\; \underline{a_2} \; \dots \; \underline{a_k}]$

$\underline{\underline{E}} = [\underline{e_1} \;\; \underline{e_2} \; \dots \; \underline{e_k}]$

$[\underline{a_1} \;\; \underline{a_2} \; \dots \; \underline{a_k}] = [\underline{e_1} \;\; \underline{e_2} \; \dots \; \underline{e_k}]$



$\underbrace{\underline{\underline{A}} = \underline{\underline{Q}}\;\underline{\underline{R}}}_{\text{Q R decomposition of } \underline{A} \text{ matrix}}$

$\underline{\underline{Q}}$ — orthonormal /orthogonal matrix

$\underline{\underline{R}}$ — upper triangle matrix

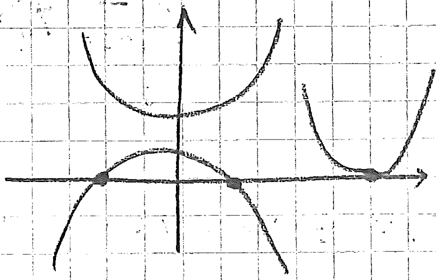
L7

26.10.2020

Positive Definite Matrices :

Assume that we have a quadratic with one dependent variable

Ex: $J_x(x) = \underset{a}{1}x^2 - \underset{b}{4}x + \underset{c}{9}$



concave up/down

- 2 roots
- single root
- no roots

Remember, sign of a coefficient immediately gives away whether quadratic is U (concave up) or ∩ (concave down)

Let's find extrema (minimum / maximum) of $J_x(x)$

$$J'_x(x) = \frac{d}{dx} J_x(x) = 2x - 4 = 0$$

$$\downarrow$$

$$x_{OPT} = 2 \longrightarrow \text{an extremum}$$

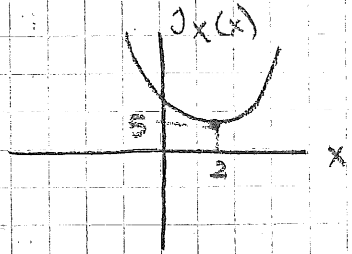
$$J''_x(x) = \frac{d^2}{dx^2} J_x(x) = 2 > 0 \longrightarrow x_{OPT} = 2 \text{ corresponds to a minimum.}$$

* translation of optimum point to the origin:

$$J_u(u) = J_x(u+2)$$

$u = x-2$

$x = u+2$



u = x-2

u is the deviation (distance) from $x_{OPT} = 2$

$$J_x(u+2) = (u+2)^2 - 4(u+2) + 9$$

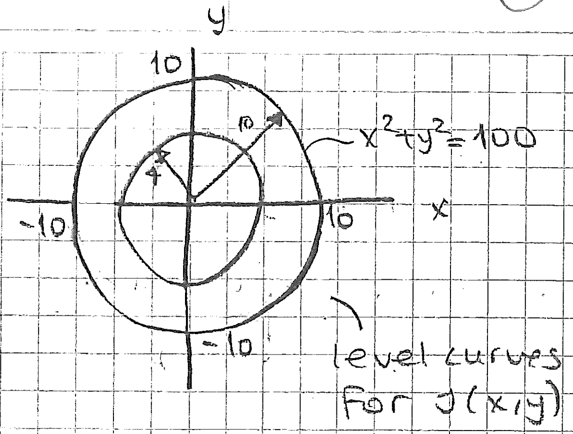
$$= u^2 + \cancel{4u} + 4 - \cancel{4u} - 8 + 9$$

$$\boxed{J_u(u) = u^2 + 5}$$ — deviation function

since $u^2 \geqslant 0$, then $u=0$ is indeed the minimum of $J_u(u)$.

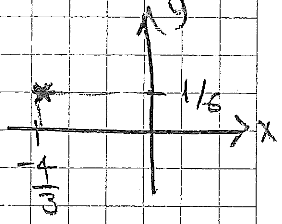
$$z = J_{circle}(x,y) = x^2 + y^2$$

locus of the points this function takes the value of 100 (cos function)



$$J(x,y) = x^2 + y^2 + 4xy + 2x + 5y + 1$$

$$\left.\begin{aligned} \frac{\partial}{\partial x} J(x,y) &= 2x + 4y + 2 = 0 \\ \frac{\partial}{\partial y} J(x,y) &= 2y + 4x + 5 = 0 \end{aligned}\right\} \quad x = -\frac{4}{3}, \; y = \frac{1}{6}$$



$$\underline{J}(\underline{x}) = \underline{x}^T \underline{\underline{A}}\, \underline{x} + \underline{x}^T \underline{b} + c$$ ← Quadratic form

$$\underline{x} = \begin{bmatrix} x \\ y \end{bmatrix}$$

$$= [x \;\; y] \begin{bmatrix} \overset{1}{a_{11}} & \overset{2}{a_{12}} \\ \underset{3}{a_{21}} & \underset{1}{a_{22}} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + [x \;\; y] \begin{bmatrix} 2 \\ 5 \end{bmatrix} + 1$$

freedom (symmetric is preferred to have orthogonal eigenvecto and real-valueo eigenvalue

$$\nabla_{\underline{x}} J(\underline{x}) = \begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \end{bmatrix} J_{\underline{x}}(\underline{x}) = (\underline{\underline{A}} + \underline{\underline{A}}^T)\underline{x} + \underline{b}$$

gradient

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}$$

$$= \begin{bmatrix} 2x a_{11} + (a_{12} + a_{21}) y \\ 2y\, a_{22} + (a_{12} + a_{21}) x \end{bmatrix} + \underline{b}$$

$$= \begin{bmatrix} 2 & 4 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} x_{OPT} \\ y_{OPT} \end{bmatrix} = \begin{bmatrix} -4/3 \\ 1/6 \end{bmatrix}$$

$$\boxed{(\underline{\underline{A}} + \underline{\underline{A}}^T)\, \underline{x}_{OPT} = -\underline{b}}$$

*) maxima or minima?

$$\underline{u} \triangleq \begin{bmatrix} x - x_{OPT} \\ y - y_{OPT} \end{bmatrix}$$

$$J_u(\underline{u}) \triangleq J_x \begin{bmatrix} u_1 + x_{OPT} \\ u_2 + y_{OPT} \end{bmatrix} = J_x(\underline{u} + \underline{x}_{OPT})$$

$$= (\underline{u} + \underline{x}_{OPT})^T \underline{\underline{A}} (\underline{u} + \underline{x}_{OPT}) + (\underline{u} + \underline{x}_{OPT})^T \underline{b} + c$$

$$= \underline{u}^T \underline{\underline{A}}\,\underline{u} + \underline{x}_{OPT}^T \underline{\underline{A}}\,\underline{u} + \underbrace{\underline{u}^T \underline{\underline{A}}\,\underline{x}_{OPT}}_{\substack{\text{scalar} \\ = \underline{x}_{OPT}^T \underline{\underline{A}}^T \underline{u}}} + \underline{x}_{OPT}^T \underline{\underline{A}}\,\underline{x}_{OPT} + \underline{u}^T \underline{b} + \underline{x}_{OPT} \underline{b} + c$$

$$\underline{x}_{OPT}^T (\underline{\underline{A}}^T + \underline{\underline{A}})\,\underline{u} = \underline{u}^T \underbrace{(\underline{\underline{A}} + \underline{\underline{A}}^T)\,\underline{x}_{OPT}}_{-\underline{b}} = -\underline{u}^T \underline{b}$$

$$\boxed{J_u(\underline{u}) = \underline{u}^T \underline{\underline{A}}\,\underline{u} + \underbrace{\underline{x}_{OPT}^T \underline{\underline{A}}\,\underline{x}_{OPT} + \underline{x}_{OPT} \underline{b} + c}_{J_x(\underline{x}_{OPT})}} \longrightarrow \text{deviation function}$$

If ≥ 0, then $\underline{x}_{OPT}$ is a minimum.

① $\underline{x}^T \underline{\underline{A}}\,\underline{x} > 0 \quad \forall \underline{x} \neq \underline{0} \longrightarrow \underline{\underline{A}}$: positive definite $(\underline{\underline{A}} > 0)$

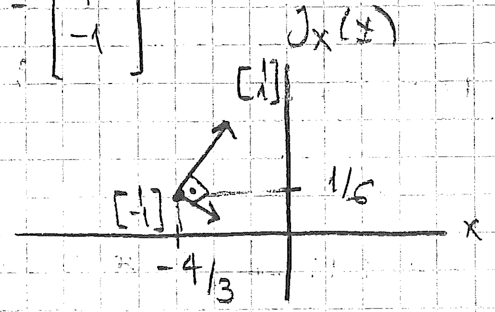
② $\underline{x}^T \underline{\underline{A}}\,\underline{x} \geq 0 \quad \forall \underline{x} \neq \underline{0} \longrightarrow \underline{\underline{A}}$: positive semi-definite $(\underline{\underline{A}} \geq 0)$

③ $-\underline{x}^T \underline{\underline{A}}\,\underline{x} > 0 \quad \forall \underline{x} \neq \underline{0} \longrightarrow \underline{\underline{A}}$: negative definite $(\underline{\underline{A}} < 0)$

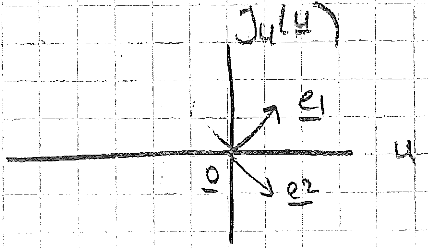
④ $-\underline{x}^T \underline{\underline{A}}\,\underline{x} \geq 0 \quad \forall \underline{x} \neq \underline{0} \longrightarrow \underline{\underline{A}}$: negative semi-definite $(\underline{\underline{A}} \leq 0)$

$$\underline{\underline{A}} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{array}{l} \nearrow \lambda = 3 \rightarrow \underline{e_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ \searrow \lambda = -1 \rightarrow \underline{e_2} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \end{array}$$

$$J_u(\underline{u}) = \underline{u}^T \underline{\underline{A}}\, \underline{u} \Big|_{\underline{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} t} = 6t^2$$



$$J_u(\underline{u}) = \underline{u}^T \underline{\underline{A}}\, \underline{u} \Big|_{\underline{u} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} t} = -2t^2$$



in $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ direction → increasing

in $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ direction → decreasing

} we see that $\underline{x}_{opt}$ is not a maxima or minima of $J_x(\underline{x})$ "saddle point"

$$J_u(\alpha \underline{e_1} + \beta \underline{e_2}) = |\alpha|^2 \|\underline{e_1}\|^2 \textcircled{\lambda_1} + |\beta|^2 \|\underline{e_2}\|^2 \textcircled{\lambda_2}$$

One of the eigenvalues is negative → trouble.

If $\underline{\underline{A}}$ matrix does not satisfy ①, ②, ③, ④,

$\underline{\underline{A}}$: indefinite matrix.

Results

① $\underline{\underline{A}} > 0$ $\longleftrightarrow$ all eigenvalues of $\underline{\underline{A}}$ should be positive

symmetric

$$\lambda_k > 0 \quad \forall k$$

Another way of checking positive definiteness is checking leading principal matrix of $\underline{\underline{A}}$.

Ex: $\begin{bmatrix} 1 & 0.1 & 3 \\ 0.1 & 4 & 6 \\ 3 & 6 & 5 \end{bmatrix} \overset{?}{>} 0$ If all D_1, D_2, D_3 are positive, then $\underline{\underline{A}} > 0$.

② $\underline{\underline{A}} \geq 0$ $\longleftrightarrow$ all eigenvalues are non-negative

$$\lambda_k \geq 0 \quad \forall k$$

Checking leading principal minors is not sufficient for deciding $\underline{\underline{A}} \geq 0$ or not.

$\boxed{\text{Note}}$ What happens when $\underline{\underline{A}}$ is not symmetric?

$$\underline{\underline{A}} = \underbrace{\frac{\underline{\underline{A}}+\underline{\underline{A}}^T}{2}}_{\substack{\text{symmetric} \\ \underline{\underline{A}}_{sym}}} + \underbrace{\frac{\underline{\underline{A}}-\underline{\underline{A}}^T}{2}}_{\substack{\text{anti-symmetric} \\ \underline{\underline{A}}_{asym}}}$$

$$\underline{x}^T\underline{\underline{A}}\,\underline{x} = \underline{x}^T\underline{\underline{A}}_{sym}\underline{x} + \cancelto{0}{\underbrace{\underline{x}^T\underline{\underline{A}}_{asym}\underline{x}}_{}}$$

$$\text{scalar} \rightarrow (\underline{x}^T\underline{\underline{A}}_{asym}\underline{x})^T = \overset{\text{scalar}}{\underline{x}^T\underline{\underline{A}}_{asym}\underline{x}}$$

$$\underline{x}^T\underbrace{\underline{\underline{A}}^T_{asym}}_{-\underline{\underline{A}}_{asym}}\underline{x} = \underline{x}^T\underline{\underline{A}}_{asym}\underline{x}$$

$$\underline{x}^T\underline{\underline{A}}_{asym}\underline{x} = 0$$

So, for the sake of quadratic form $\underline{x}^T \underline{\underline{A}}\, \underline{x}$, only

$\underline{\underline{A}}_{sym} = \frac{\underline{\underline{A}} + \underline{\underline{A}}^T}{2}$ is the matrix important for calculation.

Over-Determined Equation Systems

Let's assume, I have N equation with K unknowns, and N>K.

$$\begin{array}{r}1^{st}\,\text{eqn}\rightarrow\\ \\ \\ N^{th}\,\text{eqn}\rightarrow\end{array}\underbrace{\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1K}\\ a_{21} & a_{22} & \cdots & a_{2K}\\ \vdots & & & \\ a_{N1} & a_{N2} & \cdots & a_{NK}\end{bmatrix}}_{\text{tall-matrix}}\begin{bmatrix}x_1\\ \vdots\\ \\ x_K\end{bmatrix} = \begin{bmatrix}b_1\\ \vdots\\ \\ b_N\end{bmatrix}$$

↳ transpose: fat / short matrix

$$\underbrace{J(\underline{x})}_{\text{error vector norm square is minimized}} = \|\underline{\underline{A}}\,\underline{x} - \underline{b}\|^2 = \underbrace{(\underline{\underline{A}}\,\underline{x} - \underline{b})^T}_{\underline{x}^T\underline{\underline{A}}^T - \underline{b}^T}(\underline{\underline{A}}\,\underline{x} - \underline{b})$$

$$= \underline{x}^T \underline{\underline{A}}^T \underline{\underline{A}}\,\underline{x} - \underbrace{\underline{x}^T \underline{\underline{A}}^T \underline{b}}_{\text{scalar}} - \underline{b}^T \underline{\underline{A}}\,\underline{x} + \underline{b}^T \underline{b}$$

$$= \underline{b}^T \underline{\underline{A}}\,\underline{x} = \underline{x}^T \underline{\underline{A}}^T \underline{b}$$

If $M>0$, → minima

$$= \underline{x}^T \overset{\underline{\underline{M}}}{\underline{\underline{A}}^T \underline{\underline{A}}}\,\underline{x} - 2\underline{x}^T \underline{\underline{A}}^T \underline{b} + \underline{b}^T \underline{b}$$

$$\nabla_{\underline{x}} J(\underline{x}) = \underline{0} \longrightarrow \underbrace{(\underline{\underline{M}}^T + \underline{\underline{M}})}_{\cancel{2}\underline{\underline{A}}^T\underline{\underline{A}}}\underline{x} - \cancel{2}\underline{\underline{A}}^T \underline{b} = 0$$

$$\underline{\underline{A}}^T \underline{\underline{A}}\,\underline{x} = \underline{\underline{A}}^T \underline{b} \longrightarrow \boxed{\underline{x}_{LS} = (\underline{\underline{A}}^T \underline{\underline{A}})^{-1} \underline{\underline{A}}^T \underline{b}}$$

$$\underline{\underline{A}}\,\underline{x}_{LS} = \underbrace{\underline{\underline{A}}(\underline{\underline{A}}^T \underline{\underline{A}})^{-1} \underline{\underline{A}}^T}_{\text{projection matrix}} \underline{b}$$

Let's check $\underline{\underline{M}} > 0$ $(\underline{\underline{A}}^T \underline{\underline{A}} > 0)$ or not.

$$\underbrace{\underline{x}^T \underline{\underline{A}}^T \underline{\underline{A}}\, \underline{x}} \overset{?}{>} 0 \quad \forall \underline{x} \neq \underline{0}$$

$$\|\underline{\underline{A}}\underline{x}\|^2 \geq 0$$

and

$$\|\underline{\underline{A}}\,\underline{x}\| = 0 \iff \underset{(\underline{x} \neq \underline{0})}{\underline{\underline{A}}\,\underline{x} = \underline{0}} \iff \underline{\underline{A}} \text{ has a non-trivial null space}$$

$\Updownarrow$

Columns of $\underline{\underline{A}}$ are not linearly independent.

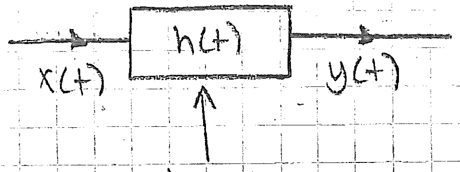
So, $\underline{\underline{A}}^T \underline{\underline{A}} \geq 0$ in general

and

if $\underline{\underline{A}}$ is full-column rank, then $\underline{\underline{A}}^T \underline{\underline{A}} > 0$.

Review of (Some) DSP Topics

Analog ⟶



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

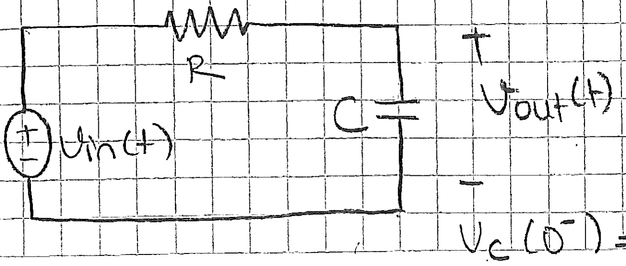
impulse response

↳ no initial conditions / energy (at rest)

LTI (linear, time-invarient system)

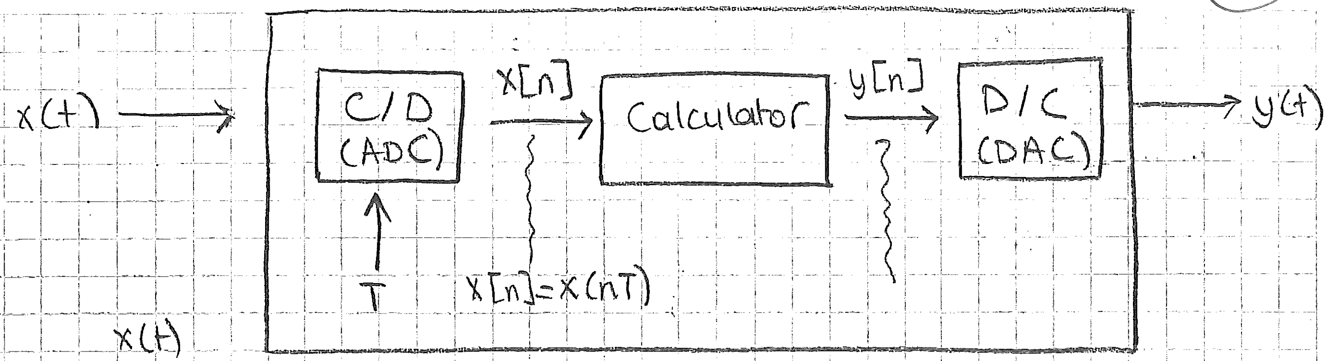
- superposition of x (integral)
- additivity ✓
- ?→ homogenity ✓ (multiplying by x)

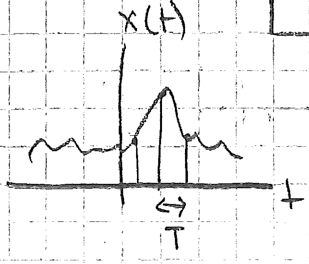
apply the input at a later time, the response is shifted to the next application time.



$$V_{out}(t) = \int_{-\infty}^{+\infty} h(\tau)\, v_{in}(t-\tau)\, d\tau$$

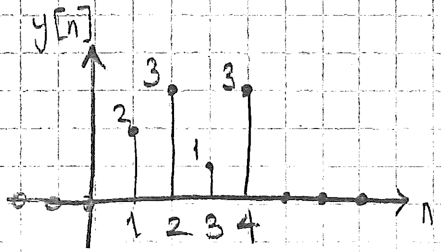
$V_c(0^-) = 0$



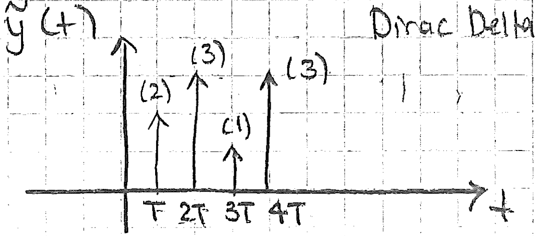


$$y[n] = \frac{x[n] + x[n-1]}{2}$$

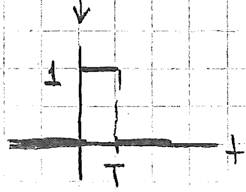
analog → Fixed quantities, not flexible while processing

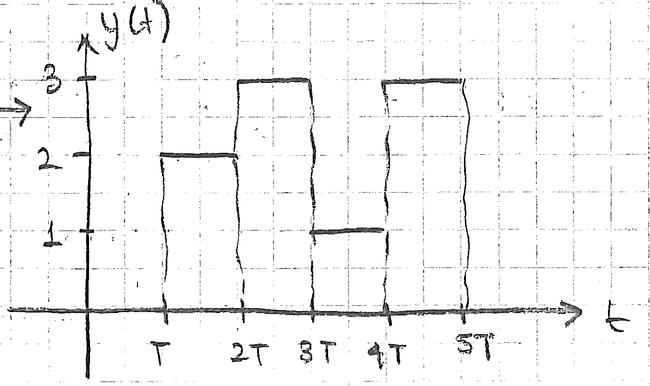


$y(t) = ?$

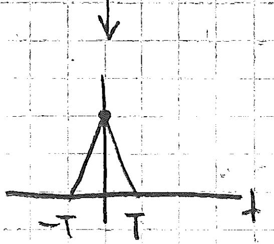


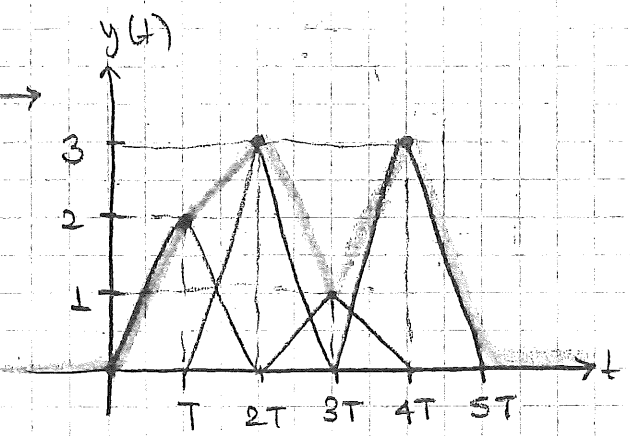
⊛ $\tilde{y}(t) * h_{ZOH}(t) \longrightarrow$





⊛ $\tilde{y}(t) * h_{LI}(t) \longrightarrow$

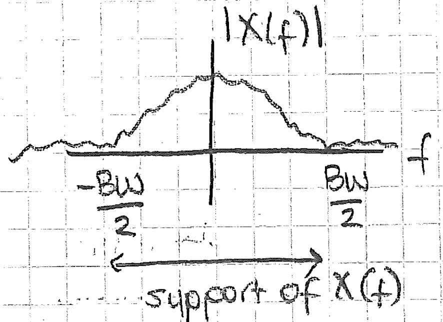




If $x(t)$ is Bandlimited, that is $\underset{\mathcal{F}\{x(t)\}}{X(f)}=0$ for $|f| > \underset{\text{bandwidth}}{BW}/2$,

then $x[n] = x(n\underset{\text{sampling period}}{T})$.

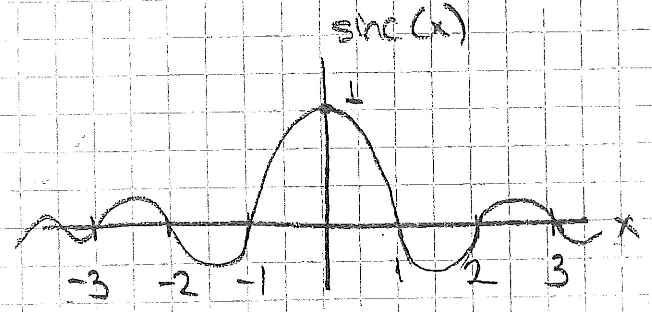
* assuming $T < \frac{1}{BW}$ OR $\underset{\text{sampling frequency}}{\boxed{\frac{1}{T}}} > \boxed{BW}$ — support of $X(f)$ → Nyquist Rate.



$$x(t) = \sum_{n=-\infty}^{+\infty} x[n] \,\text{sinc}\left(\frac{t}{T} - n\right) \longrightarrow \text{sinc interpolation}$$

↓

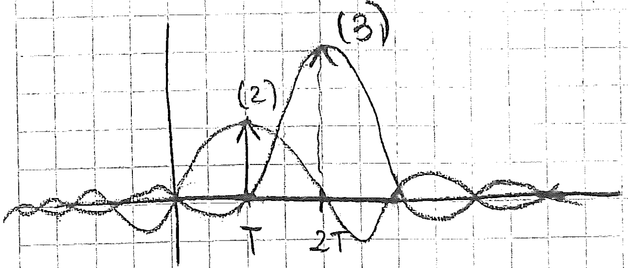
sampling theorem.



$$\text{sinc}(x) \triangleq \frac{\sin(\pi x)}{\pi x}$$

$$x(\underset{\text{integer}}{m}T) = \sum_{n=-\infty}^{+\infty} x[n] \underbrace{\text{sinc}\left(\frac{m\cancel{T}}{\cancel{T}} - n\right)} = x[m]$$

$$\delta[m-n] = \begin{cases} 1 , & m=n \\ 0 , & m\neq n \end{cases} \quad \text{Kronecker Delta}$$

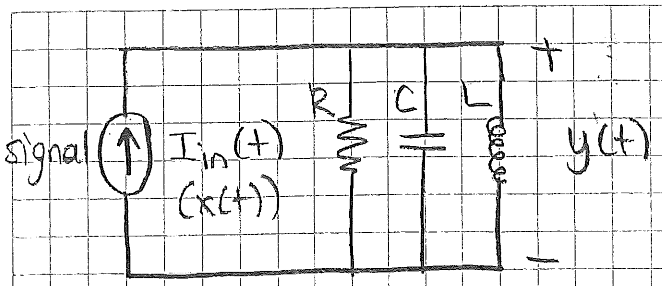


Assume $x(t)$ is bandlimited and $x[n]$ is the samples collected above Nyquist Rate.

$x[n] \longleftrightarrow x(t)$

↓

(sampling frequency is large enough
OR
sampling time is small enough)



Center frequency, passband

Components → aging, etc
(practical problems)

If $x(t)$ is bandlimited, clearly, $y(t)$ is bandlimited.

Impulse Invariance

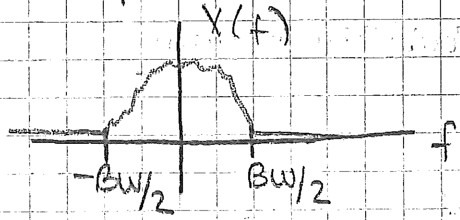
Oppenheim ~~→ Discrete-~~
time processing of
analog signals ★

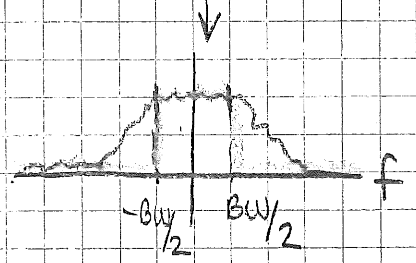
$h_c(t)$: impulse response

$h[n] = T \cdot h_c(nT)$

$Y(f) = H(f) \cdot X(f)$

→ finite support, bandlimited





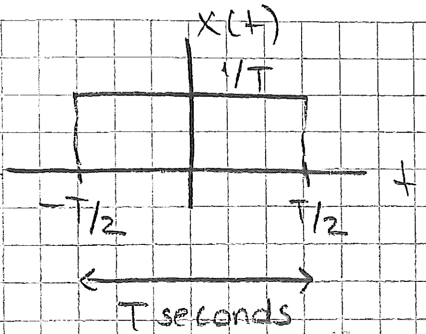
Fourier Transforms

$$X(f) = \int_{-\infty}^{+\infty} x(t)\, e^{-j2\pi f t}\, dt$$

$$x(t) = \int_{-\infty}^{+\infty} X(f)\, e^{j2\pi f t}\, df$$

$x(t) \underset{F}{\longleftrightarrow} X(f)$

ex: $F\{\underbrace{rect(\frac{t}{T})}_{x(t)}\}$



$$X(f) = \int_{-T/2}^{+T/2} \frac{1}{T} \underbrace{e^{-j2\pi ft}} dt$$

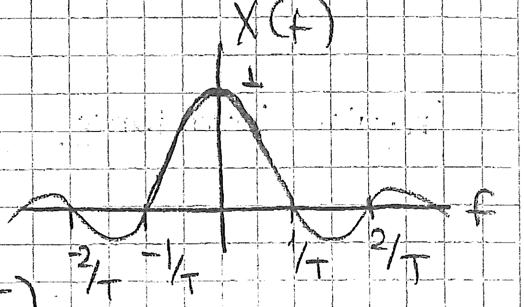
$$\downarrow$$

$$\cos(2\pi ft) - j\cancel{\sin(2\pi ft)}$$

odd

$$= \frac{\cancel{2}}{\cancel{2}\pi f T} \sin(2\pi f t) \Big|_0^{T/2}$$

$$= \frac{\sin(\cancel{2}\pi f \frac{T}{\cancel{2}})}{\pi f T} = sinc(fT)$$



Probability Review

Terminology probabilistic (random) experiment → no deterministic mapping

sample space	outcome	event	probability
throwing a die	{1}	A	a mapping from
6 faces	{2}	• a subset of S	sets(events) to
S = {{1}, {2}, {3},	{3}	• φ, {2, 4, 6}, S, ...	[0, 1].
{4}, {5}, {6}}	{4}	• a subset not	P{A} : Event space
all possible outcomes	{5}	equal to S is	↓ ↓
(universal set)	{6}	called a proper	∈ event space [0, 1]
		subset	

Kolmogorov's Axiomatic Definition:

① $P\{A\} \geq 0$

② $P\{S\} = 1$

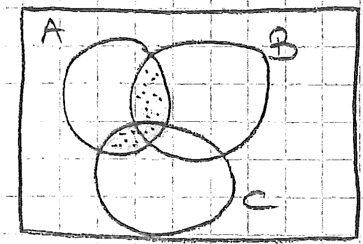
③ $P\{A \cup B\} = P\{A\} + P\{B\}$, provided that $A \cap B = \emptyset$

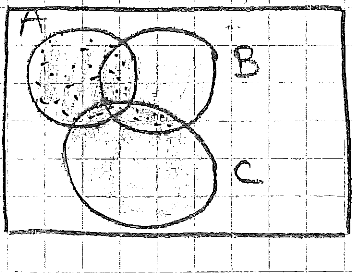
Set Operations: $A \cup B$, $A \cap B$, $\overline{B}$

DeMorgan's Laws:

① $\underline{A \cap (B \cup C)} = (A \cap B) \cup (A \cap C)$

② $\underline{A \cup (B \cap C)} = (A \cup B) \cap (A \cup C)$

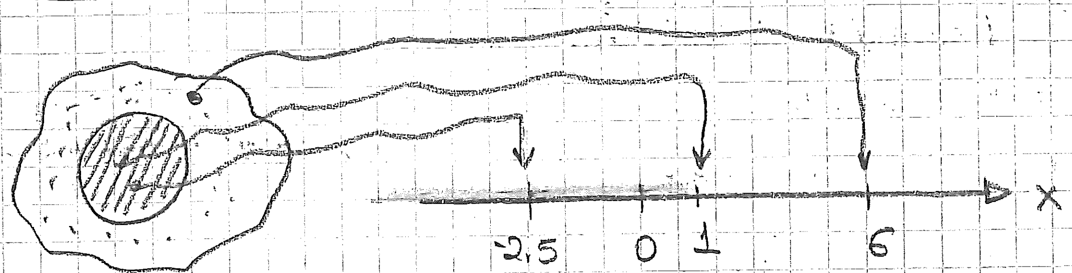




Note Sigma-Algebra

- For finite sample space such as $S = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$ the event space can be defined as all subsets of S. That is, event space consists of $2^6 = 64$ events.
- For countably infinite and uncountably infinite sample spaces, events should be defined properly such that if A and B are events, then $A \cup B$, $A \cap B$, $\overline{B}$ should also be events.
- For our purposes, the events of intervals on the real line such as $(-\infty, a]$ will be utilized and it is possible to show that all such intervals on real-line can be written as union, intersection and complement operations.

Random Variable

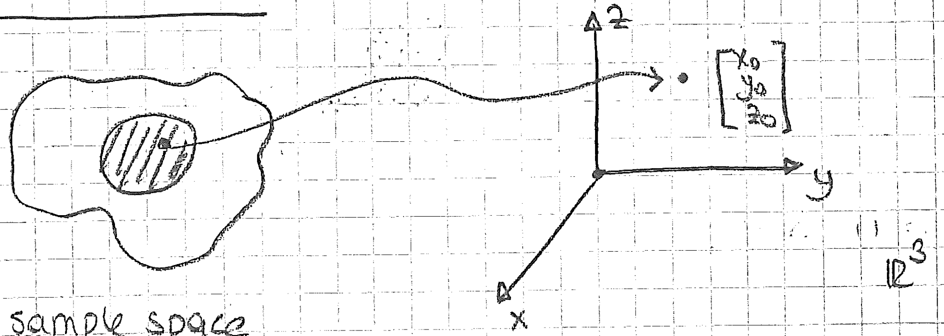


sample space

Random variables map outcomes to real numbers.

Approach: We assume the mapping is already defined in many cases and use definitions.

Random Vector



sample space

For random variables, events are union / intersection of intervals in the form $(-\infty, a]$.

$$P\{\underline{X} \in (-\infty, a]\} = P\{\underbrace{\underline{X} \leq a}_{(x\,:\, x \leq a)}\} = P\{\underset{\substack{\text{capital}\\\text{random}\\\text{variable}}}{\underline{X}} \leq \underset{\text{scalar}}{x}\}$$

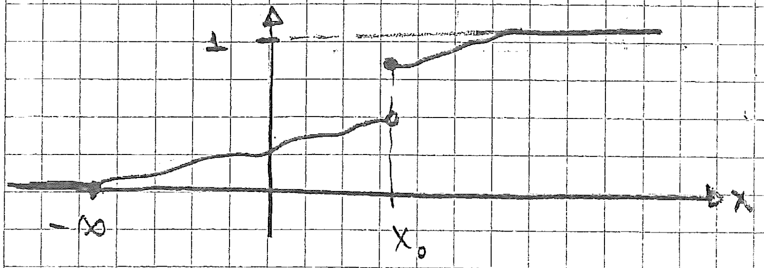
in $\underline{X}$, x is mapped to a,

y is mapped to b, ...

Instead → $\underline{X}$ is mapped to x.

c.d.f. (cumulative density function)

$$F_X(x) = P\{X \leq x\}$$



→ monotonically non-decreasing

→ right continuous

→ $F_X(-\infty)=0$, $F_X(+\infty)=1$

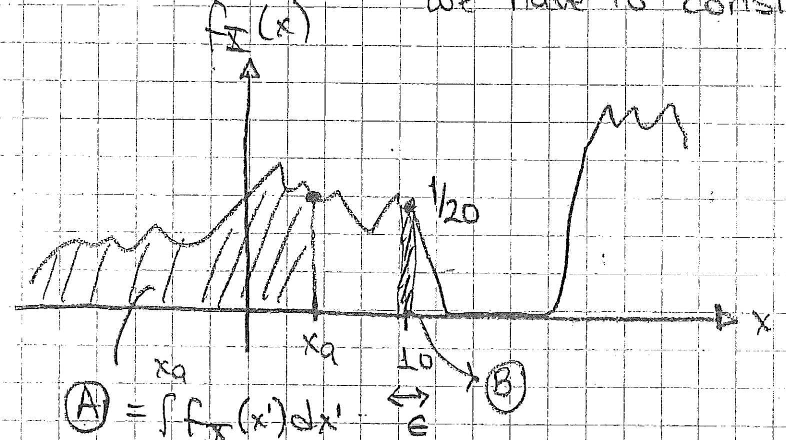
p.d.f. (probability density function)

$$f_X(x) = \frac{d}{dx} F_X(x) \longleftrightarrow F_X(x) = \int_{-\infty}^{x} f_X(x')\,dx'$$

Then, $P\{X \in Ⓐ\} = \int\limits_{x \in Ⓐ} f_X(x')\,dx' = \int_{-\infty}^{5} f_X(x')\,dx' + \int_{10}^{12} f_X(x')\,dx'$

Ⓐ : $\underbrace{(-\infty,5)}$ $\cup$ $\underbrace{(10,12]}_{(-\infty,12]-(-\infty,10]}$

→ assume c.d.f. is continuous here, otherwise we have to consider dirac delta.



Ⓐ $= \int_{-\infty}^{x_a} f_X(x')\,dx'$

$= F_X(x_a)$

* 1/20 is the value of density function at x=10, not the probability. integrate around x=10 in a short interval ϵ

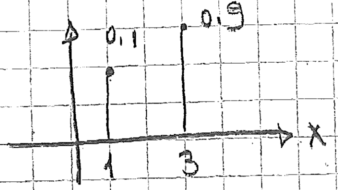
Ⓑ $\underbrace{\approx f_X(10)\cdot\epsilon}_{\text{probability}}$

units → (example) probability : unitless

x, dx : meters

$f_X(x')$: 1/meters.

⊛ discrete case



→ probability mass function

$P\{X=1\}=0.1$
$P\{X=3\}=0.9$ } easier than density functions

Independence

$P\{A\cap B\}=P\{A\}\cdot P\{B\}$ → Events A and B are independent

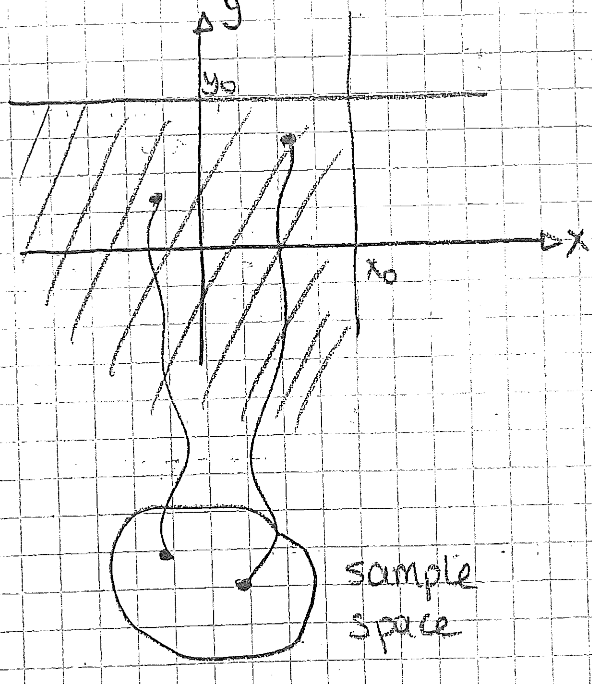
Independence for Random Variables

$$P\{X\leq x_0 , Y\leq y_0\} = P\{X\leq x_0\}\cdot P\{Y\leq y_0\} \quad ⊛$$

↓ intersection

$(X\leq x_0)\cap(Y\leq y_0)$

If ⊛ is satisfied for all $x_0, y_0 \in \mathbb{R}$, then X and Y are independent.



$$F_{X,Y}(x,y)=P\{X\leq x, Y\leq y\}$$

$$f_{X,Y}(x,y)=\frac{\partial^2}{\partial x\partial y}F_{X,Y}(x,y)$$

$$\to = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(x',y')\,dy'\,dx'$$

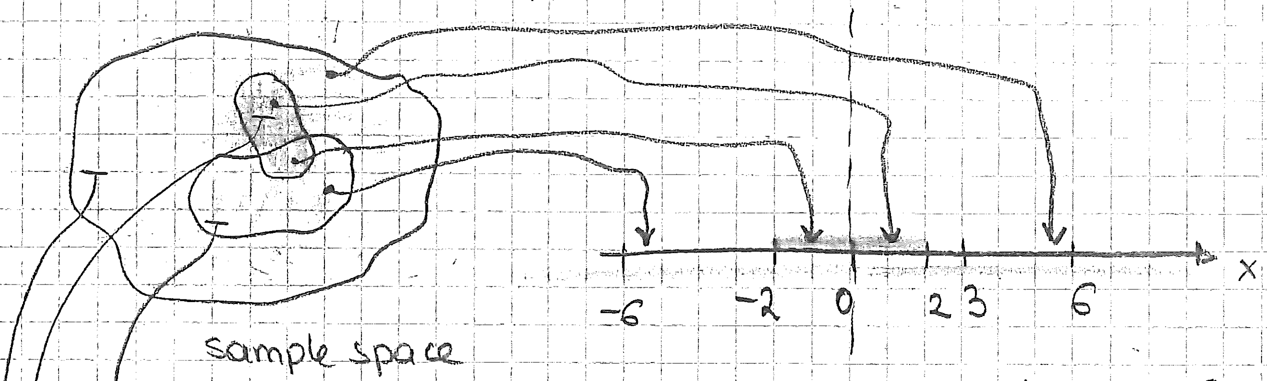
units (example) → probability: unitless

$x, y, dx', dy', \partial x, \partial y$: meters

$f_{X,Y}(x,y)$: $1/\text{meters}^2$

Conditional Probability:

P(A|B) : Probability of outcome is in A given that outcome is known to be in B.
Probability that "A happens" given that "B has happened".



gives negative outcomes Ⓝ

gives outcomes between [-2,2] Ⓨ

gives positive outcomes Ⓟ

$$P\{\underbrace{\overbrace{x\in[-2,2]}^{A}}_{Ⓨ} \mid \underbrace{\overbrace{x\geqslant 0}^{B}}_{Ⓟ}\}$$

after conditioning operation, Ⓝ gets eliminated. (not contributing to the probability)

Our sample space is reduced to Ⓟ

(without conditioning: $\frac{Ⓨ}{Ⓢ}$, with conditioning: $\frac{Ⓨ - Ⓝ}{Ⓟ} = \frac{Ⓨ \cap Ⓟ}{Ⓟ}$)

$$P\{A|B\} = \frac{P\{A\cap B\}}{P\{B\}} = \frac{P\{x\in[-2,2] \cap x\geqslant 0\}}{P\{x\geqslant 0\}} = \frac{P\{x\in[0,2]\}}{P\{x\in[0,\infty)\}}$$

If we extend $2\rightarrow\infty$ $(x\rightarrow\infty)$, the upper part becomes $\{x\in[0,\infty)\}$

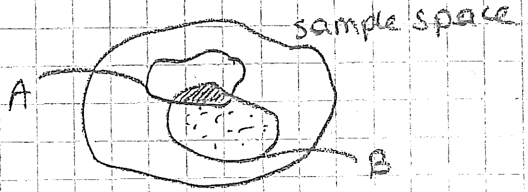
$x\in[-2|x] \mid x\geqslant 0 \longrightarrow$ has a valid c.d.f.

Bayes' Theorem

$$\left.\begin{aligned} P(A|B) &\triangleq \frac{P(A\cap B)}{P(B)} \\ P(B|A) &\triangleq \frac{P(A\cap B)}{P(A)} \end{aligned}\right\} \quad \boxed{P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)}}$$

$\boxed{2.11.2000}$

Conditional Probability $\triangleq \frac{P(A\cap B)}{P(B)}$



Bayes' Theorem $P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$

as soon as I know B has happened, my sample space is reduced to B.

2 Random Variables X and Y

Marginalization → $f_X(x_0) = \int_{-\infty}^{+\infty} f_{X,Y}(x_0,y)\,dy$ — joint probability density function

(marginalise/integrate over y)

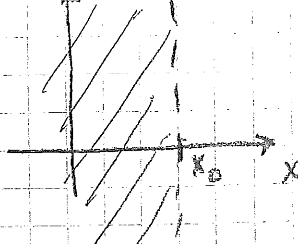
for fixed x_0, add all y probabilities in the interval ϵ.

$\epsilon \cdot f_{X,Y}(x_0,y) \sim$ probability

$$\underbrace{P\{X \leq x_0, Y < \infty\}}_{(X\leq x_0)\,\cap\,\underbrace{(Y<\infty)}_{\text{always satisfied}}} = P\{X \leq x_0\} = F_X(x_0)$$

volume

$$= \int_{-\infty}^{x_0}\int_{-\infty}^{+\infty} f_{X,Y}(x',y')\,dy'\,dx'$$



$$F_X(x_0) = \int_{-\infty}^{x_0} \underbrace{\left(\int_{-\infty}^{+\infty} f_{X,Y}(x',y')\,dy' \right)}_{g(x')} dx'$$

$$\frac{d}{dx_0} F_X(x_0) = f_X(x_0) = g(x_0) - g(-\infty)$$

$$\frac{d}{dx_0}\left(\int_{-\infty}^{x_0} g(x')\,dx'\right) = \frac{d}{dx_0}\left(G(x_0) - \cancel{G(-\infty)}^{\,0}\right) = g(x_0)$$

$$g(x_0) = f_X(x_0) = \int_{-\infty}^{+\infty} f_{X,Y}(x_0, y')\,dy'$$

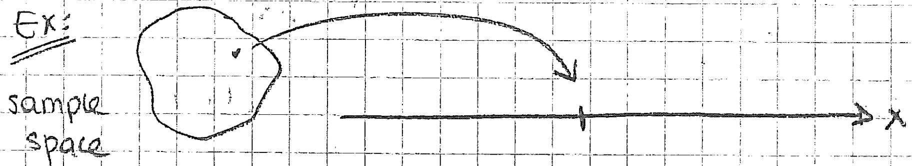
Total Probability Theorem:

$$P(X=k) = \sum_{y=-\infty}^{+\infty} \underbrace{P(X=k \mid Y=y)}_{\text{conditional probability for Y}} \underbrace{P(Y=y)}_{\text{marginal probability for Y}} \longrightarrow \text{marginalization}$$

(fix y, calculate probabilities, sum)

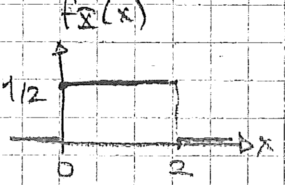
$$f_X(x) = \int_{-\infty}^{+\infty} \underbrace{f_{X|Y}(x|y)\, f_Y(y)}_{\text{joint density}}\, dy$$

Ex:



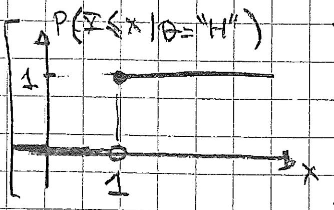
deterministic

We throw an unbiased coin. If "Heads" shows up, then $X = 1$; else, X is uniformly distributed in $[0,2]$. Find the density of X.

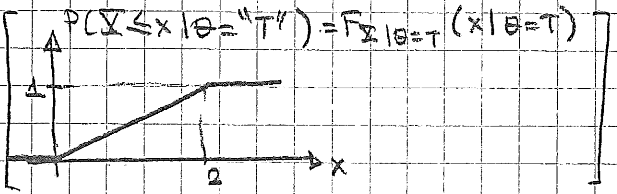


$$P(X \le x) = P(X \le x \mid \overset{\text{coin throw}}{\theta} = \text{"H"})\, P(\text{"H"}) + P(X \le x \mid \theta = \text{"T"})\, P(\text{"T"})$$

$=$

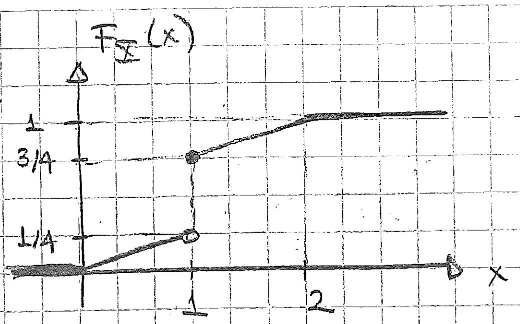


$\cdot \frac{1}{2} +$

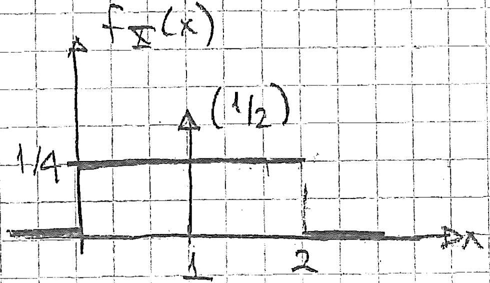


$\cdot \frac{1}{2} =$

=



$\frac{d}{dx}$



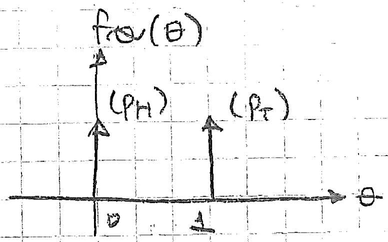
say "H" → $\Theta=0$

θ = "T" → $\Theta=1$ } random variable

$$f_X(x) = \int_{-\infty}^{+\infty} f_X(x\,|\,\Theta=\theta)\,\underbrace{f_\Theta(\theta)}\,d\theta$$

$$\downarrow$$

$$p_H\,\delta(\theta) + p_T\,\delta(\theta-1)$$

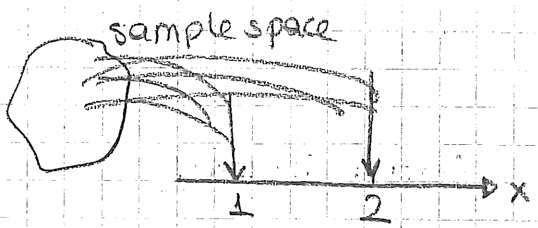


$$= p_H \underbrace{f_X(x|\Theta=0)}_{1\cdot\delta(x-1)} + p_T \underbrace{f_X(x|\Theta=1)}_{\text{unif }(x \text{ in } [0,2])}$$

→ latent (hidden) variable

⇒ mixture of 2 distributions

Expectation Operation



$$E_{\mathbb{X}}\{\mathbb{X}\} = \int_{-\infty}^{+\infty} x \cdot \underbrace{f_{\mathbb{X}}(x)}_{\downarrow} dx$$

i.e. $p_H\,\delta(x-1) + p_T\,\delta(x-2)$

$$= p_H 1 + p_T 2$$

$$\underbrace{\frac{1}{\#\text{trials}}\left(\sum_{k=1}^{\#\text{trials}} \boxed{x_k}\right)}_{\text{empirical average of trials}} = \boxed{\frac{1\cdot\#(x_k=1)}{\#\text{trials}}} + \boxed{\frac{2\cdot\#(x_k=2)}{\#\text{trials}}}$$

result at kth trial

as # trials ↑, this approaches to p_H.

as # trials ↑, this approaches to p_T.

As # trials → ∞ (independent trials) ⇒ empirical average ⇒ $E_{\mathbb{X}}\{\mathbb{X}\}$

$$E\{g(\mathbb{X})\} = \int_{-\infty}^{+\infty} g(\mathbb{X})\, f_{\mathbb{X}}(x)\, dx$$

Moments

$E\{X^k\} = m_k$ — kth moment of r.v. X

Central Moments

$$E\{\underbrace{(X - \underbrace{E\{X\}}_{m_1=\bar{X}})}_{\text{r.v. with 0-mean}}{}^k\}$$ — kth central moment of r.v. X

$m_1 = \bar{X} = E\{X\}$ is called the mean of r.v. X.

$\sigma^2 = E\{(X-\bar{X})^2\}$ is called the variance of r.v. X.

(σ: standard deviation)

Moment Generating Functions

$$\Phi(s) = E_X\{\underbrace{e^{sX}}_{g(X)}\} = \int_{-\infty}^{+\infty} e^{sx} f_X(x)\,dx$$

$$= \mathcal{L}\{f_X(x)\}(-s)$$

$$\frac{d^n}{ds^n}\Phi(s)\Big|_{s=0} = E_X\{X^n e^{sX}\}\Big|_{s=0} = E_X\{X^n\}$$ — nth moment (m_n)

Then, by Taylor series of $\Phi(s)$ at $s_0=0$, we can write

$$\Phi(s) = \sum_{n=0}^{\infty} \underbrace{\Phi^{(n)}(s_0=0)}_{m_n} \frac{(s-s_0)^n}{n!} = \sum_{n=0}^{\infty} m_n \frac{s^n}{n!}$$

So, the knowledge of all moments m_n, $n=\{1,2,\ldots\}$ is equivalent to the knowledge of $\Phi(s)$, which is equivalent to the knowledge of density.

moments $\rightarrow \Phi(s) \xrightarrow{\mathcal{L}^{-1}\{\}} f_X(x)$ (infinitely many moments are needed)

If we know moments partially, say first 2 moments, then such knowledge is called moment characterization or <u>partial</u> characterization of p.d.f.

Conditional Expectation

$$E_{\mathbf{X}|Y=y}\{\mathbf{X}|Y=y\} = \int_{-\infty}^{+\infty} x \cdot f_{\mathbf{X}|Y=y}(x|Y=y)\,dx$$

$$E_{\mathbf{X}|Y=y}\{g(\mathbf{X},Y)|Y=y_0\} = \underbrace{\int_{-\infty}^{+\infty} g(x,y_0)\, f_{\mathbf{X}|Y=y_0}(\mathbf{X}=x|Y=y_0)\,dx}_{\text{function of } y}$$

Iterated Expectation

$$E_Y\{\underbrace{E_{\mathbf{X}|Y}\{g(\mathbf{X},Y|Y=y)\}}_{\psi(y)}\} = E_{\mathbf{X},Y}\{g(\mathbf{X},Y)\}$$

$$= \int_{-\infty}^{+\infty} \psi(y)\, f_Y(y)\,dy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(\mathbf{X},Y)\underbrace{f_{\mathbf{X}|Y=y}(x|Y=y)\,dx\,f_Y(y)}_{f_{\mathbf{X},Y}(x,y)}\,dy$$

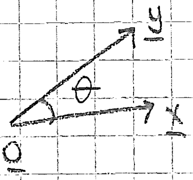
$$= \int_{-\infty}^{+\infty} g(\mathbf{X},Y)\, f_{\mathbf{X},Y}(x,y)\,dx\,dy$$

L11

Question

How to quantify the "similarity" of two random variables?

Remember that,



$$(\underline{x}, \underline{y}) = \underset{\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}}{\underline{x}^T} \underline{y} = \sum_{k=1}^{N} x_k y_k$$

$$\cos\theta = \frac{\underline{x}^T \underline{y}}{\|\underline{x}\| . \|\underline{y}\|} = \frac{\sum_{k=1}^{N} x_k \cdot y_k}{\sqrt{\sum_{k=1}^{N} x_k^2} \sqrt{\sum_{k=1}^{N} y_k^2}}$$

Assume, for now, $\underline{X}$ and Y are zero mean random variables.

$$r_{XY} \triangleq \frac{E\{\underline{X} Y\}}{\sqrt{E\{X^2\} E\{Y^2\}}}$$ ← r_{XY} definition for $E\{X\} = E\{Y\} = 0$

correlation coefficient

$$r_{XY} = \frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy\, f_{XY}(x,y)\, dx\, dy}{\sqrt{\int_{-\infty}^{+\infty} x^2 \underbrace{f_X(x)}_{\int_{-\infty}^{+\infty} f_{XY}(x,y)\, dy} dx} \sqrt{\int_{-\infty}^{+\infty} y^2 \underbrace{f_Y(y)}_{\int_{-\infty}^{+\infty} f_{XY}(x,y)\, dx} dy}} = \frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy \overset{\text{weight}}{\boxed{f_{XY}(x,y)}} dx\, dy}{\sqrt{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^2 \underset{\text{weight}}{\boxed{f_{XY}(x,y)}} dy\, dx} \sqrt{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y^2 \underset{\text{weight}}{\boxed{f_{XY}(x,y)}} dx\, dy}}$$

Properties of Correlation coefficient

① $|r_{xy}| \leq 1$

② $|r_{xy}| = 1 \longleftrightarrow Y = \underset{\text{scalar}}{\alpha} X$ (X, Y are aligned)

Proof ①

$$E\{(X+\lambda Y)^2\} \geq 0$$

$$\lambda^2 \underbrace{E\{Y^2\}}_{a} + \underbrace{2\lambda E\{XY\}}_{b} + \underbrace{E\{X^2\}}_{c} \geq 0$$

$$p(\lambda) = a\lambda^2 + b\lambda + c \geq 0$$

no 2 real roots $\longrightarrow \Delta \leq 0$

$$\cancel{4}(E\{XY\})^2 - \cancel{4}E\{Y^2\}E\{X^2\} \leq 0$$

$$\frac{|E\{XY\}|}{\sqrt{E\{X^2\}E\{Y^2\}}} \leq 1 \longrightarrow |r_{xy}| \leq 1$$

Proof ②

$$\Delta = 0 \longrightarrow \quad p(\lambda) = (\lambda - \lambda_*)^2 = E\{(X+\lambda_* Y)^2\} = 0$$

$$X + \lambda_* Y = 0 \longrightarrow X = -\underbrace{\lambda_*}_{\text{scalar}} Y$$

So, if $r_{xy} = \pm 1 \longrightarrow Y = aX$

$Y = aX \longrightarrow |r_{XY}| = 1$

$$r_{XY} = \frac{E\{X^2\}\, a}{\sqrt{a^2 E\{X^2\} E\{X^2\}}} = \frac{a}{|a|} = \text{sign}(a) = \pm 1$$

X and Y are two random variables.

r_{XY} corresponds to in a way the "angle" between X and Y.

i.e. (assume zero mean):

$r_{XY} = 0 \longrightarrow E\{XY\} = 0 \longrightarrow \theta = 90°$ (orthogonal r.v.'s)

$r_{XY} = \pm 1 \longrightarrow Y = aX \longrightarrow \theta = 0; 180°$

Ex:

$Y = aX + N$, X and N are zero mean independent r.v.'s.

Find r_{XY}.

$$E\{Y\} = E\{aX+N\} = a\,\cancel{E\{X\}} + \cancel{E\{N\}} = 0$$

$$r_{XY} = \frac{E\{XY\}}{\sqrt{E\{X^2\}E\{Y^2\}}} = \frac{E\{X(aX+N)\}}{\sqrt{E\{X^2\}E\{Y^2\}}}$$

$$= \frac{a\,E\{X^2\} + E\{XN\}}{\sqrt{\sigma_X^2 \; E\{(aX+N)^2\}}}$$

independent $= \underbrace{E\{X\}}_{0} \cdot \underbrace{E\{N\}}_{0} = 0$

σ_X^2 → central moment and moment coincides

$$= \frac{a\,E\{X^2\}}{\sqrt{\sigma_X^2\left(a^2 E\{X^2\} + 2a\,\underbrace{\cancel{E\{XN\}}}_{0} + \underbrace{E\{N^2\}}_{\sigma_N^2}\right)}}$$

$$= \frac{a \cdot \sigma_X^2}{\sqrt{\sigma_X^2 a^2 \sigma_X^2 \left(1 + \frac{\sigma_N^2}{a^2 \sigma_X^2}\right)}} = \operatorname{sign}(a) \cdot \frac{1}{\sqrt{1 + \frac{\sigma_N^2}{a^2 \sigma_X^2}}}$$

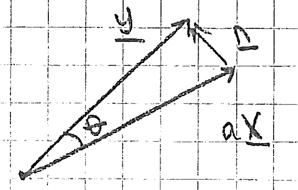
$$\text{SNR} \triangleq \frac{\text{average signal power}}{\text{average noise power}} = \frac{E\{(aX)^2\}}{E\{N^2\}} = \frac{a^2 \sigma_X^2}{\sigma_N^2}$$

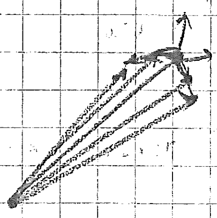
(signal to noise ratio)

$$r_{XY} = \frac{1}{\sqrt{1 + \frac{1}{\text{SNR}}}}$$

r_{XY} in this example is clearly less or equal to 1 for all SNR values.

Also, as SNR $\to \infty$, $r_{XY} \to 1$; that is, for a fixed σ_X^2, as $\sigma_n^2 \to 0$ we have $r_{XY} \to 1$.





→ observations are in high correlation with the signal. θ is small or correlation coefficient ($\cos\theta$) is large.

Correlation: $E\{XY\} \longrightarrow$ correlation of X and Y.

Correlation coefficient:
(general definition for non-zero mean X and Y)

$$r_{XY} \triangleq \frac{E\{(X-\bar{X})(Y-\bar{Y})\}}{\sqrt{E\{(X-\bar{X})^2\}\,E\{(Y-\bar{Y})^2\}}}$$

inside the expectation: zero-mean

The value of r_{XY} is independent of $\bar{X}$ and $\bar{Y}$ since r_{XY} is a function of $(X-\bar{X})$ and $(Y-\bar{Y})$.

$$r_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}}$$

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E\{(X-\bar{X})(Y-\bar{Y})\}\\ &= E\{XY\} - \bar{X}E\{Y\} - \bar{Y}\cancel{E\{\bar{X}\}}\cancel{+\bar{X}}\,\bar{Y}\\ &= E\{XY\} - \bar{X}\bar{Y}\end{aligned}$$

$$\mathrm{Cov}(X,Y) = 0 \longleftrightarrow \underbrace{E\{XY\} = E\{X\}E\{Y\}}_{\substack{\text{X and Y are called}\\ \text{uncorrelated r.v.'s}\\ \text{(but not independent!)}}}$$

(if zero mean $\rightarrow E\{XY\}=0 \rightarrow$ orthogonal

Ex:

$$X = \begin{cases} 1 & , \text{"A" happens} \\ 0 & , \text{other} \end{cases} \qquad Y = \begin{cases} 1 & , \text{"B" happens} \\ 0 & , \text{other} \end{cases}$$

(indicator function of event A) (indicator function of event B)

X and Y are binary valued r.v.'s. Find Cov(X,Y)

$$Cov(X,Y) = E_{XY}\{XY\} - E_X\{X\} E_Y\{Y\}$$

$$= \underbrace{1 . P(X=1, Y=1)}_{(\text{otherwise} \rightarrow 0)} - (1 . P(X=1))(1 . P(Y=1))$$

Case ① $Cov(X,Y) = 0$

$$P(\underbrace{X=1}_{A}, \underset{\cap}{} \underbrace{Y=1}_{B}) = P(X=1)P(Y=1)$$

*) A and B are independent.

Case ② $Cov(X,Y) \geqslant 0$

$$\frac{P(X=1, Y=1)}{P(Y=1)} \geqslant P(X=1) \quad \text{OR} \quad \frac{P(X=1, Y=1)}{P(X=1)} \geqslant P(Y=1)$$

$$P(\underbrace{X=1}_{A} | \underbrace{Y=1}_{B}) \geqslant \underbrace{P(X=1)}_{A} \quad \text{OR} \quad P(\underbrace{Y=1}_{B} | \underbrace{X=1}_{A}) \geqslant \underbrace{P(Y=1)}_{B}$$

*) if Cov(X,Y) is positive, events A and B are "happening" together.

*) knowing B has happened, probability of A increases (B affecting A)

Properties of Cov (X,Y)

$$\left.\begin{array}{ll} ① & \mathrm{Cov}(X,X) = \mathrm{Var}(X) \\ ② & \mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X) \\ ③ & \mathrm{Cov}(aX,Y) = a\,\mathrm{Cov}(X,Y) \\ ④ & \mathrm{Cov}(X+Y,Z) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z) \end{array}\right\}$$

please compare
with axioms
for inner product.

Ex: $\mathrm{Var}\left(\sum_{i=1}^{N} x_i\right)$? x_i's are jointly defined r.v.'s.

$$\mathrm{Var}\left(\sum_{i=1}^{N} x_i\right) \overset{①}{=} \mathrm{Cov}\left(\sum_{i=1}^{N} x_i, \sum_{j=1}^{N} x_j\right)$$

$$\overset{④}{=} \sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Cov}(x_i,x_j)$$

$$= \sum_{i=1}^{N} \underbrace{\mathrm{Cov}(x_i,x_i)}_{\mathrm{Var}(x_i)} + \underbrace{\sum_{\substack{i=1 \\ i\neq j}}^{N}\sum_{j=1}^{N} \mathrm{Cov}(x_i,x_j)}_{2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \mathrm{Cov}(x_i,x_j)}$$

$$\begin{array}{c|ccccc} j= & 1 & & j & & N \\ \hline i=1 & \mathrm{Cov}(x_1,x_1) & \cdots & \mathrm{Cov}(x_1,x_j) & \cdots & \mathrm{Cov}(x_1,x_N) \\ \vdots & & & & & \\ i & \mathrm{Cov}(x_i,x_1) & \cdots & \mathrm{Cov}(x_i,x_j) & \cdots & \vdots \\ \vdots & & & & & \\ N & \mathrm{Cov}(x_N,x_1) & \cdots & & \cdots & \mathrm{Cov}(x_N,x_N) \end{array}$$

diagonals → $\mathrm{Var}(x_i)$

If two random variables are independent → X and Y are uncorrelated

Uncorrelatedness → $E\{XY\} = E\{X\}E\{Y\}$

independent → $E\{XY\} = \iint xy \underbrace{f_{XY}(x,y)}_{f_X(x)f_Y(y)} dx\,dy = \underbrace{\int x f_X(x)\,dx}_{E\{X\}=\bar{X}} \underbrace{\int y f_Y(y)\,dy}_{E\{Y\}=\bar{Y}}$

Random Vectors

Previously, 2 r.v.'s such as X and Y have been discussed and the "linear dependence" between X and Y has been specified in terms of covariance / correlation of X & Y.

Now, $\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} \longrightarrow$ joint distribution of N r.v.'s is needed.

Let's define correlation matrix for a partial description.

$$\underset{\substack{(\\ \text{correlation} \\ \text{matrix}}}{\underline{\underline{R}}_X} = E\{\underline{X}\,\underline{X}^T\} = E\left\{ \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} [X_1 \cdots X_N] \right\}$$

Grammian matrix (matrix of inner products)

$$= \begin{bmatrix} E\{X_1^2\} & E\{X_1X_2\} & \cdots & E\{X_1X_n\} \\ E\{X_2X_1\} & & & \\ \vdots & & & \\ E\{X_nX_1\} & \cdots & & E\{X_n^2\} \end{bmatrix}$$

$$\underset{\substack{(\\ \text{Covariance} \\ \text{matrix}}}{\underline{\underline{C}}_X} = E\{(\underline{X}-E\{\underline{X}\})(\underline{X}-E\{\underline{X}\})^T\}$$

$$= \begin{bmatrix} Var(X_1) & Cov(X_1,X_2) & \cdots & Cov(X_1,X_n) \\ Cov(X_2,X_1) & Var(X_2) & & \vdots \\ \vdots & & & \vdots \\ Cov(X_n,X_1) & \cdots & & Var(X_n) \end{bmatrix}$$

$$E\{(X_1-\overline{X_1})(X_2-\overline{X_2})\} = Cov(X_1,X_2)$$

Properties of Covariance Matrix

① Hermitian Symmetric

$$\underline{\underline{C}}_x = \underline{\underline{C}}_x^H$$

→ eigenvectors of $\underline{\underline{C}}_x$ are orthogonal

→ eigenvalues are real

$\underline{\underline{R}}_x = E\{\underline{x}\,\underline{x}^T\}$ → real valued vectors

$\underline{\underline{R}}_x = E\{\underline{x}\,\underline{x}^H\}$ → complex valued vectors

$$\underline{\underline{R}}_x^H = E\{(\underline{x}\,\underline{x}^H)^H\} = E\{\underline{x}\,\underline{x}^H\} = \underline{\underline{R}}_x$$

② Positive Semi-Definite

$\underline{\underline{C}}_x \geq 0$ → eigenvalues are non-negative

$$\underline{z}^H \underline{\underline{R}}_x \underline{z} \overset{?}{\geq} 0 \quad \forall \underline{z} \neq \underline{0}$$

$$= \underline{z}^H E\{\underline{x}\,\underline{x}^H\} \underline{z} = E\{(\underline{z}^H \underline{x})(\underline{x}^H \underline{z})\}$$

$$= E\{|\underline{x}^H \underline{z}|^2\} \geq 0$$

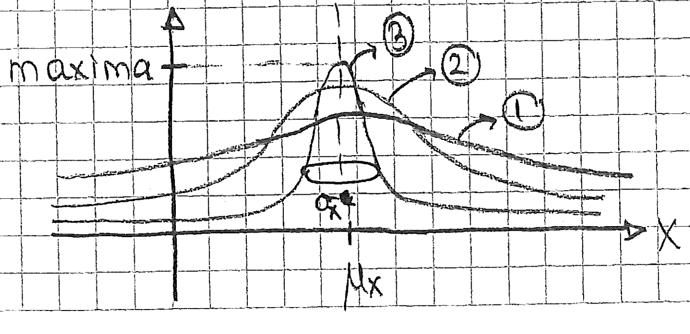
Gaussian Distribution

$\boxed{1\text{-}D}$ $\quad X \sim N(\underset{\text{mean}}{\mu_x}, \underset{\text{variance}}{\sigma_x^2})$ $\qquad f_X(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}$

$$E_X\{X\} = \mu_x$$

$$E\{(X-\mu_x)^2\} = \sigma_x^2$$

$$\sigma^2_{x③} < \sigma^2_{x②} < \sigma^2_{x①}$$



%60 of the probability around mean value lies in σ.

N-dimensional

$\underline{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ — joint distribution of $X_1, \ldots, X_n$

↓

X is jointly Gaussian distributed if density is:

$$f_{\underline{X}}(x_1,\ldots,x_n) = \frac{1}{(2\pi)^{\frac{N}{2}} \, \underbrace{|\underline{\underline{C}}_x|}_{\text{determinant of } \underline{\underline{C}}_x}^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(\underline{x}-\underline{\mu_x})^T \underline{\underline{C}}_x^{-1} (\underline{x}-\underline{\mu_x})}$$

$\underline{\mu_x}$: mean vector

$\underline{\underline{C}}_x$: covariance matrix

2-D

$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \underline{\underline{C}}_x = \left[\begin{array}{c|c} \sigma_{x_1}^2 & \text{Cov}(x_1,x_2) \\ \hline \text{Cov}(x_1,x_2) & \sigma_{x_2}^2 \end{array}\right]$

* $\text{Cov}(x_1,x_2) = r_{x_1x_2}\sigma_{x_1}\sigma_{x_2}$

* $\underline{\mu_x} = 0$

$$f_{\underline{X}}(x_1,x_2) = \frac{1}{2\pi\,|\underline{\underline{C}}_x|^{\frac{1}{2}}} \cdot \exp\left(\frac{-1}{2}\begin{bmatrix} x_1 & x_2 \end{bmatrix} \left[\begin{array}{c|c} \sigma_{x_1}^2 & r_{x_1x_2}\sigma_{x_1}\sigma_{x_2} \\ \hline r_{x_1x_2}\sigma_{x_1}\sigma_{x_2} & \sigma_{x_2}^2 \end{array}\right]^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)$$

* exponential term → $\exp\Big(\underbrace{\frac{-1}{2}\big(\underline{x}^T \underline{\underline{C}}_x^{-1} \underline{x}}_{\text{negative}}\big)\Big)$

↓ not blowing up ✓ $\qquad \underline{\underline{C}}_x > 0 \rightarrow \underline{\underline{C}}_x^{-1} > 0$

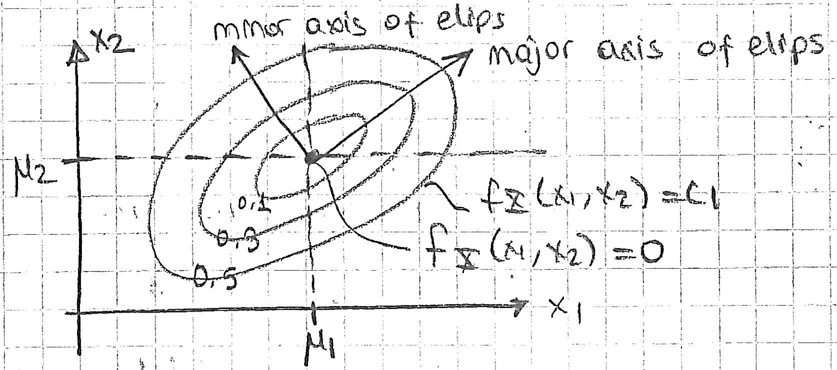
at very large values → approaches to values.

* $\exp\left(\frac{-1}{2}(\underline{x}-\underline{\mu_x})^T \underline{\underline{C}}_x^{-1}(\underline{x}-\underline{\mu_x})\right)$

↳ minima is at $\underline{x} = \underline{\mu_x}$

Level Curves

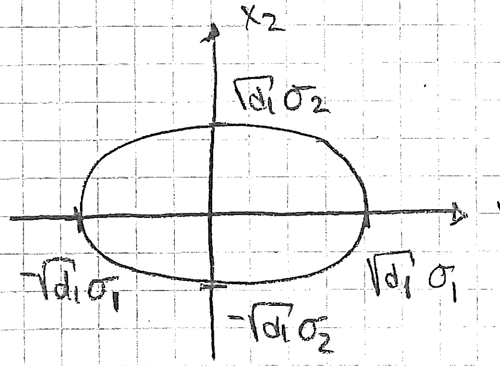
$f_{\underline{X}}(x_1,x_2) = C_1 \longrightarrow (\underline{x}-\underline{\mu_x})^T \underline{\underline{C}}_x^{-1}(\underline{x}-\underline{\mu_x})$ should be a constant.



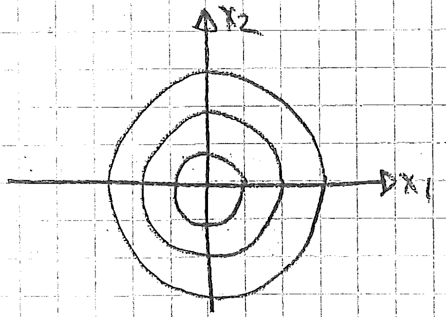
Assume

$$\underline{\underline{C}}_x = \begin{bmatrix}\sigma_1^2 & 0\\ 0 & \sigma_2^2\end{bmatrix} , \quad \underline{\mu_x} = \underline{0}$$

$$\underline{x}^T \underline{\underline{C}}_x^{-1} \underline{x} = [x_1 \ \ x_2]\begin{bmatrix}\sigma_1^2 & 0\\ 0 & \sigma_2^2\end{bmatrix}^{-1}\begin{bmatrix}x_1\\ x_2\end{bmatrix} = \frac{x_1^2}{\sigma_1^2} + \frac{x_2^2}{\sigma_2^2} = d_1$$



if $\underline{\underline{C}}_x = \begin{bmatrix}\sigma_x^2 & 0\\ 0 & \sigma_x^2\end{bmatrix} = \sigma_x^2\,\underline{\underline{I}}$



$$\frac{x_1^2}{\sigma_x^2} + \frac{x_2^2}{\sigma_x^2} = d_1$$

Some Facts on Multivariate Gaussians

① Marginalization

$X_1, X_2, \ldots, X_n$ → jointly Gaussian → marginals X_k $(k = 1, \ldots, N)$ are also Gaussian

→ pairwise (X_k, X_l) is also jointly Gaussian

(simplest proof:
moment generating function for
n variables. $(s_1, s_2, \ldots, s_n)$
$s_2 = \ldots = s_n = 0 \longrightarrow$ m.g.f. of marginal s_1)

Ex: $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\underline{\underline{C}}_x = \begin{bmatrix} 3 & 2 \\ 2 & 5 \end{bmatrix}$, $\underline{\mu}_x = \begin{bmatrix} 10 \\ 2 \end{bmatrix}$

Find density of x_1. $x_1 \sim N(10, 3)$

② Linear Processing of Gaussian vectors results in another Gaussian vector.

Ex: $\underline{X} \sim N(\underline{\mu}_x, \underline{\underline{C}}_x)$. Find $\underline{\underline{m}}\underline{X}$

Let $\underline{\underline{m}}\underline{X} = \underline{Y}$
↳ find joint density of Y vector.
$\Rightarrow Y \sim N(\underline{\mu}_y, \underline{\underline{C}}_Y)$

$\underline{\mu}_y : E\{\underline{Y}\} = E\{\underline{\underline{m}}\underline{X}\} = \underline{\underline{m}}\, E\{\underline{X}\} = \underline{\underline{m}}\, \underline{\mu}_x$

$$\underline{\underline{C}}_Y : E\{(\underline{Y} - \underline{\mu}_y)(\underline{Y} - \underline{\mu}_y)^T\} \xrightarrow{\underline{\mu}_y = 0} E\{\underline{Y}\,\underline{Y}^T\} = E\{\underline{\underline{m}}\,\underline{X}\,\underline{X}^T \underline{\underline{m}}^T\}$$

$$= \underline{\underline{m}}\, E\{\underline{X}\,\underline{X}^T\}\, \underline{\underline{m}}^T$$

$$= \underline{\underline{m}}\, \underline{\underline{C}}_x\, \underline{\underline{m}}^T \qquad (\underline{\underline{m}}\, \underline{\underline{R}}_x\, \underline{\underline{m}}^T)$$

Ex: $\mathrm{Var}\left(\sum_{i=1}^{N} x_i\right) = ?$ x_i's r.v.'s. (solution #2)

$$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}_{N\times 1} \longrightarrow z = \begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} = \underline{1}^T \underline{x}$$

Without any loss of generality, assume $\overline{x}_k = 0, \ \forall k.$

$$\mathrm{Var}\left(\sum_{i=1}^{N} x_i\right) = \mathrm{Var}(z) = E\{z^2\}$$

zero mean

$$= E\{(\underline{1}^T \underline{x})(\underline{1}^T \underline{x})\}$$

$$= E\{(\underline{1}^T \underline{x})(\underline{x}^T \underline{1})\}$$

$$= \underline{1}^T E\{\underline{x}\,\underline{x}^T\}\,\underline{1}$$

$$= \underline{1}^T \underline{\underline{R}}_x \underline{1}$$

(if not zero mean $\rightarrow \underline{1}^T \underline{\underline{C}}_x \underline{1}$)

$\boxed{09.11.2020}$

Summary

$\underline{x}$: random vector $\longrightarrow \underline{y} = \underline{\underline{A}}\,\underline{x}$

mean: $\underline{\mu}_x$

covariance: $\underline{\underline{C}}_x$

$$E\{\underline{y}\} = \underline{\underline{A}}\,\underline{\mu}_x$$

$$\underline{\underline{C}}_y = \underline{\underline{A}}\,\underline{\underline{C}}_x\,\underline{\underline{A}}^T \longrightarrow \underline{\underline{C}}_y = \underline{\underline{A}}\,\underline{\underline{C}}_x\,\underline{\underline{A}}^H$$

$$\downarrow \qquad\qquad\qquad \downarrow$$

$$\underline{\underline{R}}_y = E\{\underline{y}.\underline{y}^T\} \qquad \underline{\underline{R}}_y = E\{\underline{y}.\underline{y}^H\}$$

$$= E\{\underline{\underline{A}}\,\underline{x}\,\underline{x}^T \underline{\underline{A}}^T\}$$

$$= \underline{\underline{A}}\,E\{\underline{x}.\underline{x}^T\}\,\underline{\underline{A}}^T$$

$$= \underline{\underline{A}}\,\underbrace{\underline{\underline{R}}_x}_{\substack{\downarrow \\ \underline{\underline{C}}_x}}\,\underline{\underline{A}}^T$$

Decorrelation of Random Vectors

Given a random vector $\underline{X}$ with $\underline{\underline{R}}_x$ $(E\{\underline{x}\}=\underline{0})$, our goal is to find a $\underline{\underline{T}}$ such that $\underline{y}=\underline{\underline{T}}\underline{x}$ has a diagonal covariance / auto-correlation matrix.

① Diagonalization by Eigen Decomposition:

$$\underline{\underline{R}}_x = \underline{\underline{E}}\,\underline{\underline{\Omega}}\,\underline{\underline{E}}^H \qquad \underline{\underline{E}} = [\underline{e}_1\ \underline{e}_2\ \dots\ \underline{e}_N] \rightarrow \text{eigenvectors}$$

$$\underline{\underline{R}}_x \cdot \underline{e}_k = \lambda_k \cdot \underline{e}_k$$

We know it's Hermitian symmetric, so eigenvectors are orthogonal.

$$\underline{\underline{\Omega}} \rightarrow \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$$

(orthonormal $\rightarrow \underline{\underline{E}}\,\underline{\underline{E}}^H = \underline{\underline{E}}^H\underline{\underline{E}} = \underline{\underline{I}}$)

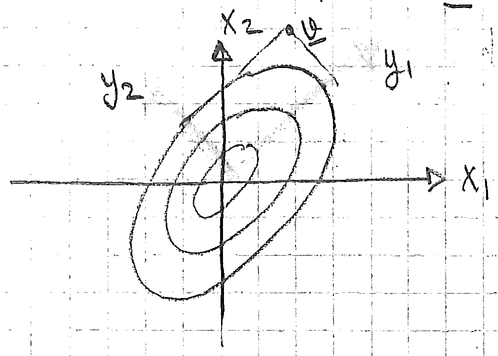
$(\underline{\underline{M}} = \underline{\underline{E}}\,\underline{\underline{\Omega}}\,\underline{\underline{E}}^{-1} \rightarrow$ general diagonalizable$)$

<u>Let</u> $\underline{\underline{T}} = \underline{\underline{E}}^H$, then $\underline{y} = \underline{\underline{T}}\,\underline{x} \longrightarrow \underline{\underline{R}}_y = \underline{\underline{T}}\,\underline{\underline{R}}_x\,\underline{\underline{T}}^H$

$$= \underline{\underline{E}}^H(\underline{\underline{E}}\,\underline{\underline{\Omega}}_x\,\underline{\underline{E}}^H)\underline{\underline{E}}$$

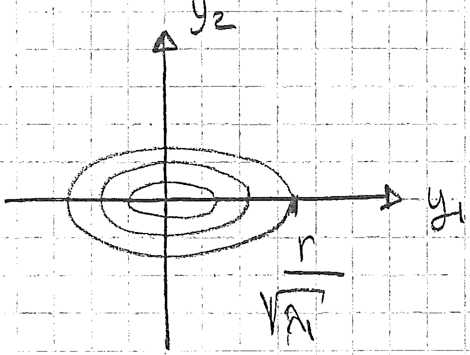
$$= \underline{\underline{\Omega}}_x \hookrightarrow \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \lambda_n \end{bmatrix}$$

Level curves of $\underline{x} \rightarrow \underline{x}^H \underline{\underline{R}}_x \underline{x} = f(\underline{x})$



Level curves of $\underline{y} \rightarrow \underline{y}^T \underline{\underline{R}}_y \underline{y}$

$$= y_1^2\lambda_1 + y_2^2\lambda_2 = r^2$$



$$f(\underline{x}) = \underline{x}^H \underline{\underline{R}}_x \underline{x} = \underbrace{\underline{x}^H \underline{\underline{E}}}_{\underline{y}^H} \underline{\underline{\Omega}} \underbrace{\underline{\underline{E}}^H \underline{x}}_{\underline{y}} \longrightarrow \underline{y} = \underline{\underline{E}}^H \underline{x} \longrightarrow \underbrace{\underline{\underline{E}}\,\underline{y} = \underline{x}}$$

$$\begin{bmatrix} \underline{e}_1 & \underline{e}_2 & \cdots & \underline{e}_N \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = I_{N \times N} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

Any vector v can be expressed in terms of canonical basis.

$$\begin{bmatrix} \underline{e}_1 & \underline{e}_2 & \cdots & \underline{e}_N \end{bmatrix} = I_{N \times N}$$

In an alternative basis using the eigenvectors, y_i's are the expansion of point v in the alternate basis.

$\underline{y} = \underline{\underline{E}}^H \underline{x}$ — expansion coefficients of v in canonical basis.

$\underline{\underline{E}}^H$ — change of basis matrix

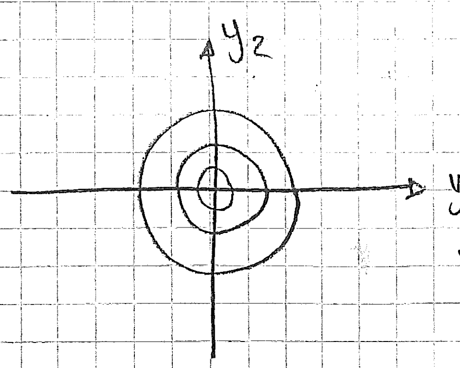
$\underline{y}$ — expansion coefficients in the eigen basis

② Diagonalization by Unitary Transformation Followed by Scaling

Let $\underline{\underline{T}} = \underset{\substack{\downarrow \\ \text{scale}}}{\underline{\underline{\Omega}}^{-1/2}} \underset{\substack{\downarrow \\ \text{decorrelate}}}{\underline{\underline{E}}^H} \longrightarrow \underline{y} = \underline{\underline{T}}\,\underline{x} \rightarrow \underline{\underline{R}}_y = \underline{\underline{T}}\,\underline{\underline{R}}_x \underline{\underline{T}}^H$

$$= \underline{\underline{\Omega}}^{-1/2} \underline{\underline{E}}^H (\underline{\underline{E}}\,\underline{\underline{\Omega}}\,\underline{\underline{E}}^H) \underline{\underline{E}}\,\underline{\underline{\Omega}}^{-1/2}$$

$$= \underline{\underline{I}}$$



level curves after "$\underline{\underline{T}}$-mapping"

③ Diagonalization by LU Decomposition

$$\underline{\underline{A}} = \underline{\underline{L}}\,\underline{\underline{U}}$$

$\underline{\underline{U}}$ → upper Δ matrix

$\underline{\underline{L}}$ → lower Δ matrix

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ -d/a & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\underline{\underline{L_1}}} \underbrace{\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}}_{\underline{\underline{A}}} = \begin{bmatrix} a & b & c \\ 0 & j & k \\ g & h & i \end{bmatrix}$$

$$\underline{\underline{L_2}}\,\underline{\underline{L_1}}\,\underline{\underline{A}} = \begin{bmatrix} a & b & c \\ 0 & j & k \\ 0 & l & m \end{bmatrix} \longrightarrow \underbrace{\underline{\underline{L_3}}\,\underline{\underline{L_2}}\,\underline{\underline{L_1}}}_{\underline{\underline{L}}_{Left}}\,\underline{\underline{A}} = \underbrace{\begin{bmatrix} a & b & c \\ 0 & j & k \\ 0 & 0 & n \end{bmatrix}}_{\underline{\underline{U}}}$$

$$\underline{\underline{A}} = \underline{\underline{L}}_{Left}^{-1}\,\underline{\underline{U}}$$

(*) $$\begin{bmatrix} 10 & 0 \\ 5 & 20 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 \\ 5/10 & 1 \end{bmatrix}} \begin{bmatrix} 10 & 0 \\ 0 & 20 \end{bmatrix} = \underline{\underline{L_u}} \cdot \underline{\underline{U}}$$

Lower Δ matrices with 1's on diagonal are called unit lower Δ matrices ($\underline{\underline{L_u}}$)

(*) $$\begin{bmatrix} 10 & 0 \\ 5 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 10 & 0 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 10 & 0 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 1/10 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 10 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} 1 & 0 \\ 5/10 & 1 \end{bmatrix} \begin{bmatrix} 10 & 0 \\ 0 & 2 \end{bmatrix}$$

Then, $\underline{\underline{R_x}} = \underline{\underline{L_1}}\,\underline{\underline{U_1}} \rightarrow \underline{\underline{R_x}} = \underline{\underline{R_x}}^H \longrightarrow \underline{\underline{R_x}}^H = \underline{\underline{U_1}}^H \underline{\underline{L_1}}^H = \underline{\underline{L_2}}\,\underline{\underline{U_2}}$

$$\underline{\underline{U_1}} = \underline{\underline{L_2}}^H$$

$\downarrow$

$$= \underline{\underline{L_1}} \cdot \underline{\underline{L_2}}^H$$

$$= \underline{\underline{L_u}}\,\underline{\underline{D}}\,\underline{\underline{L}}^H$$

$$= \underline{\underline{L_u}}\,\underbrace{\underline{\underline{D}}\,\underline{\underline{D}}^H}_{\text{diagonal with real values}}\,\underline{\underline{L_u}}^H$$

$$\underline{\underline{R}}_x = \underline{\underline{L}}_u \underline{\underline{D}}\, \underline{\underline{D}}^H \underline{\underline{L}}_u^H \quad , \quad \boxed{\underline{\underline{T}} = \underline{\underline{L}}_u^{-1}}$$

$$\underline{y} = \underline{\underline{T}}\, x \longrightarrow \underline{\underline{R}}_y = \underline{\underline{L}}_u^{-1}\, \underline{\underline{R}}_x (\underline{\underline{L}}_u^{-1})^H$$

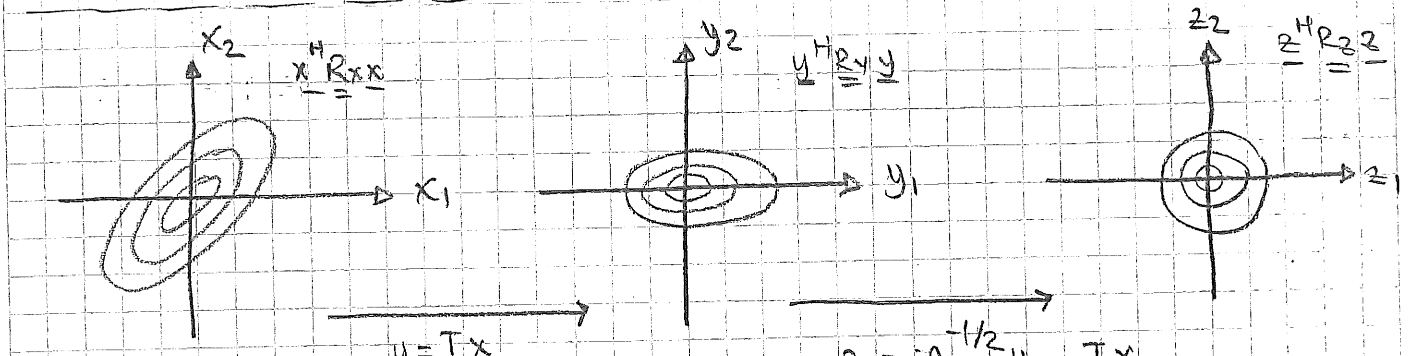
$$= \underline{\underline{L}}_u^{-1} \underline{\underline{L}}_u \underline{\underline{D}}\, \underline{\underline{D}}^H \underline{\underline{L}}_u^H (\underline{\underline{L}}_u^H)^{-1}$$

$$= \underline{\underline{D}}\, \underline{\underline{D}}^H \longrightarrow \text{forreal-valued : } \underline{\underline{D}}^2$$

$\underline{\underline{L}}_u^{-1}$ is also a lower Δ matrix. $\longrightarrow \begin{bmatrix} 1 & & & 0 \\ \alpha_1 & 1 & & \\ \beta_1 & \alpha_2 & 1 & \\ \gamma_1 & \beta_2 & \alpha_3 & 1 \end{bmatrix}$

$$\underline{y} = \begin{bmatrix} \underline{\underline{L}}_u^{-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \underbrace{\begin{bmatrix} x_1 \\ \alpha_1 x_1 + x_2 \\ \beta_1 x_1 + \alpha_2 x_2 + x_3 \\ \gamma_1 x_1 + \beta_2 x_2 + \alpha_3 x_3 + x_4 \end{bmatrix}}_{\text{causal combination of elements of } \underline{X} \text{ vector}} \longrightarrow \text{causal decorrelator}$$

$$x_1 \xrightarrow[\substack{t_1 \\ \text{seconds}}]{} x_2 \xrightarrow[\substack{t_2 \\ \text{seconds}}]{} x_3$$



$\underline{y} = \underline{\underline{T}}\, \underline{x}$

$\underline{\underline{T}} = \underline{\underline{E}}^H$ ①

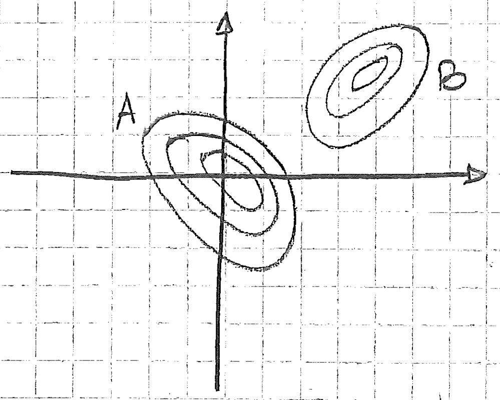
$\underline{z} = \underline{\underline{\Omega}}^{-1/2} \underline{y} = \underline{\underline{T}}\, \underline{x}$

$\underline{\underline{T}} = \underline{\underline{\Omega}}^{-1/2} \underline{\underline{E}}^H$ ②

③ $\underline{\underline{T}} = \underline{\underline{L}}_u^{-1} \qquad \underline{\underline{R}}_x = \underline{\underline{L}}_u \underline{\underline{D}}_x \underline{\underline{L}}_u^H = \underline{\underline{Q}} \cdot \underline{\underline{Q}}^H \quad (\underline{\underline{Q}} = \underline{\underline{L}}_u \underline{\underline{D}}_x^{1/2})$

Cholesky Decomposition

④ Joint Diagonalization of Two Covariance Matrices

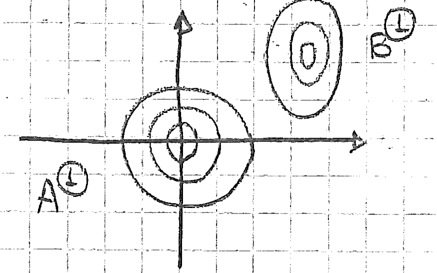


Goal: Given $\underline{\underline{R}}_A$ and $\underline{\underline{R}}_B$. Find $\underline{\underline{T}}$ such that after transformation (i.e. $\underline{\underline{T}}\,\underline{\underline{R}}_A\underline{\underline{T}}^H$ and $\underline{\underline{T}}\,\underline{\underline{R}}_B\underline{\underline{T}}^H$) we have diagonal autocorrelation matrices.

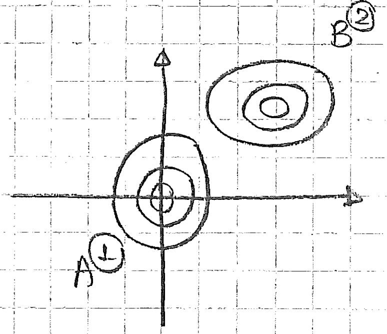
Step-1

$$\underline{\underline{T}}_1 = \underline{\underline{\Omega}}_A^{-1/2}\,\underline{\underline{E}}_A^H \qquad (\underline{\underline{R}}_A = \underline{\underline{E}}_A\,\underline{\underline{\Omega}}\,\underline{\underline{E}}_A^H)$$

At the end of Step-1



At the end of Step-2



A① → A after T1

B① → B after T1

A② → A after T2

B② → B after T2

Step-2

$$\underline{\underline{R}}_B^{\textcircled{1}} = \underline{\underline{T}}_1\,\underline{\underline{R}}_B\,\underline{\underline{T}}_1^H$$

Diagonalize $\underline{\underline{R}}_B^{\textcircled{1}}$ with eigenvectors of $\underline{\underline{R}}_B^{\textcircled{1}}$.

Question: If we diagonalize $\underline{\underline{R}}_B^{\textcircled{1}}$, is the alignment of $\underline{\underline{R}}_A^{\textcircled{1}}$ (concentric circles) lost or not?

$$\underline{\underline{R}}_B^{\textcircled{1}}\,\underline{e}_B^{\textcircled{1}} = \lambda_k \cdot \underline{e}_B^{\textcircled{1}} \longrightarrow \underline{\underline{T}}_2 = \begin{bmatrix} e_{B_1}^{\textcircled{1}\,H} \\ e_{B_2}^{\textcircled{1}\,H} \\ \vdots \\ e_{B_N}^{\textcircled{1}\,H} \end{bmatrix} = \left(\underline{\underline{E}}_B^{\textcircled{1}}\right)^H$$

$$\underline{\underline{R}}_A^{\textcircled{1}} = \underline{\underline{I}} \longrightarrow \underline{\underline{T}}_2\,\underline{\underline{R}}_A^{\textcircled{1}}\,\underline{\underline{T}}_2^H = \underline{\underline{R}}_A^{\textcircled{2}} = \underline{\underline{I}}$$

So, $\underline{\underline{R}}_A^{\textcircled{2}}$ remains decorrelated. (level curves are aligned with axes)

$\boxed{\text{Final mapping:}}$

$$\underline{\underline{T}}_{final} = \underline{\underline{T}}_2\,\underline{\underline{T}}_1 = \left[\underline{\underline{E}}_B^{\textcircled{1}}\right]^H \left[\underline{\underline{\Omega}}_A^{-1/2}\,\underline{\underline{E}}_A^H\right]$$

$$(\text{Step-2}) \rightarrow \underline{\underline{R}}_B^{\textcircled{1}}\,\underline{e}_B^{\textcircled{1}} = \lambda_k\,\underline{e}_B^{\textcircled{1}}$$

$$\underline{\underline{T}}_1\,\underline{\underline{R}}_B\,\underline{\underline{T}}_1^H \hookleftarrow$$

$$\downarrow$$

$$\underline{\underline{R}}_B\,\underline{\underline{T}}_1^H\,\underline{e}_B^{\textcircled{1}} = \lambda_k\,\underline{\underline{T}}_1^{-1}\,\underline{e}_B^{\textcircled{1}}$$

$$= \lambda_k \underbrace{\overbrace{(\underline{\underline{E}}_A\,\underline{\underline{\Omega}}_A^{1/2})}(\underline{\underline{\Omega}}_A^{1/2}\,\underline{\underline{E}}_A^H}_{\underline{\underline{R}}_A}\;\underbrace{\underline{\underline{E}}_A\,\underline{\underline{\Omega}}_A^{-1/2})}_{\underline{\underline{T}}_1^H}\,\underline{e}_B^{\textcircled{1}}$$

$$\underline{\underline{R}}_B\,\underline{\underline{T}}_1^H\,\underline{e}_B^{\textcircled{1}} = \lambda_k\,\underline{\underline{R}}_A\,\underbrace{\underline{\underline{T}}_1^H\,\underline{e}_B^{\textcircled{1}}}_{\underline{f}_k}$$

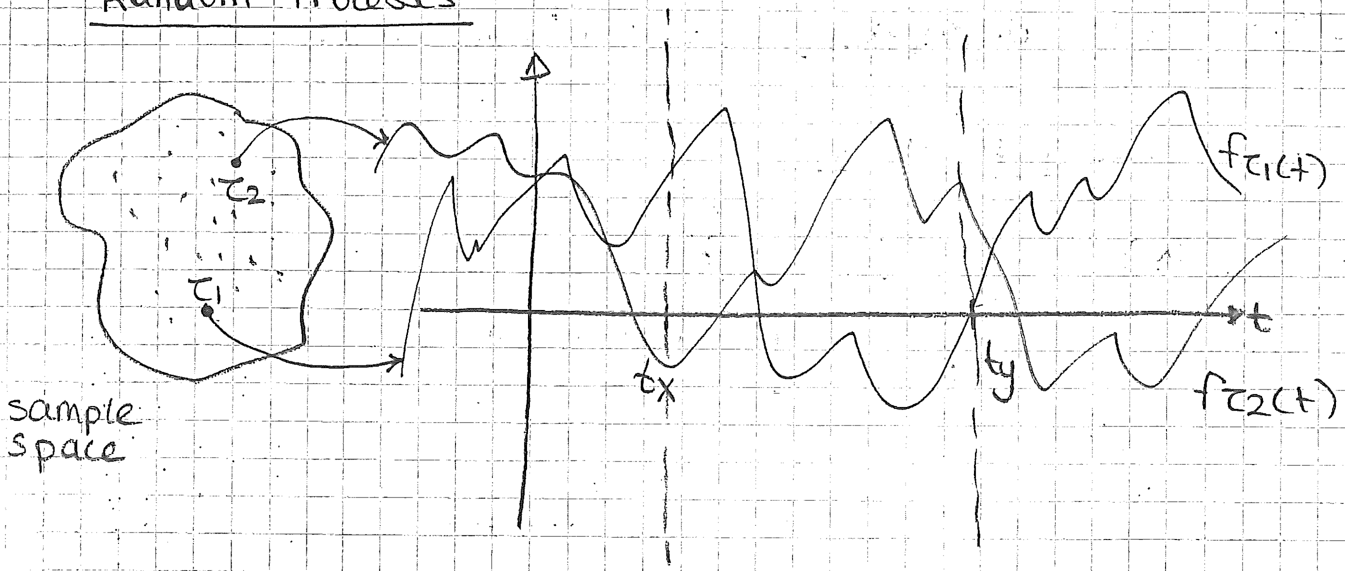
$$\boxed{\underline{\underline{R}}_B\,\underline{f}_k = \lambda_k\,\underline{\underline{R}}_A\,\underline{f}_k} \longrightarrow$$ generalized eigenvector of $\underline{\underline{R}}_A$ and $\underline{\underline{R}}_B$ (eig(B,A))

$$\underline{\underline{F}} = [\underline{f}_1\;\underline{f}_2\;\ldots\;\underline{f}_N]$$

$$\underline{\underline{T}}_{final} = \begin{bmatrix} E_{B1}^{\textcircled{1}\,H} \\ E_{B2}^{\textcircled{1}\,H} \\ \vdots \\ E_{B_N}^{\textcircled{1}\,H} \end{bmatrix} \cdot \underline{\underline{T}}_1 = \begin{bmatrix} \underline{f}_1^H \\ \vdots \\ \underline{f}_N^H \end{bmatrix} = \underline{\underline{F}}^H$$

Therrien's Textbook

Random Processes



$f_{\tau_1}(t), f_{\tau_2}(t) \rightarrow$ realizations of random process

$t \rightarrow$ process variable.

Ex: Random Frequency Cosine

$x(t) = \cos|\underset{\sim}{w}t+30^\circ|$

random variable.

Comments

① For a fixed random experiment outcome, τ_k, $f_{\tau_k}(t)$ is a function of time.

② If "t" is fixed to "t_x" $(f_{\tau_1}(t_x), f_{\tau_2}(t_x)) \rightarrow \underset{\sim}{z} = f(t_x)$ random variable

(fix 2 random times → random vector)

③ If $z_1 = \underset{\sim}{f}(t_1)$ and $z_2 = \underset{\sim}{f}(t_2)$, $\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \underset{\sim}{z} \longrightarrow$ random vector

($\underset{\sim}{f}$: random process)

72

Description of Random Processes

① Joint pdf description

$\underset{\sim}{x}(t)$: random process

***First Order pdf Description:**

$f_{\underset{\sim}{x}(t_1)}(x_1) \quad \forall t_1$ → r.p. evaluated at t_1 → density of x_1

(if evaluation time changes, density changes)

***Second Order pdf Description:**

$f_{\underset{\sim}{x}(t_1)\underset{\sim}{x}(t_2)}(x_1, x_2) \quad \forall t_1, \forall t_2$

$\left.\begin{array}{l} X_1 = \underset{\sim}{x}(t_1) \\ X_2 = \underset{\sim}{x}(t_2) \end{array}\right\}$ random variables.

***N^{th} Order pdf Description:**

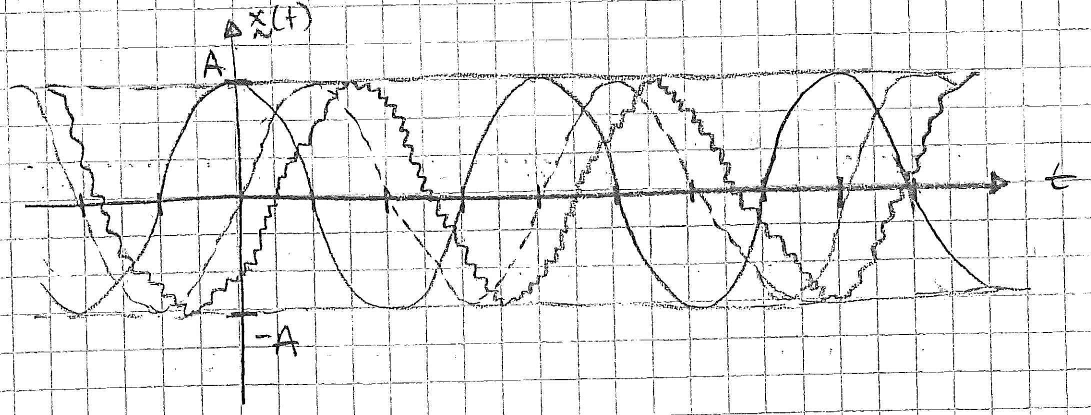
$f_{\underset{\sim}{x}(t_1)\underset{\sim}{x}(t_2)\cdots\underset{\sim}{x}(t_N)}(x_1, x_2, \ldots, x_N) \quad \forall t_1, \forall t_2, \ldots, \forall t_N$

Example

$\underset{\sim}{x}(t) = A\cos(2\pi f t + \underset{\sim}{\theta})$ (θ → generating random process)

$\underset{\sim}{\theta}$: uniform $[0, 2\pi)$

- say $\theta = 5$ → $\underset{\sim}{x}(t)$ is a realization of the random process



3 realizations.

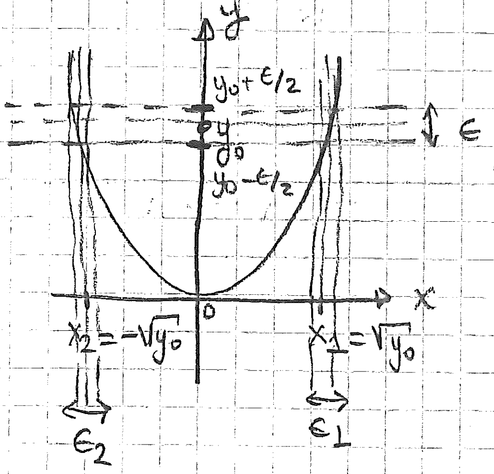
Find joint pdf description of $\underset{\sim}{x}(t)$

$\boxed{1^{st}\text{ order pdf}}$

fixed — only θ is random

$$\underset{\text{r.v.}}{X_1} = \underset{\sim}{x}(t_1) = A\cos(2\pi f \overset{\text{fixed}}{t_1} + \underset{\text{r.v.}}{\underset{\sim}{\theta}}) \longrightarrow f_{\underset{\sim}{x}(t_1)}\, x_1 = ?$$

* One function of one r.v



$$\underset{\sim}{y} = f(\underset{\sim}{x}) \text{ — } f_X(x)$$

$$\longrightarrow \underset{\sim}{y} = \underset{\sim}{x}^2 = g(\underset{\sim}{x})$$

Given density of x, goal is finding the density of y ($f_{\underset{\sim}{y}}(y)$)

$$f_y(y_0)\cdot\epsilon \approx P\left(y\in\left(y_0-\frac{\epsilon}{2}, y_0+\frac{\epsilon}{2}\right)\right) \approx f_x(\sqrt{y_0})\,\epsilon_1 + f_x(-\sqrt{y_0})\,\epsilon_2$$

for a small point around $(\sqrt{y_0}, y_0)$, approximate to a straight line.

$(+\sqrt{y_0}, y_0) \longrightarrow g'(x)$ at x_1

$(-\sqrt{y_0}, y_0) \longrightarrow g'(x)$ at x_2

$$f_y(y_0) \approx \frac{f_x(\sqrt{y_0})}{\epsilon/\epsilon_1} + \frac{f_x(-\sqrt{y_0})}{\epsilon/\epsilon_2}$$

as $\epsilon \longrightarrow 0$

$$\boxed{f_y(y_0) = \frac{f_x(\sqrt{y_0})}{\underbrace{|g'(x_1)|}_{2\sqrt{y_0}}} + \frac{f_x(-\sqrt{y_0})}{\underbrace{|g'(x_2)|}_{2\sqrt{y_0}}} \quad \text{where } x_k\text{'s satisfy } g(x_k)=y_0,\ k=\{1,2\}}$$

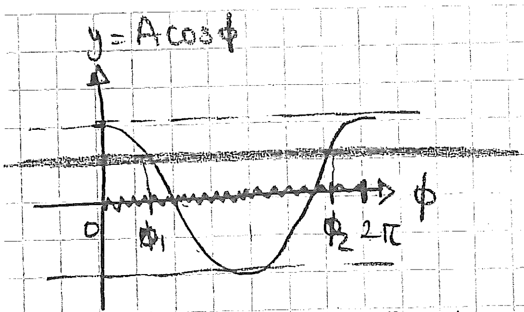
$$X_1 = A\cos(2\pi f t_1 + \underset{\sim}{\theta}) \longrightarrow \overset{d}{=} A\cos(\underset{\sim}{\phi})$$

$\underset{\sim}{\theta}$ → uniform $[0, 2\pi)$

→ shifted version of the uniform density

$\underset{\sim}{\phi}$ → uniform $[0, 2\pi)$

because of cosine, we are always in a 2π-interval.

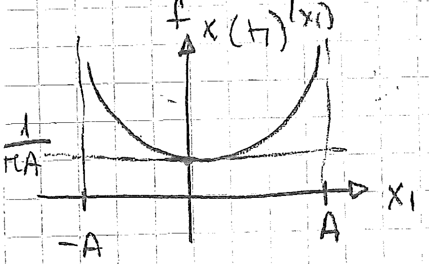


For a y value, we have two intervals.

$$f_{x(t_1)}(x_1) = \begin{cases} \dfrac{1/2\pi}{|-A\sin\phi_1|} + \dfrac{1/2\pi}{|-A\sin\phi_2|} , & |x_1| \leq A \\ 0 , & |x_1| > A \end{cases} \quad (\forall t_1)$$

$$= \begin{cases} \dfrac{1/2\pi}{\sqrt{A^2-x_1^2}} + \dfrac{1/2\pi}{\sqrt{A^2-x_1^2}} = \dfrac{1/\pi}{\sqrt{A^2-x_1^2}} , & |x_1| \leq A \\ 0 , & |x_1| > A \end{cases} \quad (\forall t_1)$$

$$\left(A\cos\phi_1 = x_1 \longrightarrow -A\sin\phi_1 = -A\sqrt{1-\cos^2\phi_1} = -\sqrt{A^2-x_1^2}\right)$$



If we are looking for values around $x(t_1)=0$, slope is very large (at cosine) r.p. is spending very little time.

At higher points, slope is small, r.p. is spending lots of time.

2nd Order pdf

$$f_{x(t_1)x(t_2)}(x_1,x_2) = f_{x(t_2)|x(t_1)}(x_2|x_1) \cdot f_{x(t_1)}(x_1) \qquad (t_1 \neq t_2)$$

$$\left.\begin{aligned} x_1 &= A\cos(2\pi f t_1 + \theta) \\ x_2 &= A\cos(2\pi f t_2 + \theta) \end{aligned}\right\} \quad \begin{aligned} x_2 &= A\cos(2\pi f t_2 + 2\pi f(t_1 - t_1) + \theta) \\ &= A\cos(\underbrace{2\pi f t_1 + \theta}_{\textcircled{I}} + \underbrace{2\pi f(\underbrace{t_2 - t_1}_{\Delta t})}_{\textcircled{II}}) \end{aligned}$$

$$= A(\cos\textcircled{I}\cos\textcircled{II} - \sin\textcircled{I}\sin\textcircled{II}) = (x_1)\cos\textcircled{II} - \left(\pm\sqrt{A^2-x_1^2}\right)\sin\textcircled{II}$$

$$x_2 = x_1\cos 2\pi f\Delta t \mp \sqrt{A^2-x_1^2}\,\sin(2\pi f\Delta t)$$

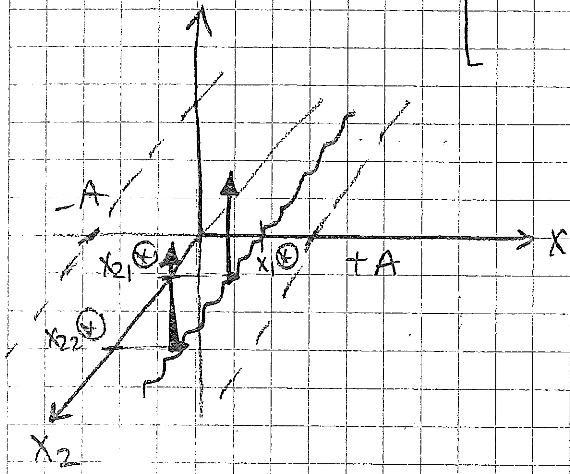
Call x_{21} and x_{22} as two possible values for x_2.

Claim : Since θ is uniform $\rightarrow$ x_{21} and x_{22} are equilikely.

$$f_{x(t_2)|x(t_1)}(x_2|x_1) = \frac{1}{2}\left[\delta(x_2 - x_{21}) + \delta(x_2 - x_{22})\right]$$

functions of Δt

$$f_{x(t_1)x(t_2)}(x_1,x_2) = \begin{cases} \frac{1}{2}\left[(\delta(x_2 - x_{21}) + \delta(x_2 - x_{22})\right]\frac{1/\pi}{\sqrt{A^2 - x_1^2}} & , |x_1| < A \\ 0 & , \text{o.w.} \end{cases}$$



- for every x_1, there are two impulses.
- weight becomes larger towards $\mp A$ (density of x_1)
- joint density depends on $\Delta t = t_2 - t_1$

3rd Order pdf

$$f_{x(t_1)x(t_2)x(t_3)}(x_1,x_2,x_3) = \underbrace{f_{x(t_3)|x(t_2),x(t_1)}(x_3|x_2,x_1)}_{\text{conditional density}}\ \underbrace{f_{x(t_1)x(t_2)}(x_1,x_2)}_{\checkmark}$$

from knowledge of

$x_1 = A\cos(2\pi f t_1 + \theta)$

$x_2 = A\cos(2\pi f t_2 + \theta)$,

θ can be uniquely found.

$x_3 = A\cos(2\pi f t_3 + \theta)$

$\hookrightarrow$ becomes deterministic.

$\delta(x_3 - \text{func}(x_1, x_2))$

There are two possibilities. If we know which one has occured, we know all the parameters of cosine. We have identified the cosine.

$(x_1^{\circledast}, x_{21}^{\circledast})$ or $(x_1^{\circledast}, x_{22}^{\circledast})$

Any two observations sufficiently fix another time value of the cosine function.

L16

16.11.2020

Gaussian Processes

A random process $\underset{\sim}{x}(t)$ is called Gaussian process $\forall N$ if every N samples of $\underset{\sim}{x}(t)$ is jointly Gaussian distributed.

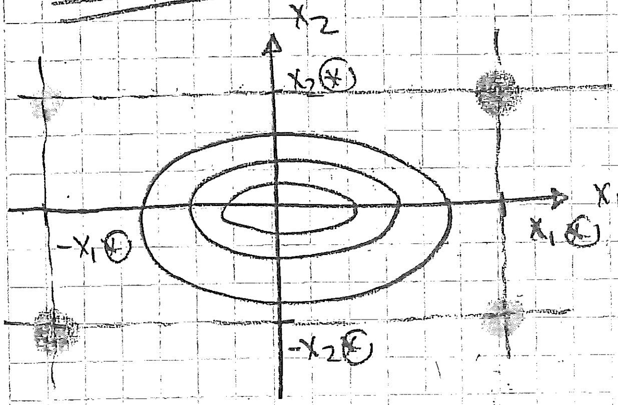
$$f_{x(t_1)x(t_2)\dots x(t_N)}(x_1, x_2, \dots, x_N) = \frac{1}{(2\pi)^{N/2}\, |\underline{\underline{C}}_x|^{1/2}} e^{-\frac{1}{2}[\underline{x}-\underline{\mu_x}]^T \underline{\underline{C}}_x^{-1} [\underline{x}-\underline{\mu_x}]}$$

$\underline{\mu_x} : E\{\underline{x}\}$

$$\underline{x} = \begin{bmatrix} x(t_1) \\ \vdots \\ x(t_N) \end{bmatrix} \qquad [\underline{\underline{C}}_x]_{k,l} = \text{Cov}(x_k, x_l)$$

k^{th} row, l^{th} column entry

Example (6.1 from Papoulis)

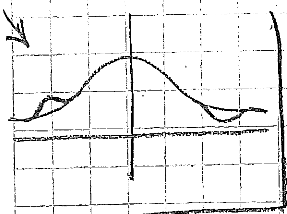


$\mathbb{X}_1, \mathbb{X}_2$ are jointly Gaussian.
The marginals $\mathbb{X}_1$ and $\mathbb{X}_2$ are also Gaussian.

Subtract from ⟶ Add to

$(x_1⊛, -x_2⊛) \longrightarrow (x_1⊛, x_2⊛)$

$(-x_1⊛, x_2⊛) \longrightarrow (-x_1⊛, -x_2⊛)$



Marginal distribution of x_1⊛ stays the same.
Marginal distribution of x_2⊛ stays the same.

If you only know $\mathbb{X}_1$ and $\mathbb{X}_2$ are marginally Gaussian distributed, this does not say anything about their joint distribution.

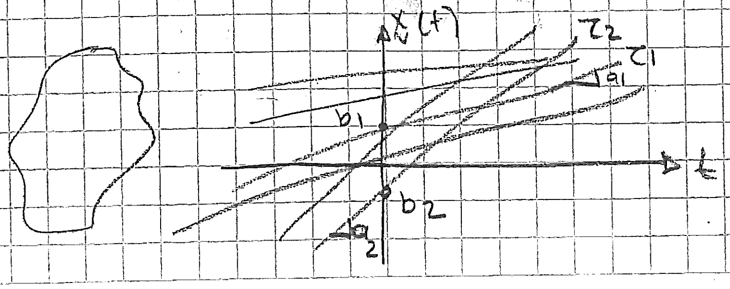
Example $\underset{\sim}{x}(t) = \underset{\sim}{a}t + \underset{\sim}{b}$ $\underset{\sim}{a}, \underset{\sim}{b} \rightarrow$ independent

⊛ $\underset{\sim}{a} \sim N(0, \sigma_a^2)$

⊛ $\underset{\sim}{b} \sim N(\mu_b, \sigma_b^2)$

⊛ $(\mu_b > 0)$
↳ more likely to have a positive b value.

Find joint pdf characterization of $\underset{\sim}{x}(t)$.



1st Order Pdf

$\underset{\sim}{x}(t_1) = x_1$ $\quad \underbrace{\underset{\sim}{a}\, t_1 + \underset{\sim}{b} = \underset{\sim}{x_1}}$ → one function of two r.v.s.

$\downarrow$ linear mapping

$$[t_1 \quad 1]\begin{bmatrix} \underset{\sim}{a} \\ \underset{\sim}{b} \end{bmatrix}$$

↙ random vector

If $\underset{\sim}{a}$ and $\underset{\sim}{b}$ are jointly Gaussian, I know that $\underset{\sim}{x_1}$ is also Gaussian r.v. (since linear combinations of jointly Gaussian r.v.s results in jointly Gaussian r.v.s)

In this example, $\underset{\sim}{a}$ and $\underset{\sim}{b}$ are jointly Gaussian distributed

since $f_{A,B}(a,b) = f_A(a)\, f_B(b) \longrightarrow a \cdot f_A(a) \cdot b \cdot f_B(b) = ab\, f_{A,B}(a,b)$

x_1 is a Gaussian r.v $\longrightarrow E\{x_1\} = E\{\underset{\sim}{a}\, t_1 + \underset{\sim}{b}\} = \mu_b$

$$\longrightarrow \sigma_{x_1}^2 = t_1^2 \sigma_a^2 + \sigma_b^2$$

$$x_1 \sim N(\mu_b,\ t_1^2\sigma_a^2 + \sigma_b^2) \qquad \forall t_1$$

2nd Order Pdf

$x_1 = x(t_1)$

$x_2 = x(t_2)$ $\longrightarrow$

$$\begin{bmatrix} \underset{\sim}{x_1} \\ \underset{\sim}{x_2} \end{bmatrix} = \overbrace{\begin{bmatrix} t_1 & 1 \\ t_2 & 1 \end{bmatrix}}^{\underline{\underline{T}}} \begin{bmatrix} \underset{\sim}{a} \\ \underset{\sim}{b} \end{bmatrix}$$

jointly Gaussian vector $\underline{x}$

$$\underset{2\times1}{\underline{\mu_x}} = E\{\underline{x}\} = E\{\underline{\underline{T}}\begin{bmatrix} a \\ b \end{bmatrix}\} = \underline{\underline{T}}\, E\{\begin{bmatrix} a \\ b \end{bmatrix}\} = \underline{\underline{T}}\begin{bmatrix} 0 \\ \mu_b \end{bmatrix} = \mu_b \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\underset{2\times2}{\underline{\underline{C_x}}} = \begin{bmatrix} Cov(x_1,x_1) & Cov(x_1,x_2) \\ Cov(x_2,x_1) & Cov(x_2,x_2) \end{bmatrix}$$

$\longrightarrow \underset{\sim}{a}\, t_l + \underbrace{\underset{\sim}{b} - \mu_b}_{\underset{\sim}{b}_{zm}}$ — zero-mean

$$Cov(x_k, x_l) = E\{(\underset{\sim}{x}(t_k) - \bar{x}(t_k))(\underset{\sim}{x}(t_l) - \bar{x}(t_l))\}$$

$$= E\{(\underset{\sim}{a}\, t_k + \underset{\sim}{b}_{zm})(\underset{\sim}{a}\, t_l + \underset{\sim}{b}_{zm})\} = t_k t_l \sigma_a^2 + \sigma_b^2$$

3rd Order Pdf

$$\begin{matrix} \underset{\sim}{x_1} = \underset{\sim}{x}(t_1) \\ \underset{\sim}{x_2} = \underset{\sim}{x}(t_2) \\ \underset{\sim}{x_3} = \underset{\sim}{x}(t_3) \end{matrix} \longrightarrow \underset{3\times1}{\underline{\mu_x}} = \mu_b \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} , \quad \underset{3\times3}{\underline{\underline{C_x}}} = = \begin{bmatrix} , & , & , \\ Cov(x_2,x_1) & , & , \\ Cov(x_3,x_1) & , & Cov(x_3,x_3) \end{bmatrix}$$

Remember, earlier example:

$$f_{x(t_1),x(t_2)}(x_1,x_2) = f_{x(t_2)|x(t_1)}(x_2|x_1)\, f_{x(t_1)}(x_1)$$

How do we know $\underline{C_x}$ is invertible? (To be Gaussian)

$$f_{\underset{\sim}{x}(t_1),\underset{\sim}{x}(t_2),\underset{\sim}{x}(t_3)}(x_1,x_2,x_3) = f_{\underset{\sim}{x}(t_3)|\underset{\sim}{x}(t_1),\underset{\sim}{x}(t_2)}(x_3|x_1,x_2)\, f_{\underset{\sim}{x}(t_1)\underset{\sim}{x}(t_2)}(x_1,x_2)$$

$$\begin{matrix} x_1 = \underset{\sim}{x}(t_1) = \underset{\sim}{a}t_1 + \underset{\sim}{b} \\ x_2 = \underset{\sim}{x}(t_2) = \underset{\sim}{a}t_2 + \underset{\sim}{b} \end{matrix} \longrightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} t_1 & 1 \\ t_2 & 1 \end{bmatrix} \begin{bmatrix} \underset{\sim}{a} \\ \underset{\sim}{b} \end{bmatrix} \qquad (t_1 \neq t_2)$$

We observe two values, so a and b can be found.

$$\begin{bmatrix} a* \\ b* \end{bmatrix} = \begin{bmatrix} t_1 & 1 \\ t_2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \longrightarrow \underbrace{x_3 = a*t_3 + b*}_{\text{deterministicly found}} \quad \text{NOT GAUSSIAN}$$

$$f_{x(t_3)|x(t_1),x(t_2)}(x_3|x_1,x_2) = \delta(x_3 - a*t_3 - b*)$$

$\downarrow$

C_x will not be invertible (becomes degenerate)

$$\underset{3\times3}{\underline{\underline{C_x}}} = \begin{bmatrix} \sigma_a^2 t_1^2 + \sigma_b^2 & \sigma_a^2 t_1 t_2 + \sigma_b^2 & \sigma_a^2 t_1 t_3 + \sigma_b^2 \\ \sigma_a^2 t_2 t_1 + \sigma_b^2 & \sigma_a^2 t_2^2 + \sigma_b^2 & \sigma_a^2 t_2 t_3 + \sigma_b^2 \\ \sigma_a^2 t_3 t_1 + \sigma_b^2 & \sigma_a^2 t_3 t_2 + \sigma_b^2 & \sigma_a^2 t_3^2 + \sigma_b^2 \end{bmatrix}$$

$$= \sigma_a^2 \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \begin{bmatrix} t_1 & t_2 & t_3 \end{bmatrix} + \sigma_b^2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \longrightarrow \text{Rank} = 2$$

$$\underline{X} = \begin{bmatrix} x(t_1) \\ x(t_2) \\ x(t_3) \end{bmatrix} = \underbrace{\begin{bmatrix} t_1 & 1 \\ t_2 & 1 \\ t_3 & 1 \end{bmatrix}}_{\underline{T}} \underbrace{\begin{bmatrix} \underset{\sim}{a} \\ \underset{\sim}{b} \end{bmatrix}}_{\underline{A}}$$

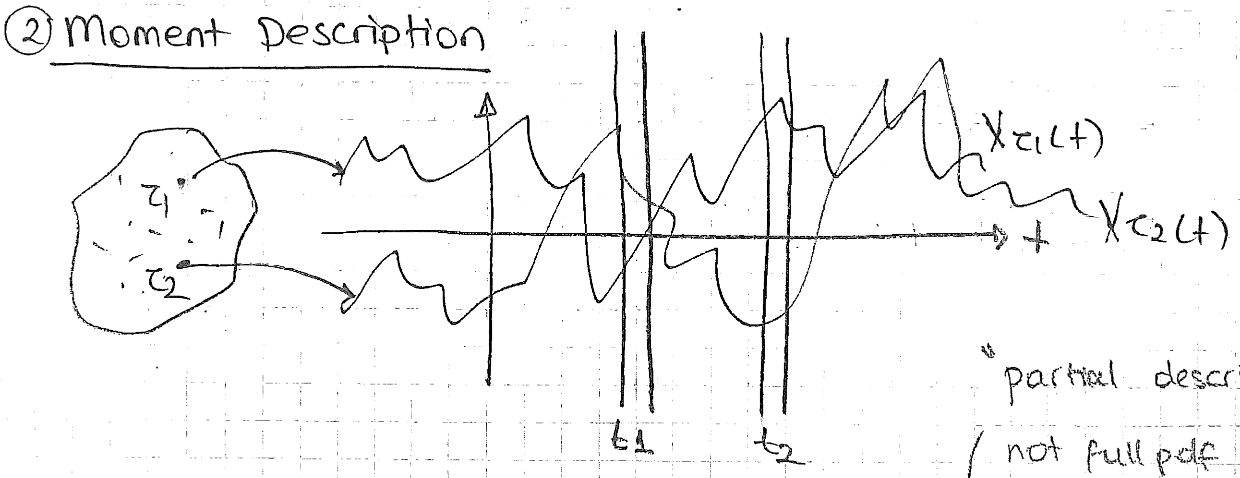
$$\underline{\underline{C}}_X = E\{\underline{X} . \underline{X}^T\} = \begin{bmatrix} t_1 & 1 \\ t_2 & 1 \\ t_3 & 1 \end{bmatrix} \begin{bmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{bmatrix} \begin{bmatrix} t_1 & t_2 & t_3 \\ 1 & 1 & 1 \end{bmatrix} = \underline{\underline{T}}\, \underline{\underline{C}}_A . \underline{\underline{T}}^T$$

In 3D, $[t_1\ t_2\ t_3]$ and $[1\ 1\ 1]$ are independent and there's a 3rd vector which is orthogonal to both of them.

So → null space is at least 1 dimensional.

$\underline{\underline{C}}_X$ is not invertible.

② Moment Description



"partial descriptors"
(not full pdf description!
mean of the pdf)

* First Order Moment Description

$\mu_X(t) = E\{\underset{\sim}{X}(t)\}$ $\forall t$ → mean function

* Second Order Moment Description

$R_X(t_1, t_2) = E\{x(t_1)\, x(t_2)\}$ $\forall t_1, \forall t_2$

Auto-correlation function

$C_X(t_1, t_2) = E\{(x(t_1) - \mu_X(t_1))(x(t_2) - \mu_X(t_2))\}$ $\forall t_1, \forall t_2$

Covariance function

Comments

① Calculation of moment descriptions are much easier than joint pdf descriptions.

② For Gaussian processes, first two moment descriptions (mean function and covariance functions) are sufficient to write joint pdf description for N^{th} order descriptions.

③ In practice, pdf estimation is difficult but moment estimation is much more practical.

Example

Let $\underset{\sim}{x}(t)$ be a r.p. with $\mu_x(t) = 3$ and $R_x(t_1,t_2) = 9+4e^{-0,2|t_1-t_2|}$

$\underset{\sim}{z} \triangleq \underset{\sim}{x}(5)$, $\underset{\sim}{w} \triangleq \underset{\sim}{x}(8)$

Find $E\{\underset{\sim}{z}\}$, $E\{\underset{\sim}{w}\}$, $E\{\underset{\sim}{z}^2\}$, $E\{\underset{\sim}{w}^2\}$, $E\{zw\}$.

$E\{\underset{\sim}{z}\} = E\{\underset{\sim}{x}(5)\} = \mu_x(5) = 3$

$E\{\underset{\sim}{w}\} = E\{\underset{\sim}{w}(8)\} = \mu_x(8) = 3$

$E\{\underset{\sim}{z}^2\} = E\{\underset{\sim}{x}(5)\,\underset{\sim}{x}(5)\} = R_x(5,5) = 13$

$E\{\underset{\sim}{w}^2\} = E\{\underset{\sim}{x}(8)\,\underset{\sim}{x}(8)\} = R_x(8,8) = 13$

$E\{\underset{\sim}{z}\,\underset{\sim}{w}\} = E\{\underset{\sim}{x}(5)\,\underset{\sim}{x}(8)\} = R_x(5,8) = 9+4e^{-0.6}$

Example

$\underset{\sim}{z} = \underset{\sim}{x}(t_1) + \underset{\sim}{x}(t_2)$, Find $E\{\underset{\sim}{z}^2\}$

$E\{(\underset{\sim}{x}(t_1)+\underset{\sim}{x}(t_2))^2\} = E\{\underset{\sim}{x}^2(t_1)\} + 2E\{\underset{\sim}{x}(t_1)\,\underset{\sim}{x}(t_2)\} + E\{\underset{\sim}{x}^2(t_2)\}$

$= R_x(t_1,t_1) + 2R_x(t_1,t_2) + R_x(t_2,t_2)$

Example

$$s = \int_a^b \underset{\sim}{x}(t)\,dt \qquad a)\ E\{\underset{\sim}{s}\} = ? \qquad b)\ E\{\underset{\sim}{s}^2\} = ?$$

(stochastic process)

$$a)\ E\{\underset{\sim}{s}\} = \int_a^b E\{\underset{\sim}{x}(t)\}\,dt = \int_a^b \mu_x(t)\,dt$$

$$b)\ E\{\underset{\sim}{s}^2\} = E\left\{ \int_a^b \underset{\sim}{x}(t)\,dt \cdot \int_a^b \underset{\sim}{x}(t')\,dt' \right\}$$

$$= \int_a^b \int_a^b E\{\underset{\sim}{x}(t)\ \underset{\sim}{x}(t')\}\,dt\,dt'$$

$$= \int_a^b \int_a^b R_x(t,t')\,dt\,dt'$$

Example

$$\underset{\sim}{x}(t) = \underset{\sim}{A}\ \cos(\omega t + \underset{\sim}{\theta})$$

$\underset{\sim}{A}, \underset{\sim}{\theta}$ → independent r.v.s

$\underset{\sim}{\theta}$ → uniform $[0, 2\pi)$

$\underset{\sim}{A}$ → pdf not given

a) $\mu_x(t) = ?$

b) $R_x(t_1, t_2) = ?$

$$a)\ \mu_x(t) = E\{\underset{\sim}{x}(t)\} = E\{\underset{\sim}{A}\ \cos(\omega t + \underset{\sim}{\theta})\}$$

$$= E\{\underset{\sim}{A}\}\ E\{\cos(\omega t + \underset{\sim}{\theta})\}$$

$$= \mu_A \cdot \frac{1}{2\pi} \cancelto{0}{\int_0^{2\pi} \cos(\omega t + \theta)\,d\theta}$$

$$= 0$$

$$b)\ R_x(t_1, t_2) = E\{\underset{\sim}{x}(t_1)\ \underset{\sim}{x}(t_2)\}$$

$$= E\{A^2 \cos(\omega t_1 + \theta)\cos(\omega t_2 + \theta)\}$$

$$= E\{A^2\}\left(\cancelto{0}{E\{\tfrac{1}{2}\cos(\omega(t_1+t_2) + 2\theta)\}} + E\{\tfrac{1}{2}\cos(\omega(t_1 - t_2))\} \right\}'$$

$$= \frac{E\{A^2\}}{2} \cdot \cos(\omega(t_1 - t_2))$$

Notice that RHS is a function of $\Delta = t_2 - t_1$. $R_x(t_1, t_1 + \Delta) = R_x(t_1', t_1' + \Delta)$

L18

Notes (Definitions for complex valued r.p.'s)

If we have complex valued processes, such as $\underset{\sim}{x}(t)=\underset{\sim}{A}\,e^{jwt+\underset{\sim}{\theta}}$,

then $\mu_x(t) \triangleq E\{x(t)\}$ and $R_x(t_1,t_2) \triangleq E\{x(t_1)\,\underline{x^*(t_2)}\}$

second argument has a conjugate.

Example

Same conditions for $\underset{\sim}{A}$ and $\underset{\sim}{\theta}$ as in previous example.

$\underset{\sim}{x}(t)=\underset{\sim}{A}\,e^{j(wt+\underset{\sim}{\theta})}$. Find $\mu_x(t)$, $R_x(t_1,t_2)$.

a) $\mu_x(t)=E\{\underset{\sim}{A}\}\,E\{e^{j(wt+\underset{\sim}{\theta})}\}=E\{\underset{\sim}{A}\}\,E\{\overset{0}{\cancel{\cos(wt+\underset{\sim}{\theta})}}+j\,\overset{0}{\cancel{\sin(wt+\theta)}}\}$

$=0$

b) $R_x(t_1,t_2)=E\{x(t_1)x^*(t_2)\}=E\{\underset{\sim}{A}^2\}\,E\{e^{j(wt_1+\underset{\sim}{\theta})}\,e^{-j(wt_2+\underset{\sim}{\theta})}\}$

$=E\{\underset{\sim}{A}^2\}\,E\{e^{jw\underbrace{(t_1-t_2)}_{\Delta}}\}$

$=E\{\underset{\sim}{A}^2\}\,e^{jw\Delta}$

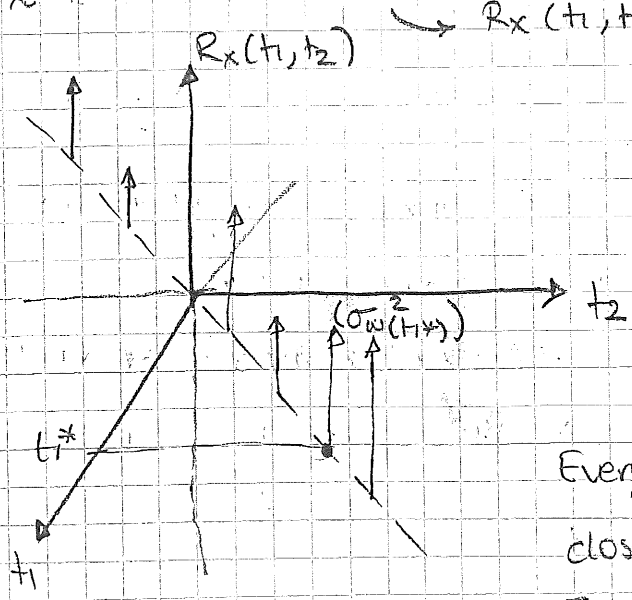
— White Noise Processes —

A process x(t) is called White Noise if its mean function is equal to zero and its autocorrelation is an impulse function.

$\underset{\sim}{x}(t)$: white noise $\rightarrow E\{\underset{\sim}{x}(t)\}=0 \quad \forall t$

$\rightarrow R_x(t_1,t_2)=E\{x(t_1)\,x^*(t_2)\}$

$=\sigma^2_{w(t_1)}\,\delta(t_1-t_2)$



Stationary White Noise:

1) $E\{\underset{\sim}{x}(t)\}=0$

2) $R_x(t_1,t_2)=\sigma_w^2\,\delta(t_1-t_2)$

Every 2 samples (no matter how close they are) are uncorrelated. They are not predictable in a linear sense. $(t_1\neq t_2)$

Example

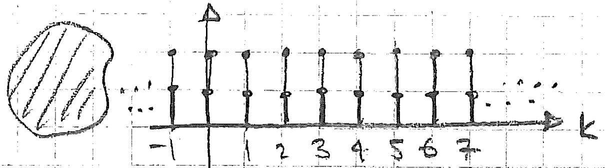
Two random processes $\underset{\sim}{x}_1[k]$ and $\underset{\sim}{x}_2[k]$ are given.

a) $\underset{\sim}{x}_1[k] = \underset{\sim}{w}$ where $\underset{\sim}{w}$ is Gaussian distributed with $N(0, \sigma_w^2)$

b) $\underset{\sim}{x}_2[k] = \underset{\sim}{w}_k$ where $\underset{\sim}{w}_k$ is iid (independent and identically distributed) with $N(0, \sigma_w^2)$

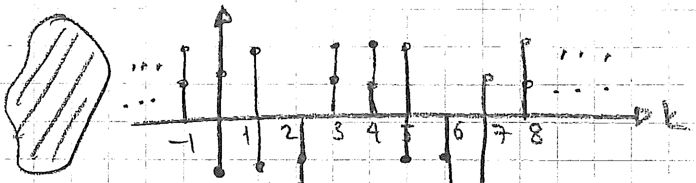
Find pdf, moment descriptions.

$x_1[k] = \underset{\sim}{w}$



remains constant

$x_2[k] = \underset{\sim}{w}_k$



have to have infinite length vector to define this process

First Order Pdf Descriptions

$x_1[k_1] \triangleq x_{k_1}^{①} \rightarrow f_{x_1[k_1]}(x_{k_1}^{①}) \sim N(0, \sigma_w^2)$

$x_2[k_1] \triangleq x_{k_1}^{②} \rightarrow f_{x_2[k_1]}(x_{k_1}^{②}) \sim N(0, \sigma_w^2)$

Second Order Pdf Descriptions

$$f_{x_1[k_1]x_1[k_2]}(x_{k_1}^{①}, x_{k_2}^{①}) =$$

$$\underbrace{f_{x_1[k_1]}(x_{k_1}^{①})}_{N(0,\sigma_w^2)} \cdot \underbrace{f_{x_1[k_2]|x_1[k_1]}(x_{k_2}^{①} | x_{k_1}^{①})}_{\rightarrow \delta(x_{k_2}^{①} - x_{k_1}^{①})}$$

$$= N(x_{k_1}^{①}; 0, \sigma_w^2) \cdot \delta(x_{k_2}^{①} - x_{k_1}^{①})$$

If we know one value, we know all values

$$f_{x_2[k_1]x_2[k_2]}(x_{k_1}^{②}, x_{k_2}^{②}) =$$

iid

$$f_{x_2[k_1]}(x_{k_1}^{②}) \cdot f_{x_2[k_2]}(x_{k_2}^{②})$$

$$= N(x_{k_1}^{②}; 0, \sigma_w^2) \cdot N(x_{k_2}^{②}; 0, \sigma_w^2)$$

Mean Functions (First moments)

$E\{x_1[k]\} = 0 \quad \forall k_1$

$E\{x_2[k_1]\} = 0 \quad \forall k_1$

Autocorrelation functions (Second moments)

$$R_{x_1}(k_1, k_2) = E\{x_1[k_1] x_1^*[k_2]\}$$
$$= E\{|x_1[k_1]|^2\}$$
$$= \sigma_w^2$$

$$R_{x_2}(k_1, k_2) = E\{x_2[k_1] x_2^*[k_2]\}$$

$$= \begin{cases} E\{|x_2[k]|^2\}, & k_1 = k_2 \\ \underbrace{\cancel{E\{x_2[k_1]\}}}_{0} \underbrace{\cancel{E\{x_2^*[k_2]\}}}_{0}, & k_1 \neq k_2 \end{cases}$$

$$= \sigma_w^2 \, \delta[k_1 - k_2]$$

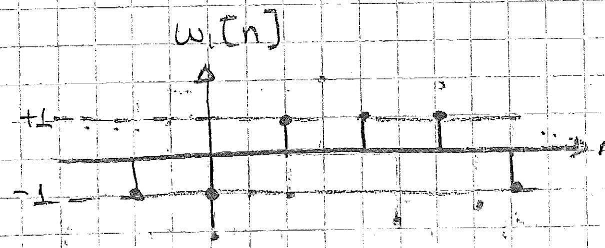
So, $x_2[k]$ is a discrete time white noise process and $x_1[k]$ is not a white noise process since its correlation is constant, i.e. not an impulse function.

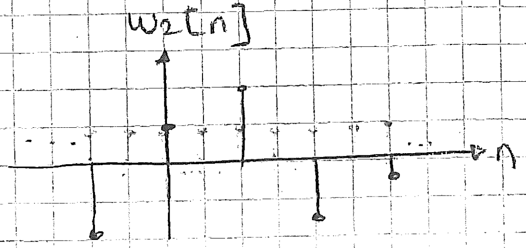
Example

$w_1[n] \in \{\pm 1\}$ iid with equal probability.

$w_2[n] \sim N(0,1)$ iid for every "n".

Find pdf / moment characterizations.





First Order Pdf Descriptions

$$f_{w_1[k_1]}(w_{k_1}) = \frac{1}{2}\delta(w_{k_1}-1) + \frac{1}{2}\delta(w_{k_1}+1)$$

= (Kronecker da olabilirdi)

$$f_{w_2[k_1]}(w_{k_2}) \sim N(w_{k_1}; 0, 1)$$

Second Order Pdf Descriptions

$$f_{w_1[k_1]w_1[k_2]}(w_{k_1}, w_{k_2}) \overset{iid}{=} \underbrace{f_{w_1[k_1]}(w_{k_1})}_{\left(\frac{1}{2}\delta(w_{k_1}-1)+\frac{1}{2}\delta(w_{k_1}+1)\right)} \underbrace{f_{w_1[k_2]}(w_{k_2})}_{\left(\left(\frac{1}{2}\delta(w_{k_2}-1)+\frac{1}{2}\delta(w_{k_2}+1)\right)\right)}$$

$$f_{w_2[k_1]w_2[k_2]}(w_{k_1}, w_{k_2}) \overset{iid}{=} \underbrace{f_{w_2[k_1]}(w_{k_1})}_{N(w_{k_1};0,1)} \cdot \underbrace{f_{w_2[k_2]}(w_{k_2})}_{N(w_{k_2};0,1)}$$

Mean Functions (First Moments)

$E\{w_1[k]\} = 0 \quad \forall k$

$E\{w_2[k]\} = 0 \quad \forall k$

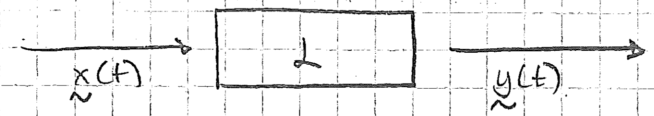
Auto Correlation Functions (Second Moments)

$$R_{w_1}[k_1,k_2] \overset{iid}{=} E\{w_1[k_1]\}\, E\{w_1[k_2]\} = \begin{cases} 1, & k_1 = k_2 \\ 0, & k_1 \neq k_2 \end{cases}$$

$$R_{w_2}[k_1,k_2] \overset{iid}{=} E\{w_2[k_1]\}\, E\{w_2[k_2]\} = \begin{cases} 1, & k_1 = k_2 \\ 0, & k_1 \neq k_2 \end{cases}$$

Both $w_1[n]$ and $w_2[n]$ are discrete time white noise processes, but it should be clear that process outcomes are wildly different from each other. Hence, one more time, we see that moments only give us a partial description. In some discussions, AWGN acronym is used to stand for additive white Gaussian noise processes.

$\boxed{23.11.2020}$



$h(t,\tau)$: response of the system
$h(t-\tau)$: to an impulse at $t=\tau$

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

"linear system definition"

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

"linear and time-invariant system definition" (LTI)

→ convolution integral $(x(t) * h(t))$

$$[y]_k = [\underset{m\times N}{\underline{H}}\, x]_k$$

$$y_k = \sum_{k'=1}^{N} H(k,k')\, x(k') \xrightarrow[\text{invariancy}]{\text{with time}} H(k,k') = \hat{H}(k-k')$$

$$= \begin{bmatrix} h_0 & 0 & 0 & \\ h_1 & h_0 & 0 & \\ \vdots & h_1 & h_0 & \\ \vdots & \vdots & h_1 & \\ \vdots & \vdots & \vdots & \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 0 \\ \vdots \\ \vdots \end{bmatrix}$$

convolution matrix

Toeplitz matrix

$\boxed{L19}$

Linear Systems with Stochastic Inputs

Moment Descriptions

Mean function at the output $\longrightarrow$ $E\{y(t)\} = E\{L\{x(t)\}\} = E\{\int_{-\infty}^{+\infty} h(t,\tau)\, x(\tau)\, d\tau\}$

$$= \int_{-\infty}^{+\infty} h(t,\tau)\, \underbrace{E\{x(\tau)\}}_{\mu_x(\tau)}\, d\tau = L\{\mu_x(t)\} = \mu_y(t)$$

Basic assumption $\longrightarrow$ We assume that anylinear operation and expectation operation can be interchanged, that is

$$E\{L\{x(t)\}\} = L\{E\{x(t)\}\}$$

Auto-correlation function $\longrightarrow$ $R_y(t_1, t_2) = R_{yy}(t_1,t_2) = ?$

$\boxed{\text{Step 1}}$ $R_{xy}(t_1,t_2) = E\{x(t_1)\, y(t_2)\}$ ← cross correlation function

$$= E\{x(t_1)\, L\{x(t)\}\big\downarrow_{t=t_2}\}$$

$$= E\{\underset{\text{fixed}}{\underline{x(t_1)}}\, L\{x(t)\}\}\big\downarrow_{t=t_2}$$

$$= E\{L\{x(t_1)\, x(t)\}\}\big\downarrow_{t=t_2}$$

$$= L\{E\{x(t_1)\, x(t)\}\}\big\downarrow_{t=t_2}$$

$$= L\{R_x(t_1,t)\}\big\downarrow_{t=t_2}$$

$$= \int_{-\infty}^{+\infty} h(t,\tau)\, R_x(t_1,\tau)\, d\tau \big\downarrow_{t=t_2}$$

$$\boxed{R_{xy}(t_1,t_2) = L\{R_x(t_1,t)\}\big\downarrow_{t=t_2} = \int_{-\infty}^{+\infty} h(t_2,\tau)\, R_x(t_1,\tau)\, d\tau}$$

$\boxed{\text{Step 2}}$ $R_y(t_1,t_2) = E\{y(t_1)\, y(t_2)\}$

$$= E\{ L\{x(t)\}\downarrow_{t=t_1} \underbrace{y(t_2)}_{\text{fixed}}\}$$

$$= E\{L\{x(t)\, y(t_2)\}\downarrow_{t=t_1}\}$$

$$= E\{L\{x(t)\, y(t_2)\}\}\downarrow_{t=t_1}$$

$$= L\{E\{x(t)\, y(t_2)\}\}\downarrow_{t=t_1}$$

$$= L\{R_{xy}(t, t_2)\}\downarrow_{t=t_1}$$

$$= \int_{-\infty}^{+\infty} h(t_1,\tau')\, R_{xy}(\tau',t_2)\, d\tau'$$

Combine ① $\xrightarrow{\qquad}$ $= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(t_1,\tau')\, R_x(\tau',\tau)\, h(t_2,\tau)\, d\tau\, d\tau'$

$$\boxed{R_y(t_1,t_2) = L\{R_{xy}(t,t_2)\}\downarrow_{t=t_1} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(t_1,\tau')\, R_x(\tau',\tau)\, h(t_2,\tau)\, d\tau\, d\tau'}$$

Example: $\underset{\sim}{x}(t) \longrightarrow \boxed{d/dt} \longrightarrow \underset{\sim}{y}(t)$ Find $\mu_y(t)$, $R_{yy}(t_1,t_2)$

$\mu_y(t) = L\{\mu_x(t)\} = \frac{d}{dt}\mu_x(t)$

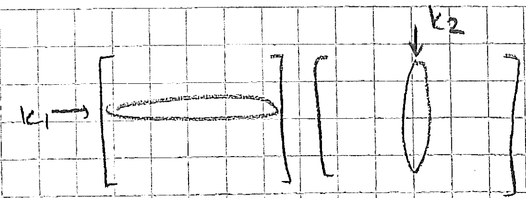
$R_{yy}(t_1,t_2) \Rightarrow$ $R_{xy}(t_1,t_2) = L\{R_x(t_1,t)\}\downarrow_{t=t_2} = \frac{\partial}{\partial t} R_x(t_1,t)\downarrow_{t=t_2}$

$$= \frac{\partial}{\partial t_2} R_x(t_1,t_2)$$

$R_{yy}(t_1,t_2) = L\{R_{xy}(t,t_2)\}\downarrow_{t=t_1} = \frac{\partial}{\partial t}\left(\frac{\partial}{\partial t_2} R_x(t_1,t_2)\right)\downarrow_{t=t_1}$

$$= \frac{\partial^2}{\partial t_1\, \partial t_2} R_x(t_1,t_2)$$

Discrete-time



$\underline{y} = \underline{\underline{H}}\,\underline{x}$ $\qquad \underline{\underline{R}}_y = \underline{\underline{H}}\,\underline{\underline{R}}_x \underline{\underline{H}}^T$

$$[\underline{\underline{R}}_y]_{k_1,k_2} = [\underline{\underline{H}}\,\underline{\underline{R}}_x \underline{\underline{H}}^T]_{k_1,k_2} = \sum_{k'} \underline{\underline{H}}(k_1,k')\,[\underline{\underline{R}}_x\underline{\underline{H}}^T]_{k',k_2}$$

$$= \sum_{k'} \underline{\underline{H}}(k_1,k') \sum_{k''} \underline{\underline{R}}_x(k',k'')\,\underbrace{\underline{\underline{H}}^T(k'',k_2)}_{\underline{\underline{H}}(k_2,k'')}$$

$$\boxed{[\underline{\underline{R}}_y]_{k_1,k_2} = \sum_{k'}\sum_{k''} \underline{\underline{H}}(k_1,k')\,\underline{\underline{R}}_x(k',k'')\,\underline{\underline{H}}(k_2,k'')}$$

* If LTI and Wide sense Stationary $\underset{\sim}{x}(t)$:

$$R_{xy}(t_1,t_2) = \int_{-\infty}^{+\infty} h(\underbrace{t_2-\tau}_{\tau'})\,\hat{R}_x(t_1-\tau)\,d\tau \longrightarrow \text{seems like convolution}$$

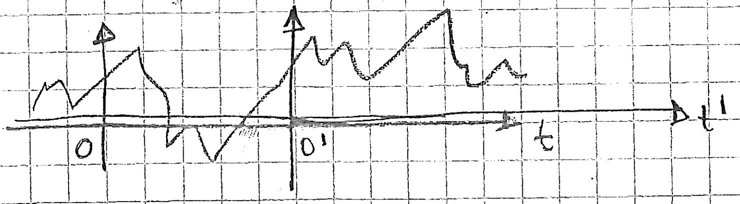
$$= \int_{-\infty}^{+\infty} h(\underbrace{\tau'}_{-\tau''})\,\hat{R}_x(t_1-t_2+\tau')\,d\tau$$

$$= \int_{-\infty}^{+\infty} h(-\tau'')\,\hat{R}_x(t_1-t_2-\tau'')\,d\tau''$$

$$= h(-\gamma) * \hat{R}_x(\gamma) \quad \big\downarrow \; \gamma = t_1 - t_2$$

Stationary Random Processes

A process is called stationary if it has no dependence on initial time.



$$\xrightarrow[\underset{\sim}{x}(t)]{} \boxed{\text{Delay by } \Delta} \xrightarrow[\underset{\sim}{x}(t-\Delta)]{}$$

The delayed process $\underset{\sim}{x}(t-\Delta)$ has the same description with $\underset{\sim}{x}(t)$ for all delays Δ, then the process is a stationary process in the sense of that description (pdf / moment) $\forall \Delta$.

① Stationarity in pdf description

1st Order Stationarity

$$f_{x(t_1)}(x_1) = f_{x(t_1+\Delta)}(x_1) \quad \forall \Delta$$

2nd order Stationarity

$$f_{x(t_1)\,x(t_2)}(x_1, x_2) = f_{x(t_1+\Delta)\,x(t_2+\Delta)}(x_1, x_2) \quad \forall \Delta$$

Nth Order Stationarity

$$f_{x(t_1)\,x(t_2)\ldots x(t_N)}(x_1, x_2, \ldots, x_N) = f_{x(t_1+\Delta)\,x(t_2+\Delta)\ldots x(t_N+\Delta)}(x_1, x_2, \ldots, x_N) \quad \forall \Delta$$

If a process is Nth order stationary for every N, then process is called strict sense stationary.

② Stationarity in moments description

Stationarity in the mean function

$$\left.\begin{array}{l} \mu_x(t_1) = E\{x(t_1)\} \\ \mu_x(t_2) = E\{x(t_2)\} \end{array}\right\} \begin{array}{l} \text{If } \mu_x(t_1) = \mu_x(t_2) \;\; \forall\, t_1, t_2, \\ \text{then } \mu_x(t) \text{ is constant.} \end{array}$$

Stationarity in the autocorrelation function

$$\left.\begin{array}{l} R_x(t_1, t_2) = E\{x(t_1)\,x^*(t_2)\} \\ R_x(t_1+\Delta, t_2+\Delta) = E\{x(t_1+\Delta)\,x^*(t_2+\Delta)\} \end{array}\right\} \text{the same for all } \Delta$$

If $R_x(t_1, t_2) = R_x(t_1+\Delta, t_2+\Delta) \;\; \forall \Delta$, then $x(t)$ is stationary in the autocorrelation.

Observe that, I can set $\Delta = -t_2$,

$$R_x(t_1, t_2) = R_x(t_1 - t_2, 0) = func(t_1 - t_2)$$

$t_1 - t_2$ is the time difference between two sampling instances and it's called the lag of $\tau = t_1 - t_2$.

$$r_x(t_1, \overset{t_2}{\overbrace{t_1 - \tau}}) = E\{x(t_1)\,x^*(t_1 - \tau)\} = func(\tau) = r_x(\tau) \rightarrow \text{autocorrelation function for stationary processes.}$$

$$\boxed{r_x(\tau) \triangleq E\{x(t)\,x^*(t-\tau)\}}$$

L20

Stationarity in joint pdf

$$f_{\underset{\sim}{x}(t_1)\underset{\sim}{x}(t_2)\cdots\underset{\sim}{x}(t_N)}(x_1,x_2,\dots,x_N) =$$

$$f_{\underset{\sim}{x}(t_1+\Delta)\underset{\sim}{x}(t_2+\Delta)\cdots\underset{\sim}{x}(t_N+\Delta)}(x_1,x_2,\dots,x_N)$$

$$\forall \Delta$$

1st order $\longrightarrow f_{\underset{\sim}{x}(t_1)}(x_1) = f_{\underset{\sim}{x}(\underbrace{t_1+\Delta}_{t_2})}(x_1)$

$\forall t_2 \ (\forall \Delta)$

$\boxed{\text{Stationary for all orders} \rightarrow \text{SSS}}$

Stationarity in moments

$\xrightarrow[\text{(mean)}]{\text{1st Order}}$ $E\{x(t_1)\} = E\{x(\overset{t_1+\Delta}{t_2})\} = c$

$\forall t_1, t_2 \ (\forall \Delta)$

$\xrightarrow[\text{(autocorrelation)}]{\text{2nd Order}}$ $E\{x(t_x)\, x^*(t_x-\tau)\}$

$= E\{x(t_y)\, x^*(t_y-\tau)\}$

$\forall t_x, t_y$

$= \text{func}(\tau)$

If we have stationarity in first order and second order moments, then the process is called WSS.

$\boxed{\text{Wide sense stationary}}$

Notes

① SSS → WSS , WSS $\not\rightarrow$ SSS

② If Process is Gaussian and WSS → SSS

example $\underset{\sim}{x}(t) = \underset{\sim}{a}\cos\omega t + \underset{\sim}{b}\sin\omega t$

Find conditions on $\underset{\sim}{a}$ and $\underset{\sim}{b}$ r.v. s.t. $\underset{\sim}{x}(t)$ is WSS.

1) Stationarity in the mean:

$E\{\underset{\sim}{x}(t)\} = c = E\{\underset{\sim}{a}\}\cos\omega t + E\{\underset{\sim}{b}\}\sin\omega t$ constant ∀t.

$$\left.\begin{array}{l} \omega t=0 \rightarrow E\{\underset{\sim}{a}\} \\ \omega t=90 \rightarrow E\{\underset{\sim}{b}\} \\ \omega t=180 \rightarrow -E\{\underset{\sim}{a}\} \end{array}\right\} \text{should be equal} \rightarrow E\{\underset{\sim}{a}\} = E\{\underset{\sim}{b}\} = 0$$

(cosωt and sinωt → linearly independent → constancy can not be satisfied unless $E\{\underset{\sim}{a}\} = E\{\underset{\sim}{b}\} = 0$)

2) Stationarity in Auto-correlation

$$r_x(t_1,t_2) = E\{x(t_1)x^*(t_2)\}$$

$$\left.\begin{aligned} r_x(t_1,t_1) &= E\{x^2(t_1)\} \\ r_x(t_2,t_2) &= E\{x^2(t_2)\} \end{aligned}\right\}$$ should be equal to each other due to WSS ($\tau=0$)

$$\left.\begin{aligned} \omega t_1 = 0 &\longrightarrow r_x(t_1,t_1) = E\{\underset{\sim}{a}^2\} \\ \omega t_2 = \frac{\pi}{2} &\longrightarrow r_x(t_2,t_2) = E\{\underset{\sim}{b}^2\} \end{aligned}\right\} \quad E\{\underset{\sim}{a}^2\} = E\{\underset{\sim}{b}^2\}$$

$$\begin{aligned} r_x(t,t-\tau) &= E\{(a\cos\omega t + b\sin\omega t)(a\cos\omega(t-\tau) + b\sin\omega(t-\tau))\} \\ &= E\{a^2\}(\cos\omega t\cos\omega(t-\tau)) + \overset{E\{a^2\}}{\boxed{E\{b^2\}}}(\sin\omega t\sin\omega(t-\tau)) \\ &\quad + E\{ab\}(\sin\omega t\cos\omega(t-\tau) + \cos\omega t\sin\omega(t-\tau)) \\ &= E\{a^2\}\cos\omega\tau + E\{ab\}\sin(2\omega \underset{\equiv}{t} - \omega\tau) \end{aligned}$$

$\longrightarrow E\{ab\}=0$ for $r_x(t,t-\tau)$ to be a function of τ but not t.

Results

$E\{a\} = E\{b\} = 0$ So, $\underset{\sim}{a}$ and $\underset{\sim}{b}$ are uncorrelated,

$E\{a^2\} = E\{b^2\}$ zero-mean random variables

$E\{ab\} = 0$ with identical variance.

For SSS of the same example, check p. 301 of Papoulis.

Example:

$\underset{\sim}{x}[n]$ is a discrete rp.

At even indices, $\underset{\sim}{x}[2k] \sim \text{unif}[-\sqrt{3}, +\sqrt{3}]$.

At odd indices, $\underset{\sim}{x}[2k+1] \sim N(0,1)$.

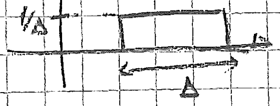
All samples are independent from each other.

Q: Find $\underset{\sim}{x}[n]$ is SSS/WSS or not.

WSS: 1)
$$E\{x[n]\} = \begin{cases} 0, & n: \text{even} \\ 0, & n: \text{odd} \end{cases}$$ $\longrightarrow$ stationarity in the mean.

2) $$E\{x[n]\,x[n-k]\} = \begin{cases} E\{x^2[n]\} & , k=0 \\ \underbrace{E\{x[n]\}}_{0}\,\underbrace{E\{x[n-k]\}}_{0} & , k\neq 0 \end{cases}$$

Since $E\{(x[2n])^2\} = \frac{\Delta^2}{12} = \frac{(2\sqrt{3})^2}{12} = 1$



and $E\{(x[2n+1])^2\} = 1$,

$E\{x[n]\,x[n-k]\} = \delta[k]$

So, $x[n]$ is WSS, (white noise)

SSS: $f_{x[2n]}(x)$, $f_{x[2n+1]}(x)$ } not equal. not even 1st order stationary

Jointly WSS Random Processes

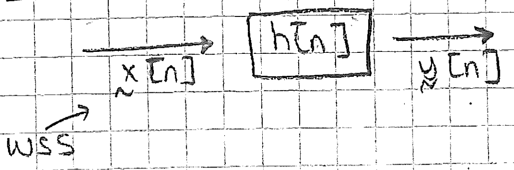
If ① $x[n]$ and $y[n]$ are WSS processes and

② $R_{xy}[n, n-k] = E\{x[n]\,y^*[n-k]\} = func(k) \quad \forall n$,

crosscorrelation

then, $x[n]$ and $y[n]$ are called jointly WSS.

Linear Time-Invariant Processing of WSS Processes



$$y[n] = \sum_{k=-\infty}^{+\infty} h[k]\,x[n-k]$$

$$H(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} h[n]\,e^{-j\omega n}$$

Q: Is $y[n]$ WSS?

$$H(z) = \sum_{n=-\infty}^{+\infty} h[n]\,z^{-n}$$

① Stationarity in the mean

$$E\{y[n]\} = \sum_{k=-\infty}^{+\infty} h[k]\,E\{\underbrace{x[n-k]}_{\mu_x}\} = \mu_x \sum_{k=-\infty}^{+\infty} h[k] = \mu_x H(e^{j\omega})\Big|_{\omega=0} = \mu_x H(1)$$

Since x is WSS, sampling time doesn't matter, μ_x is constant

So, $y[n]$ is stationary in the mean.

② Stationarity in Auto Correlation

$$r_y[n, n-k] = E\{y[n]\, y^*[n-k]\} \overset{?}{=} \text{func}(k)$$

Step 1

$$r_{xy}[n, n-k] = E\{x[n]\, y^*[n-k]\} = E\{x[n] \overbrace{\sum_{k'=-\infty}^{+\infty} h^*[k']\, x^*[n-k-k']}^{y^*[n-k]}\}$$

$$= \sum_{k'=-\infty}^{+\infty} h^*[k'] \underbrace{r_x[n, n-k-k']}_{x \text{ is WSS} \longrightarrow r_x[k+k']}$$

$$= \sum_{k'=-\infty}^{+\infty} h^*[k']\, r_x[k+k']$$

$k'' = -k'$, sum is commutative

$$= \sum_{k''=-\infty}^{+\infty} h^*[-k'']\, r_x[k-k'']$$

$$\boxed{r_{xy}[k] = h^*[-k] * r_x[k]}$$

Step 2

$$E\{y[n]\, y^*[n-k]\} = E\{\underbrace{\sum_{k'=-\infty}^{+\infty} h[k'] \overbrace{x[n-k']\, y^*[n-k]}^{r_{xy}[k-k']}}_{y[n]}\}$$

$$= \sum_{k'=-\infty}^{+\infty} h[k']\, r_{xy}[k-k']$$

$$\boxed{r_y[k] = h[k] * r_{xy}[k]}$$

So, $\underset{\sim}{y}[n]$ is WSS and $\underset{\sim}{x}[n]$, $\underset{\sim}{y}[n]$ are also jointly WSS.

Power Spectral Density

$$r_x[k] \xleftrightarrow{\text{DTFT}} S_x(e^{j\omega})$$

$r_x[k]$ ↑ autocorrelation of a WSS process

$S_x(e^{j\omega})$ ↑ power spectral density of WSS process $\underset{\sim}{x}[n]$

$$\boxed{S_x(e^{j\omega}) = \sum_{k=-\infty}^{+\infty} r_x[k]\, e^{-j\omega k}}$$ → periodic function of ω (2π)

$$\boxed{r_x[k] = \frac{1}{2\pi} \int_0^{2\pi} S_x(e^{j\omega})\, e^{j\omega k}\, d\omega}$$

Notes

① $r_{xy}[k] = r_x[k] * h^*[-k]$

$r_y[k] = r_x[k] * h^*[-k] * h[k]$

② $$\text{DTFT}\{h^*[-n]\} = \left(\left(\sum_{n=-\infty}^{+\infty} h^*[-n]\, e^{-j\omega n}\right)^*\right)^*$$

$$= \left(\sum_{n=-\infty}^{+\infty} h[-n]\, e^{j\omega n}\right)^*$$

$n' = n$

$$= \left(\sum_{n'=-\infty}^{+\infty} h[n']\, e^{-j\omega n'}\right)^*$$

$$= H^*(e^{j\omega})$$

$$S_{xy}(e^{j\omega}) = \text{DTFT}\{r_{xy}[k]\} = S_x(e^{j\omega})\, H^*(e^{j\omega})$$

$$S_y(e^{j\omega}) = \text{DTFT}\{r_y[k]\} = S_x(e^{j\omega})\, H^*(e^{j\omega})\, H(e^{j\omega})$$

$$\boxed{S_y(e^{j\omega}) = S_x(e^{j\omega})\, |H(e^{j\omega})|^2}$$

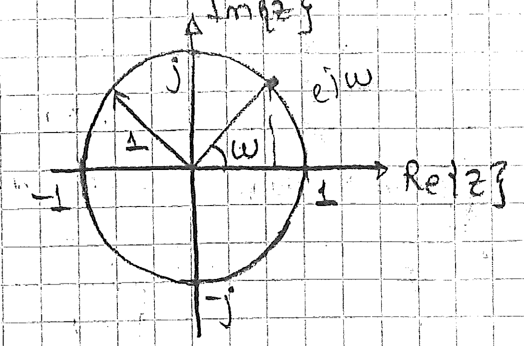
output psd — input psd — filter's magnitude response (squared)

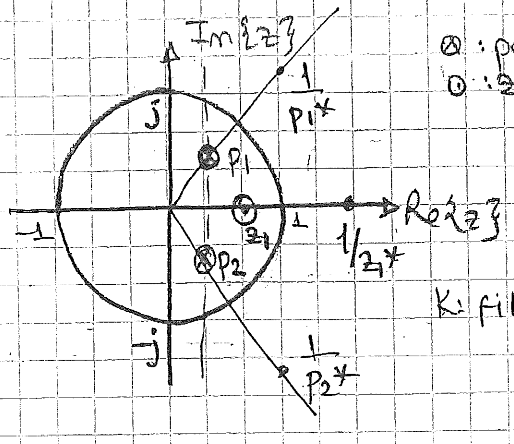
Also, in z-domain:

$\mathcal{Z}\{h^*[-n]\} = H^*\left(\frac{1}{z^*}\right)$ $\qquad H(z) \triangleq \mathcal{Z}\{h[n]\} = \sum_n h[n]\, z^{-n}$

$$S_{xy}(z) = \mathcal{Z}\{r_{xy}[k]\} = S_x(z) \cdot H^*(1/z^*)$$

$$S_y(z) = \mathcal{Z}\{r_y[k]\} = S_x(z) \cdot \underbrace{H(z)\, H^*(1/z^*)}_{H_{comb}(z)}$$





⊗ : pole
⊙ : zero

K: filter gain

pole-zero diagram of H(z)

$z_1 = 1/2 \longrightarrow H(\frac{1}{2}) = 0$ $\quad H(z)\downarrow_{z=\{p_1,p_2\}}$ has a singularity.

if $H(z_k) = 0 \longrightarrow H^*(\frac{1}{z^*})\downarrow_{z=\frac{1}{z_k^*}} = 0$, zero at $z_k \longrightarrow$ zero at $1/z_k^*$

if $H(p_k) = \infty \longrightarrow H^*(\frac{1}{z^*})\downarrow_{z=\frac{1}{p_k^*}} = \infty$, pole at $p_k \longrightarrow$ pole at $1/p_k^*$

$\left(1/p_k^* = \frac{1}{|p_k|} \cdot e^{j\angle p_k}\right)$

So, $H_{comb}(z) = H(z)\,H^*(1/z^*)$ has poles and zeros in conjugate reciprocal pairs.

$1/p_1^*$, $1/p_2^*$, $1/z_1^*$ $\rightarrow$ poles & zeros of $H^*(1/z^*)$

Power Spectral Density- Properties

$$S_x(e^{j\omega}) = \sum_{k=-\infty}^{+\infty} r_x[k] e^{-j\omega k}$$

① * $r_x[k]$ is Hermition symmetric $(r_x[k] = r_x^*[-k])$

$$r_x[k] = E\{x[n]\,x^*[n-k]\} = \left(E\{x^*[n]\,x[n-k]\}\right)^*$$
$$= \left(E\{x[n-k]\,x^*[n]\}\right)^* = r_x^*[-k]$$

* $S_x(e^{j\omega})$ is real valued. $(S_x(e^{j\omega}) = S_x^*(e^{j\omega}))$

* if $r_x[k]$ is real, then it is even. $(r_x[k] = r_x[-k])$

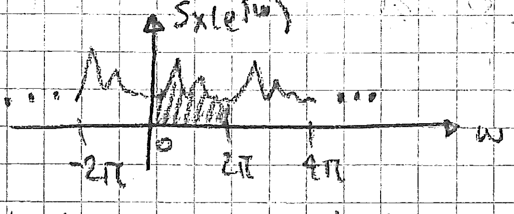
② $S_x(e^{j\omega})$ is always non negative. (More on this later!)

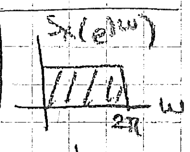
③ "Area" under $S_x(e^{j\omega})$ is the "power" of rp $x[n]$.

"power" of rp $\longrightarrow E\{x^2[n]\} = r_x[0]$

$$r_x[k]\downarrow_{k=0} = \text{DTFT}^{-1}\{S_x(e^{j\omega})\}\downarrow_{k=0} = \frac{1}{2\pi}\int_0^{2\pi} S_x(e^{j\omega})e^{j\omega k}\,d\omega \downarrow_{k=0}$$

$$r_x[0] = \frac{1}{2\pi}\underbrace{\int_0^{2\pi} S_x(e^{j\omega})\,d\omega}_{\text{"Area"}}$$





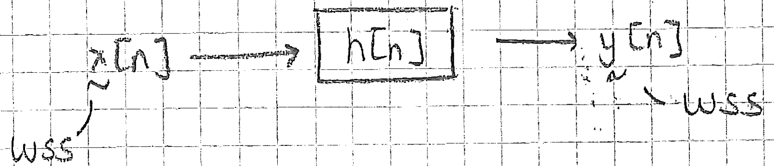
↓

white noise (power is equally distributed)

At peaks → resonant frequency. high power concentration around neighborhoo

L22

30.11.2020



$r_x[k] = E\{x[n]\,x^*[n-k]\}$

$\updownarrow$ DTFT

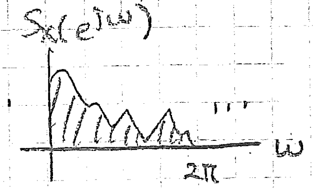
$S_x(e^{j\omega})$: Power Spectral Density

$$\boxed{\begin{array}{c} r_y[k] = r_x[k] * h[k] * h^*[-k] \\ \updownarrow \text{ DTFT} \\ S_y(e^{j\omega}) = S_x(e^{j\omega})\,|H(e^{j\omega})|^2 \end{array}}$$

$S_x(e^{j\omega}) \geq 0$	PSD acts like a density
- real-valued	- PSD ≥ 0
- non-negative	- Real-valued
- $r_x[k] = r_x^*[-k]$	- Area gives total power

$$\underbrace{r_x[0] = \frac{1}{2\pi}\int_0^{2\pi} S_x(e^{j\omega})\,d\omega}_{} \longrightarrow \text{total power across the spectrum}$$

sample value at $k=0$ $\longleftrightarrow$ Area under psd



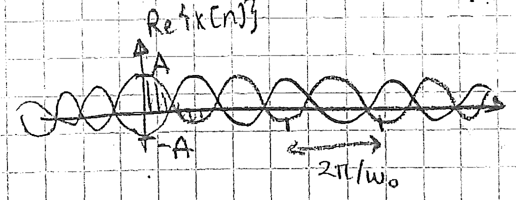
if process is zero-mean $\longrightarrow$ $r_x[0] = \int psd$ becomes variance.

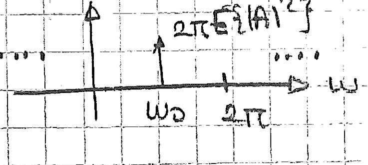
(LP kind of filter)

example $\underset{\sim}{x}[n] = \underset{\sim}{A}\, e^{j(\omega_0 n + \underset{\sim}{\phi})}$ A, ϕ : independent r.v's

Find $S_x(e^{j\omega})$.

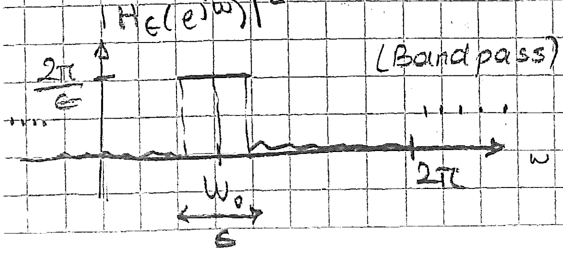
$r_x[k] = E\{|\underset{\sim}{A}|^2\}\, e^{j\omega_0 k}$ $\xleftrightarrow[\text{DTFT}]{}$ $2\pi\, E\{|\underset{\sim}{A}|^2\}\,\delta(\omega - \omega_0) = S_x(e^{j\omega})$

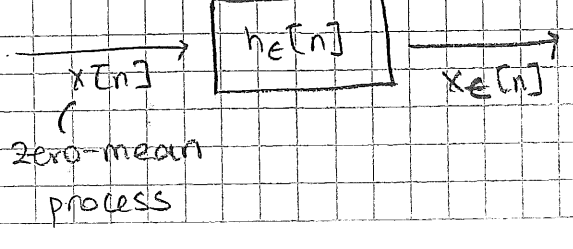




In all of the realizations, $\frac{2\pi}{\omega_0}$ is common. Every realization has the power focused on ω_0.

Proof of $S_x(e^{j\omega}) \geq 0$ let's have a filter $H_\epsilon(e^{j\omega})$ with mag spectrum:





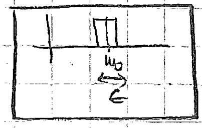
$$r_{x_\epsilon}[k] = r_x[k] * h_\epsilon[k] * h_\epsilon^*[-k] \xleftrightarrow[\text{DTFT}]{} S_{x_\epsilon}(e^{j\omega}) = S_x(e^{j\omega})\,|H_\epsilon(e^{j\omega})|^2$$

$r_{x_\epsilon}[0] = \mathrm{Var}(x_\epsilon[n])$ (since zero-mean)

$$r_{x_\epsilon}[0] = \frac{1}{2\pi}\int_0^{2\pi} S_{x_\epsilon}(e^{j\omega})\,d\omega = \frac{1}{\cancel{2\pi}}\int_{|\omega-\omega_0|\leq\frac{\epsilon}{2}} S_x(e^{j\omega})\,\frac{\cancel{2\pi}}{\epsilon}\,d\omega$$

for small ϵ $\rightarrow \approx \frac{S_x(e^{j\omega})}{\cancel{\epsilon}}\cdot\cancel{\epsilon} = S_x(e^{j\omega}) = r_{x_\epsilon}[0] = \mathrm{Var}(x_\epsilon[n]) \geq 0$

This result is not only a theoretical result of importance, but it is also critical in the interpretation of $S_x(e^{j\omega})$.

$\underset{\sim}{x}[n] \rightarrow$  $\rightarrow \underset{\sim}{y}[n]$ $\quad \mathrm{var}(y[n]) \approx \frac{1}{N}\sum_N |y[n]|^2$

So, $S_x(e^{j\omega})$ gives us "$\underset{(\epsilon \text{ is small})}{\text{finely}}$ filtered" $\underset{\sim}{x}[n]$ process output variance and no surprises $S_x(e^{j\omega}) \geq 0$. Also, no surprises the area under $S_x(e^{j\omega})$ gives the total power of the process $x[n]$.

Some important facts

$\underset{\text{(valid auto-correlation sequence)}}{r_x[k]} \xleftrightarrow[\text{iff}]{} \underset{\sim}{R}_x^{N\times N} \geq 0 \quad \forall N \xleftrightarrow[\text{iff}]{} S_x(e^{j\omega}) \geq 0 \quad \forall\omega$

*) $r_x[k] = r_x^*[-k]$

$$\underline{\underline{R_x}} = E\left\{\begin{bmatrix} x[n] \\ x[n-1] \\ x[n-2]\end{bmatrix}\begin{bmatrix} x^*[n] \\ x^*[n-1] \\ x^*[n-2]\end{bmatrix}^T\right\} = \begin{bmatrix} r_x[0] & r_x[1] & r_x[2] \\ r_x[-1] & r_x[0] & r_x[1] \\ r_x[-2] & r_x[-1] & r_x[0]\end{bmatrix}$$

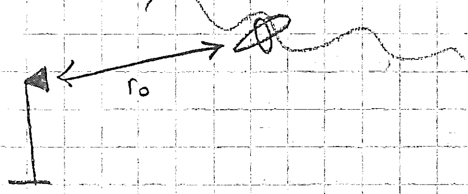
*) $|r_x[0]| \geq |r_x[k]| \quad \forall k$
(for det ≥ 0)

Question Given a positive valued function, can I always construct a random process whose PSD is that function?

(Papoulis, Sec. 10.24, p. $\overset{322}{\cancel{232}}$)

$s(t) = A e^{j\omega_0\left(t - \frac{r(t)}{c}\right)}$, $\quad r(t) = r_0 + \overset{\text{rv}}{\underset{\sim}{v}}\,t$

$s(t)$: continuous variable $\quad$ (c: speed of light)



Given $f_{\underset{\sim}{v}}(v)$, find $S_s(j\omega)$.

$\frac{\text{distance}}{c} \rightarrow$ delay

L23a

$$r_s(\tau) = E\{s(t)s^*(t-\tau)\} = |A|^2 E\{\underbrace{e^{jw_0\left(t-\frac{r(t)}{c}\right)} e^{-jw_0\left(t-\tau-\frac{r(t-\tau)}{c}\right)}}_{e^{jw_0\tau}\, e^{-j\frac{w_0}{c}\underbrace{(r(t)-r(t-\tau))}_{v\tau}}}\}$$

$$r_s(\tau) = |A|^2 e^{jw_0\tau} \underbrace{E_{\underset{\sim}{v}}\{e^{-j\frac{w_0}{c}\tau \underset{\sim}{v}}\}}$$

$$\downarrow$$

$$\int_{-\infty}^{+\infty} e^{-j\frac{w_0}{c}\tau v} f_v(v)\,dv = \int_{-\infty}^{+\infty} e^{j\left(\frac{-w_0\tau}{c}\right)v} f_v(v)\,dv \overset{2\pi}{=} \mathcal{F}^{-1}_{v\to x}\{f_v(v)\}\left(\frac{-w_0\tau}{c}\right)_x$$

$$\Big/\frac{1}{2\pi}$$

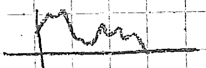
$$S_s(jw) = \mathcal{F}_{\tau\to w}\{r_s(\tau)\} = 2\pi|A|^2 \mathcal{F}_{\tau\to w}\{\underset{Ⓜ}{e^{jw_0\tau}}\, \mathcal{F}^{-1}_{v\to x\to\tau}\{f_v(v)\left(\frac{-w_0\tau}{c}\right)\}\}$$

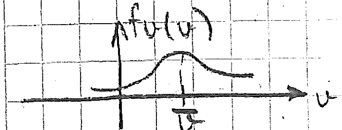
Ⓜ: frequency modulation
↓
shift in the other domain

$$\underset{X(f)}{f_v(v)} \longrightarrow \boxed{\mathcal{F}^{-1}} \xrightarrow{x(t)} \boxed{\text{scaling by } 2} \xrightarrow{x(2t)} \boxed{\mathcal{F}} \longrightarrow X\left(\frac{f}{2}\right)\cdot\frac{1}{|2|}$$

If Ⓜ is absent, then $S_s'(jw) = \frac{|A|^2 c}{w_0} f_v\left(\frac{-c}{w_0}w\right)2\pi$ density function

When Ⓜ is present, then $S_s(jw) = \frac{|A|^2 c}{w_0} f_v\left(\frac{c}{w_0}(w_0-w)\right)2\pi$ → after scaling and shift

So, given a non-negative function , I can form a density as $f_{\underset{\sim}{v}}(v)$ and insert in this example to get a random process whose psd is related with original function -



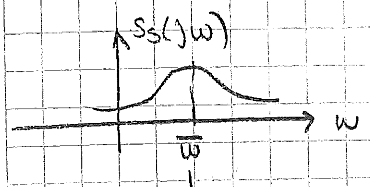
The area under $S_s(jw) = 2\pi\int_{-\infty}^{+\infty} |A|^2 \frac{c}{w_0} f_v\left(\frac{c}{w_0}(w_0-w)\right)dw$

$w_0 - w = w' \rightarrow -dw = dw'$

$\frac{cw'}{w_0} = v \rightarrow \frac{c}{w_0}dw' = dv'$

$$= 2\pi\int_{-\infty}^{+\infty} |A|^2 f_v(v)\,dv$$

$$= |A|^2 2\pi$$



center - doppler frequency due to target motion

$f_v(\bar{v})$ is the peak value of the density. Then $S_s(jw)$ has the peak at $\frac{c}{w_0}(w_0-\bar{w}) = \bar{v} \rightarrow \bar{w} = w_0 - \frac{w_0\bar{v}}{c} = \boxed{w_0\left(1-\frac{\bar{v}}{c}\right)}$ - doppler frequency shift

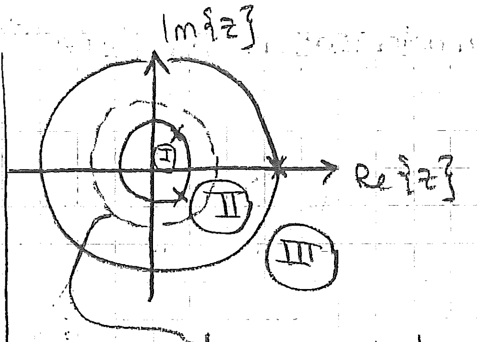
Example (Hayes 3.4.1)

$$x[n] \longrightarrow \boxed{h[n]} \longrightarrow y[n] \qquad H(z) = \frac{1}{1-\frac{1}{4}z^{-1}} \qquad z \in R.O.C$$

$x[n]$ ↳ WSS white noise with variance 1

Find $r_y[k]$.

$$x[n] = \underset{z}{\longleftrightarrow} \sum_{n=-\infty}^{+\infty} x[n] z^{-n}$$



R.O.C. never includes poles

	unit circle	infinity		
Ⓘ	X	X	unstable	anti-causal
Ⓘⓘ	✓	X	stable	
Ⓘⓘⓘ	X	✓	unstable	causal

H(z) has a single pole at 1/4.

$$r_y[k] = \underbrace{r_x[k]}_{\delta[k]} * h[k] * h^*[-k]$$

$$= h[k] * h^*[-k]$$

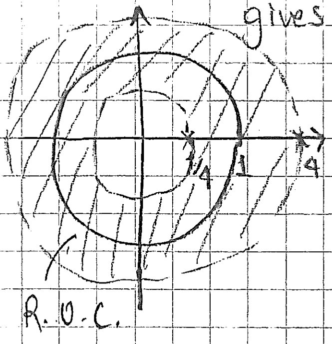
$$S_y(z) = H(z)\, H^*(1/z^*)$$

$$r_y[k] = z^{-1}\{S_y(z)\}$$

$$= z^{-1}\left\{ \frac{1}{1-\frac{1}{4}\underset{z_1}{z^{-1}}} \cdot \frac{1}{1-\frac{1}{4}\underset{z_1^{-1}}{z}} \right\} = z^{-1}\left\{ \frac{-4}{z_1-4} \cdot \frac{z_1}{z_1-\frac{1}{4}} \right\}$$

$$= z^{-1}\left\{ \frac{-4 \cdot 4 \cdot 4/15}{z_1-4} + \frac{-4 \cdot 1/4 \cdot -4/15}{z_1 - 1/4} \right\} = z^{-1}\left\{ \frac{-64/15}{z^{-1}-4} + \frac{4/15}{z^{-1}-1/4} \right\}$$

$$= z^{-1}\left\{ \underbrace{\frac{16/15}{1-\frac{1}{4}z^{-1}}}_{\text{gives causal one}} + \frac{-16/15}{1-4z^{-1}} \right\} = \frac{16}{15}\left(\frac{1}{4}\right)^k u[k] + \frac{16}{15}(4)^k u[-k-1]$$



R.O.C.

1/4 → ROC → infinity → causal

4 → ROC → zero → anticausal

1/4, 4 → conjugate reciprocal pairs

$$z\{a^n u[n]\} = \frac{1}{1-az^{-1}}$$

causal ROC includes ∞

$$z\{-a^n u[-n-1]\} = \frac{1}{1-az^{-1}}$$

anticausal

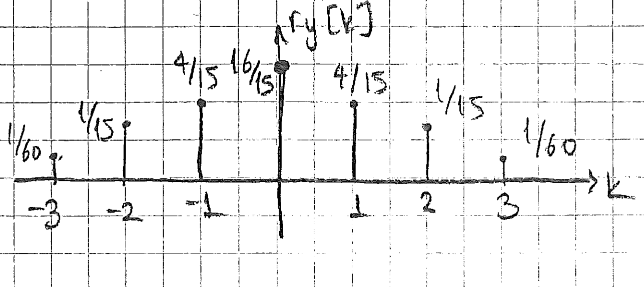
ry[k] is a double sided sequence. (real → even)

$$\underbrace{r_y[k] = r_y[-k]}$$

not causal
not anti-causal

$$r_y[0] = \frac{16}{15} \qquad r_y[1] = r_y[-1] = \frac{4}{15} \qquad r_y[2] = r_y[-2] = \frac{1}{15}$$

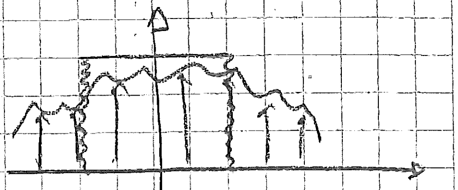
$$\boxed{r_y[k] = \frac{16}{15}\,\frac{1}{4}^{|k|}}$$



A 2nd Characterization for $S_x(e^{j\omega})$

If $\underset{\substack{\downarrow \\ \text{continuum of} \\ \omega(-\pi\leq\omega\leq\pi)}}{X_N(e^{j\omega})} = \underset{\substack{| \\ \text{DTFT}}}{F}\{\underset{\sim}{x}[n]\underset{\substack{| \\ \text{window}}}{w_N[n]}\}$ where $w_N[n] = \begin{cases} 1, & -N\leq n\leq N \\ 0, & \text{o.w.} \end{cases}$)

then $S_x(e^{j\omega}) = \lim_{N\to\infty} \frac{1}{2N+1} \text{Var}(X_N(e^{j\omega}))$



$\underline{X(e^{j\omega_0})} = \sum_{-\infty}^{+\infty} x[n]e^{-j\omega_0 n}$ → weighted sum of $x[n]$ for $\omega=\omega_0$.
do it for all ω_0.

Another rp with process variable ω.

The Releation Between WSS r.p.'s and Fourier Transforms

$$X(e^{j\omega}) = \sum_{-\infty}^{+\infty} \underset{\sim}{x}[n]e^{-j\omega n}.$$

Assume $\underset{\sim}{x}[n]$ is zero-mean and WSS r.p. Find $\text{Cov}(X(e^{j\omega_1}), X(e^{j\omega_2}))$

$$\text{Cov}(X(e^{j\omega_1}), X(e^{j\omega_2})) = E\{X(e^{j\omega_1})X^*(e^{j\omega_2})\}$$

$$= \sum_{n_1}\Big(\sum_{n_2} \underbrace{E\{x[n_1]x^*[n_2]\}}_{r_x[\underbrace{n_1-n_2}_{n_2'}]} e^{+j\omega_2 n_2}\Big)e^{-j\omega_1 n_1}$$

$$= \sum_{n_1}\Big(\sum_{n_2'} r_x[n_2']\, e^{j\omega_2(n_1-n_2')}\Big)e^{-j\omega_1 n_1}$$

$$= \sum_{n_1}\big(S_x(e^{j\omega_2})\, e^{j\omega_2 n_1}\big)\, e^{-j\omega_1 n_1}$$

$$= S_x(e^{j\omega_2}) \underbrace{\sum_{n_1} e^{-j(\omega_1-\omega_2)n_1}}_{2\pi\delta(\omega_1-\omega_2)}$$

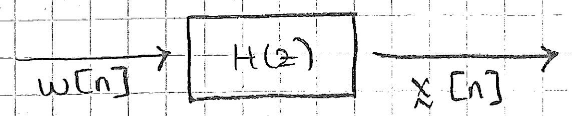
$$= 2\pi S_x(e^{j\omega_2})\,\delta(\omega_1-\omega_2)$$

Fourier Transform is decorrelating every WSS process.

Samples of WSS process in spectrum $X(e^{j\omega_1}), X(e^{j\omega_2})$ are uncorrelated $(\omega_1\neq\omega_2)$. Hence, the process in Fourier domain is non-stationary white noise process where process variable is frequency ω.

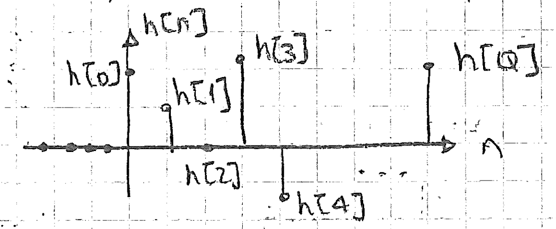
Types of WSS Random Processes

① Moving Average Process (MA Process)



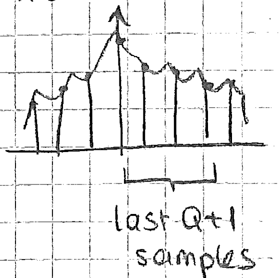
white noise
with variance σ_w^2

H(z) is a causal FIR filter. → 0 for negative indices



$$x[n] = w[n] * h[n] = \sum_{k=0}^{Q} h[k]w[n-k]$$

"weighted sum"



$$= h[0]\,w[n] + h[1]\,w[n-1] + \dots + h[Q]\,w[n-Q]$$

$$r_x[k] = \underbrace{r_w[k]}_{\sigma_w^2\delta[k]} * h^*[-k] * h[k]$$

DTFT $\updownarrow$

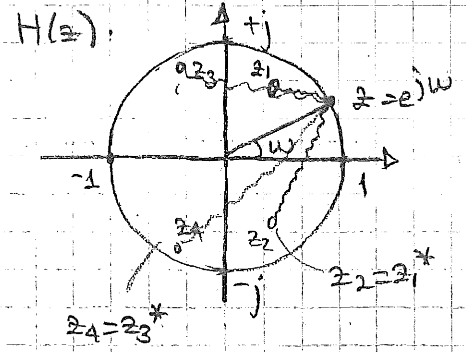
$$S_x(e^{jw}) = \underbrace{S_w(e^{jw})}_{\sigma_w^2}\,|H(e^{jw})|^2$$

Remarks

1) MA filters are all zero filters, they do not have any poles other then $z=0$, $z=\infty$.

2) Pole-zero diagram: H(z) → MA filter.
(to make real-valued impulse response, complex conjugates of zero locations are required.)

H(z):



4-th order system ⟶ Q=4

$$H(z) = h_0 + h_1 z^{-1} + \dots + \underbrace{h_Q}_{h[Q]} z^{-Q}$$

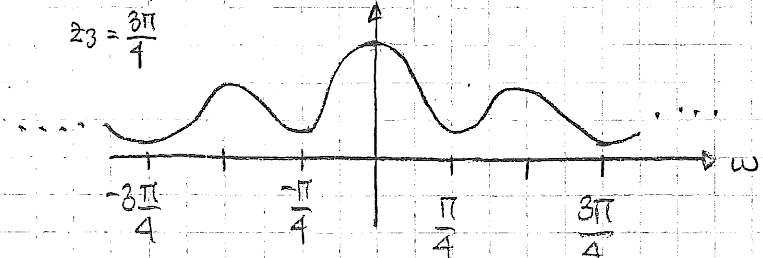
$$= K.\prod_{k=1}^{Q}(z-z_k) \qquad z_k\text{: zeros of } H(z)$$

$$\rightarrow |H(e^{jw})| = |K|.\prod_{k=1}^{Q}|e^{jw}-z_k|$$

Say $\angle z_1 = \frac{\pi}{4}$

$\angle z_3 = \frac{3\pi}{4}$

$|H(e^{jw})|$



Because of zeros around $z=\{e^{j\frac{\pi}{4}}, e^{j\frac{3\pi}{4}}\}$, we observe valleys/dips around these frequencies.

So, for MA processes ; we expect to see dips/valleys in the power spectral density. Next, try to calculate MA process autocorrelation.

$$r_x^{MA}[k] = \sigma_w^2 (h[k] * h^*[-k]) = \sigma_w^2 \sum_l h^*[-l] h[k-l] = \sigma_w^2 \sum_l h^*[l] h[k+l]$$

$$\left(\left(\; r_x^{MA}[k]^* \Big\downarrow_{k \to -k} = \left(\sigma_w^2 \sum_l h^*[l] h[k+l]\right)^* \Big\downarrow_{k \to -k}\right.\right.$$

$$r_x^{MA*}[-k] = \sigma_w^2 \underbrace{\sum_l h[l] h^*[l-k]}_{\text{deterministic autocorrelation}} = r_x^{MA}[k]$$

Deterministic AutoCorrelation

$xcorr(\underline{h}) \longrightarrow \underline{h}[n] = [\cdots 0\ 0\ 0\ \overset{\overset{n}{\downarrow}}{h_0}\ h_1\ h_2\ 0\ 0\ 0 \cdots]$

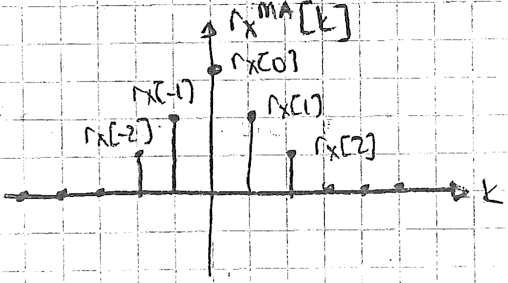
(MATLAB)

$\underline{h}[n-1] = [\cdots 0\ 0\ 0\ 0\ 0\ h_0\ h_1\ h_2\ 0\ 0 \cdots]$

$(Q=2)$

$\underline{h}[n-2] = [\cdots 0\ 0\ 0\ 0\ 0\ 0\ h_0\ h_1\ h_2\ 0 \cdots]$

*)
$$\begin{cases} k=0 \rightarrow r_x[0] = \sigma_w^2 (h_0^2 + h_1^2 + h_2^2) \\ k=1 \rightarrow r_x[1] = \sigma_w^2 (h_1 h_0 + h_2 h_1) \\ k=2 \rightarrow r_x[2] = \sigma_w^2 (h_2 h_0) \\ k=3 \rightarrow r_x[3] = 0 \end{cases}$$

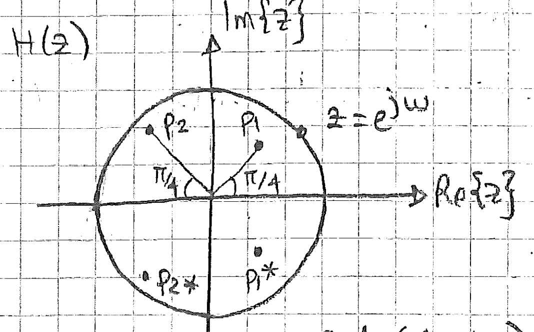


Observations

1) Finding $r_x[k]$ given $h[n]$ is simply by deterministic autocorrelation for MA process.

2) Finding $H(z)$ filter having a desired $r_x^{MA}[k]$ autocorrelation sequence requires solving non-linear equations as in (**) The solution of non-linear equations is more difficult (but can be done by spectral factorization) than linear equation system.
(later on)

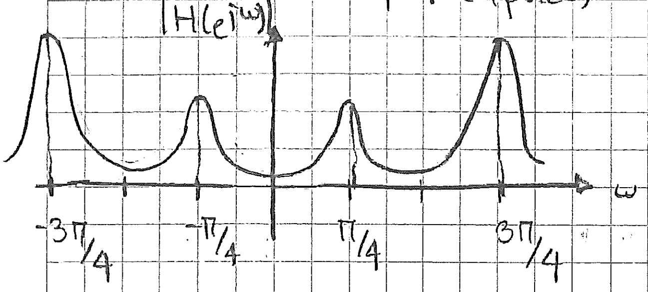
② Autoregressive Processes (AR Process)



$$H^{AR}(z) = \frac{b_0}{1+\sum_{k=1}^{p} a_k z^{-k}}$$ "All pole system"

$$|H(e^{j\omega})| = \frac{|b_0|}{\left|1+\sum_{k=1}^{p} a_k z^{-k}\right|} = \frac{K}{\prod_{k=1}^{p} |z-p_k|}$$

$p=4$ (4 poles)

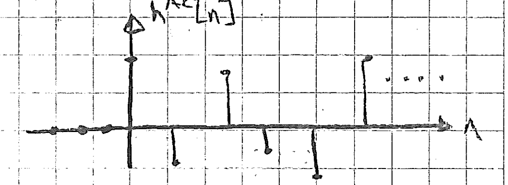


If $e^{j\omega}$ is close to pole locations, we observe a peaky response around $w = w_x$. So, AR processes $S_x(e^{j\omega}) = \sigma_w^2 |H(e^{j\omega})|^2$ is "peaky" and has some "resonance" type shape.

Autocorrelation Calculation for AR Processes

$$r_x[k] = \sigma_w^2 \left(h^{AR}[k] * h^{AR}[k]^*\right) \qquad h^{AR}[k] = z^{-1}\{H^{AR}(z)\}$$

Assumption #1: We assume that $H^{AR}(z)$ corresponds to a causal all-pole filter.



$$\frac{X(z)}{W(z)} = H^{AR}(z) = \frac{b_0}{1+\sum_{k=1}^{p} a_k z^{-k}} \rightarrow \boxed{\text{corresponds to a causal } h^{AR}[n]}$$

z^{-1}

$$x[n] + \sum_{k=1}^{p} a_k x[n-k] = b_0 w[n] \longrightarrow \text{recursion running in forward "n"}$$

$$x[n] + \sum_{k'=1}^{p} a_{k'} x[n-k'] = b_0 w[n] \longrightarrow \text{multiply with } x^*[n-k],\ k \geq 0,\ \text{then } E\{\cdot\}$$

$$E\{x[n]x^*[n-k]\} + \sum_{k'} a_{k'} E\{x[n-k']x^*[n-k]\} = b_0 E\{\underbrace{w[n]}_{\text{new sample}}\ \underbrace{x^*[n-k]}_{\text{interested in old samples}}\}$$

independent (WN)

$$\boxed{r_x[k] + \sum_{k'=1}^{p} a_{k'} r_x[k-k'] = |b_0|^2 \sigma_w^2 \delta[k] \quad (k \geq 0)}$$

$$= \begin{cases} b_0^* \sigma_w^2, & k=0 \\ 0, & k>0 \end{cases} \quad (k \geq 0)$$

$$E\{w[n]\ \underbrace{x^*[n]}_{(b_0 w[n] - \underbrace{\sum_{k'=1}^{p} a_{k'} x[n-k']}_{\text{earlier}})^*}\}$$

$$= E\{b_0^* |w[n]|^2\}$$

$$= b_0^* r_w[0] = b_0^* \sigma_w^2$$

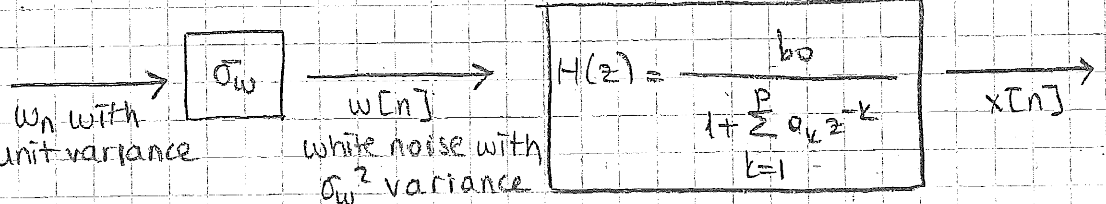
L25

104

07.12.2020

$$\begin{array}{c|cccc} & k'=0 & k'=1 & \cdots & k'=p \\ \hline k=0 & r_x[0] & r_x[-1] & \cdots & r_x[-p] \\ k=1 & r_x[1] & r_x[0] & \cdots & r_x[-p+1] \\ k=2 & r_x[2] & r_x[1] & \cdots & r_x[-p+2] \\ \vdots & \vdots & \vdots & & \vdots \\ k=p & r_x[p] & r_x[p-1] & \cdots & r_x[0] \\ k=p+1 & r_x[p+1] & & & r_x[1] \\ \vdots & \vdots & & & \vdots \end{array} \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} |b_0|^2 \sigma_w^2 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \end{bmatrix}$$

Yule-Walker Equations



① Finding an $H^{AR}(z)$ such that a valid autocorrelation sequence, $r_x^{AR}[k]$ is realized. So, in Yule-Walker equations, we have the auto-correlation matrix but do not know the transfer function coefficients.

For $p=2$ $\rightarrow$

$$\begin{array}{c|ccc} k=0 \rightarrow & r_x[0] & r_x[-1] & r_x[-2] \\ \hline \text{Ⓘ} & r_x[1] & r_x[0] & r_x[-1] \\ & r_x[2] & r_x[1] & r_x[0] \\ \hline & r_x[3] & r_x[2] & r_x[1] \end{array} \begin{bmatrix} 1 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sigma_w^2 |b_0|^2 \\ 0 \\ 0 \\ \vdots \\ \vdots \end{bmatrix}$$

From Ⓘ ($k=1, k=2$ equations) $\rightarrow \begin{bmatrix} r_x[0] & r_x[-1] \\ r_x[1] & r_x[0] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = - \begin{bmatrix} r_x[1] \\ r_x[2] \end{bmatrix}$

Solve a_1, a_2 from Ⓘ and then insert in the $k=0$ equation to get $\sigma_w^2 |b_0|^2$ value.

② Assume we are given $H^{AR}(z)$ and we would like to calculate auto-correlation sequence.

For P=2 case $\longrightarrow$ Assume $r_x[k]$ is real valued $\rightarrow$ $r_x[k] = r_x[-k] \quad \forall k$

$$\begin{bmatrix} 1 & a_1 & a_2 \\ a_1 & (1+a_2) & 0 \\ a_2 & a_1 & 1 \end{bmatrix} \begin{bmatrix} r_x[0] \\ r_x[1] \\ r_x[2] \end{bmatrix} = \begin{bmatrix} \sigma_w^2 |b_0|^2 \\ 0 \\ 0 \end{bmatrix}$$

degree of freedom $\rightarrow$ 3

Solve for $r_x[0], r_x[1], r_x[2]$

To get $r_x[3], r_x[4], \dots$ $\longrightarrow$ $k=3 \rightarrow r_x[k] + \sum_{k'=1}^{P} a_{k'} r_x[k-k'] = \sigma_w^2 |b_0|^2 \delta[k]$

$$r_x[3] + a_1 r_x[2] + a_2 r_x[1] = 0$$

Example

AR(1) process, $x[n] = a\, x[n-1] + w[n]$. Find $r_x[k]$.

$$H(z) = \frac{1}{1-az^{-1}} \longrightarrow$$ causal system running in the forward direction

$x[n] \xrightarrow{x} x[n-k], k>0$

$\downarrow E\{.\}$

$$r_x[k] = a\, r_x[k-1] + E\{\underbrace{\cancel{w[n]x[n-k]}}_{k>0}\}^{0}$$

uncorrelated. $\rightarrow$ Since filtering operation is a causal operation, $x[n-k]$ does not contain any input samples from future ($w[n]$) $x[n-k]$ contains w's up to $n-k$.

$r_x[k] = a\, r_x[k-1] \;, k>0$ $\quad$ $w[n]$ independent $w[n-k]$

$$\left.\begin{aligned} r_x[1] &= a \cdot r_x[0] \\ r_x[2] &= a \cdot r_x[1] = a^2 \cdot r_x[0] \end{aligned}\right\} \; r_x[k] = a^k \cdot r_x[0] \quad k>0$$

$\longrightarrow$ for $k<0 \rightarrow r_x[k] = r_x[-k]$

for $k=0 \rightarrow$?

$$x[n] \xrightarrow{(.)^2} x^2[n] \xrightarrow{E\{.\}} r_x[0] = E\{a^2 x^2[n-1] + w^2[n] + 2a\, \cancel{x[n-1]w[n]}^{\,0}\}$$

$$r_x[0] = a^2 r_x[0] + \sigma_w^2$$

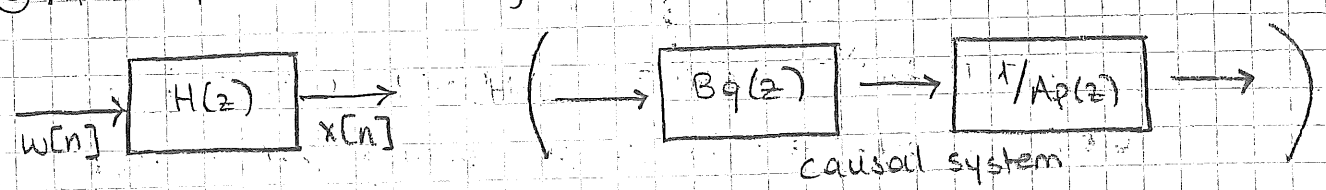
$$r_x[0] = \frac{\sigma_w^2}{1-a^2}$$

$$r_x[k] = a^k \cdot r_x[0] = \frac{\sigma_w^2}{1-a^2} \cdot a^k \quad , k \geq 0$$

Since $r_x[k] = r_x[-k]$,

$$r_x[k] = \frac{\sigma_w^2}{1-a^2}\, a^{|k|} \quad \forall k$$

③ ARMA Processes : Autoregressive and moving Average Processes



Q zeros $\rightarrow B_q(z) = b_q(0) + b_q(1) z^{-1} + b_q(2) z^{-2} + \dots + b_q(Q) z^{-Q}$

P poles $\rightarrow A_p(z) = 1 + a_p(1) z^{-1} + a_p(2) z^{-2} + \dots + a_p(P) z^{-P}$

$$H(z) = \frac{X(z)}{W(z)} = \frac{B_q(z)}{A_p(z)} \qquad x[n] + \sum_{k'=1}^{P} a_p(k') x[n-k'] = \sum_{k'=0}^{Q} b_q(k') w[n-k']$$

$$E\{\cdot\} \leftarrow x^*[n-k]$$

$$r_x[k] + \sum_{k'=1}^{P} a_p(k') r_x[k-k'] = \underbrace{\sum_{k'=0}^{Q} b_q(k') \underbrace{r_{wx}[k-k']}_{\sigma_w^2 h^*[k'-k]}}_{\triangleq C_Q[k]}$$

$$\boxed{\begin{aligned} r_{wx}[k] &= \\ \underbrace{r_w[k]}_{\sigma_w^2 \delta[k]} &* h^*[-k] \\ &= \sigma_w^2 h^*[-k] \end{aligned}}$$

$$C_Q[k] = \sigma_w^2 \sum_{k'=0}^{Q} b_q(k') h^*[k'-k]$$

$$= \sigma_w^2 \sum_{k'=-k}^{Q-k} b_q(k'+k) h^*[k'] = \sigma_w^2 \sum_{k'=0}^{Q-k} b_q(k'+k) h^*[k'] = \begin{cases} \text{non-zero}, & k \leq Q \\ 0, & k > Q \end{cases}$$

(causality) $h[k]$ is the impulse response of the $H^{ARMA}(z)$ system.

$$r_x[k] + \sum_{k'=1}^{P} a_p(k') r_x[k-k'] = C_Q[k]$$

$$\begin{array}{cc}
 & \begin{array}{ccccc} k'=0 & k'=1 & k'=2 & - - - & k'=P \end{array} \\
\begin{array}{c} k=0 \\ k=1 \\ k=2 \\ \vdots \\ k=Q \\ k=Q+1 \\ \vdots \\ k=N \end{array} &
\left[\begin{array}{ccccc}
r_x[0] & r_x[-1] & r_x[-2] & \cdots & r_x[-p] \\
r_x[1] & r_x[0] & r_x[-1] & \cdots & r_x[-p+1] \\
\vdots & \vdots & \vdots & & \vdots \\
\vdots & \vdots & \vdots & & \vdots \\
r_x[Q] & r_x[Q-1] & r_x[Q-2] & \cdots & r_x[Q-p] \\
\hline
r_x[Q+1] & r_x[Q] & r_x[Q-1] & & r_x[Q-p+1] \\
\vdots & & & & \\
r_x[N] & r_x[N-1] & r_x[N-2] & \cdots & r_x[N-p]
\end{array}\right]
\begin{bmatrix} 1 \\ a_p(1) \\ a_p(2) \\ \vdots \\ a_p(P) \end{bmatrix}
=
\begin{bmatrix} C_Q[0] \\ C_Q[1] \\ C_Q[2] \\ \vdots \\ C_Q[Q] \\ 0 \\ \vdots \\ 0 \end{bmatrix}
\end{array}$$

Solution of denominator coefficients ($a_p(k)$ values) is as in AR modeling, but again finding numerator coefficients ($b_q(k)$ values) is difficult. Since by inserting $a_q(k)$ values in the upper part of the dotted line, we can get $c_q[k]$ values but not $b_q(k)$! (We don't know the filter, only denominator values) And solving for $b_q(k)$ from $c_q[k]$ is difficult! So, only in AR modeling, we have simple linear equations in play! MA and ARMA results in non-linear equations in modeling.

④ Periodic (Harmonic) Processes:

Let's remember that $|r_x[k]| \leqslant r_x[0]$ $\qquad E\{\overset{(N+1)\times 1}{\underline{x}}\,\underline{x}^T\} \geqslant 0$ where $\underline{x} = \begin{bmatrix} x[n] \\ x[n+1] \\ \vdots \\ x[n+N] \end{bmatrix}$

① $$\underset{\uparrow k}{[1\ 0 \dots 0 \pm 1\ 0 \dots]} \begin{bmatrix} r_x[0] & r_x[-1] & \dots & r_x[-k] & \dots & r_x[-N] \\ r_x[1] & r_x[0] & & \vdots & & \vdots \\ r_x[2] & r_x[1] & & \vdots & & \vdots \\ \vdots & & & & & \\ r_x[k] & \dots & \dots & r_x[0] & & \vdots \\ \vdots & & & & & \\ r_x[N] & \dots & \dots & \dots & & r_x[0] \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ \pm 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \overset{\leftarrow k}{} \geqslant 0$$

$$2r_x[0] \pm r_x[k] \pm \underbrace{r_x[-k]}_{=r_x[k]} \geqslant 0$$

$$\pm r_x[k] \leqslant r_x[0]$$

$$|r_x[k]| \leqslant r_x[0]$$

② $|E\{zw\}| \leqslant \sqrt{E\{z^2\}\, E\{w^2\}}$ where $z = x[n]$, $w = x[n-k]$ $\rightarrow |r_x[k]| \leqslant r_x[0]$
(Cauchy-Schwarz or $|r_{zw}| \leqslant 1$)

③ $$|r_x[k]| = \left|\frac{1}{2\pi}\int_{-\pi}^{+\pi} S_x(e^{j\omega}) e^{j\omega k}\, d\omega\right| \leqslant \frac{1}{2\pi}\int_{-\pi}^{+\pi} \underbrace{|S_x(e^{j\omega})\, e^{j\omega k}|}_{|S_x(e^{j\omega})| - \text{nonnegative, real}}\, d\omega = \underbrace{\frac{1}{2\pi}\int_{-\pi}^{+\pi} S_x(e^{j\omega})\, d}_{= r_x[0]}$$

triangle inequality.

Question

What happens when $r_x[T] = r_x[0]$ for $T \neq 0$?

Answer

Definition (Mean-Square Periodic Processes): A r.p. is said to be Mean-Square (ms) periodic if $E\{(x[n+T] - x[n])^2\} = 0 \;\forall n, \exists\, T \neq 0$.

Note 1 $E\{(x[n+T]-x[n])^2\} = E\{x^2[n+T]-2x[n]x[n+T]+x^2[n]\}$

$$\overset{\text{wss}}{=} 2r_x[0]-2r_x[-T] = 2r_x[0]-2r_x[T] = 0$$

∴ r.p. is ms periodic $\longleftrightarrow$ $r_x[0]=r_x[T]$ $\exists\, T\neq 0$.

Note 2 Claim: if $r_x[0]=r_x[T]$ (ms periodic r.p.) $\longrightarrow$ $r_x[k]$ is periodic with T. ($r_x[k]=r_x[k+T]$ $\forall k$)

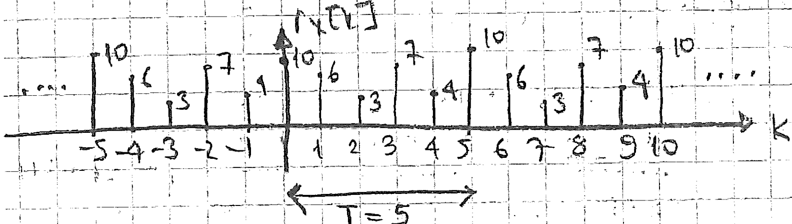
Proof → $|E\{zw\}| \leq \sqrt{E\{z^2\}E\{w^2\}}$ where $z = x[n+k+T]-x[n+k]$

$w = x[n]$

$$\big(\underbrace{r_x[k+T]-r_x[k]}_{=0}\big)^2 \leq \underbrace{(2r_x[0]-2r_x[T])}_{=0}\ \underbrace{r_x[0]}_{\neq 0}$$

$\longleftarrow$

$r_x[k+T]=r_x[k]$ $\forall k$, So, if $r_x[0]=r_x[T]$, $r_x[k]$ is periodic with T.



Then, assume T=3, $r_x[k]=r_x[k+3]$

$$\underline{x} = \begin{bmatrix} x[n] \\ x[n+1] \\ x[n+2] \\ x[n+3] \end{bmatrix}$$

$$\underline{\underline{R}}_x = E\{\underline{x}\,\underline{x}^T\} = \begin{bmatrix} \boxed{r_x[0]} & r_x[-1] & r_x[-2] & \boxed{r_x[-3]} \\ \boxed{r_x[1]} & r_x[0] & r_x[-1] & \boxed{r_x[-2]} \\ \boxed{r_x[2]} & r_x[1] & r_x[0] & \boxed{r_x[-1]} \\ \boxed{r_x[3]} & r_x[2] & r_x[1] & \boxed{r_x[0]} \end{bmatrix}$$

same $\longrightarrow$ $\underline{\underline{R}}_x$ is singular.

Clearly, if $r_x[k]=r_x[k+3]$ $\forall k$, then $\underline{\underline{R}}_x$ with 4x4 dimensions (or higher) is singular, that is $\underline{\underline{R}}_x \geq 0$. $\underline{\underline{R}}_x$ is positive semi-definite but not positive definite. So, there exists a non-trivial vector ($v\neq 0$) s.t. $\underline{\underline{R}}_x \underline{v} = \underline{0}$.

$$\underline{v}^T \underline{\underline{R}}_x \underline{v} = \underline{v}^T E\{\underline{x}\,\underline{x}^T\}\underline{v} = E\{(\underline{x}^T\underline{v})^2\} = 0$$

There's a non-trivial linear combination vector $\underline{v}$ such that one of the samples in $\underline{x}$ vector becomes perfectly predictable from others.

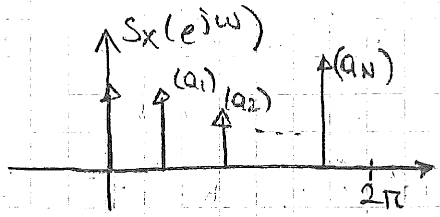
$$0 = [x[n]\ \ x[n+1]\ \ x[n+2]\ \ x[n+3]]\,\underline{v}$$

Finally, if $r_x[k]$ is periodic by T, that is

$$r_x[k] = \sum_{l=0} a_l e^{j\frac{2\pi}{T}lk}$$

$$\updownarrow$$

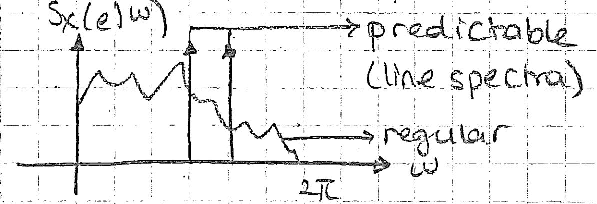
$$S_x(e^{j\omega}) = 2\pi \sum_{l=0} a_l \, \delta(\omega - \frac{2\pi}{T} l)$$



impulsive p.s.d.

Wold's Decomposition Theorem says that an arbitrary process can be written as $x[n] = x_p[n] + x_r[n]$ where $x_p[n]$ is the predictable part and $x_r[n]$ is the regular part where $x_p[n]$ and $x_r[n]$ are orthogonal to each other.

$E\{x_p[m] \, x_r^*[n]\} = 0 \quad \forall m, n.$ (p.107 Hayes)



impulsive → there's some parts making autocorrelation matrix singular (addition of two different autocorrelation matrices, one of them is perfectly predictable. they are uncorrelated, so their autocorrelation matrices add up)

Example

$$x[n] = \sum_{l=1}^{L} \underset{\sim}{a_l} \cdot e^{j(\omega_l n + \underset{\sim}{\theta_l})}$$

$\underset{\sim}{a_l}, \underset{\sim}{\theta_l}$ → independent

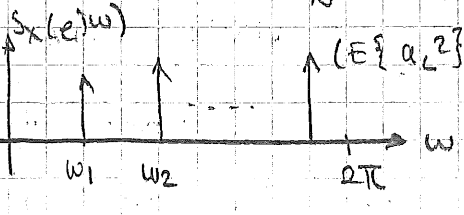
$\underset{\sim}{a_l}$ iid r.v's

$\underset{\sim}{\theta_l}$: unif $[0, 2\pi)$

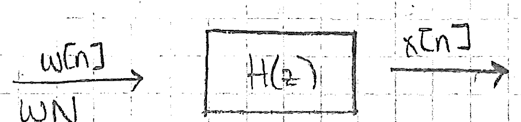
$$r_x[k] = \sum_{l=1}^{L} E\{\underset{\sim}{a_l}^2\} e^{j\omega_l k}$$

$$\updownarrow$$

$$S_x(e^{j\omega}) = \sum_{l=1}^{L} E\{\underset{\sim}{a_l}^2\} 2\pi \, \delta(\omega - \omega_l)$$



Spectral Factorization

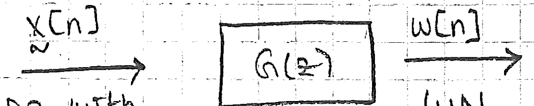
① 

WN

$S_w(e^{j\omega}) = \sigma_w^2 = 1$

Synthesis

- Input: white
- Filter: fixed
- Output: becomes a r.p. with $S_x(z) = H(z)H^*(1/z^*)$

② 

r.p. with $S_x(z) = H(z)H^*(1/z^*)$

WN

Whitening

- Input: x[n] with $S_x(z)$
- Output: white

– How to find G(z) whitening filter given $S_x(z)$?

– It's like decorrelation operation but with infinite dimensional vector.

Example $S_x(e^{j\omega}) = P_x(e^{j\omega}) = \dfrac{5-4\cos\omega}{10-6\cos\omega}$ (always positive → valid)

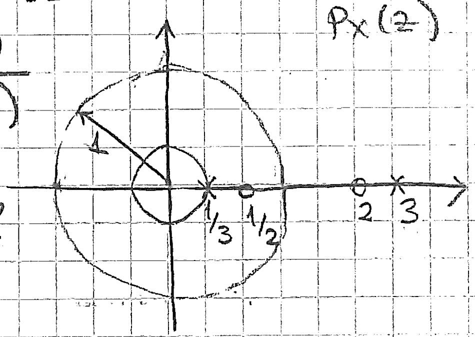
1) Find $H(z)$ when excited with white noise with unit variance. The output process has a psd. as $P_x(e^{j\omega})$ → synthesis.

2) Find $G(z)$ s.t. when $x[n]$ with p.s.d. $P_x(e^{j\omega})$ is at the input of $G(z)$, the output is white noise. → whitening

$$P_x(e^{j\omega}) = \frac{5-4\cos\omega}{10-6\cos\omega} = \frac{5-4\left(\frac{e^{j\omega}+e^{-j\omega}}{2}\right)}{10-6\left(\frac{e^{j\omega}+e^{-j\omega}}{2}\right)} = \frac{5-2(z+z^{-1})}{10-3(z+z^{-1})} \Big\downarrow_{z=e^{j\omega}}$$

$$P_x(z) = H(z)H^*(1/z^*) = \frac{5-2z-2z^{-1}}{10-3z-3z^{-1}} \cdot \frac{-z}{-z} = \frac{2z^2-5z+2}{3z^2-10z+3}$$

$$= \frac{(2z-1)(z-2)}{(3z-1)(z-3)} \cdot \frac{(-z^{-1})}{(-z^{-1})} = \frac{(2z-1)(2z^{-1}-1)}{(3z-1)(3z^{-1}-1)}$$



We want a causal and stable $H(z)$.

$$H(z) = \left\{ \frac{2z-1}{3z-1}, \frac{2z^{-1}-1}{3z^{-1}-1}, \frac{2z-1}{3z^{-1}-1}, \frac{2z^{-1}-1}{3z-1} \right\}$$

poles:	1/3	3	3	1/3
zeros:	1/2	2	1/2	2
	↓ stable & causal outside of 1/3 right-sided Unit circle included	stability and causality are not jointly possible		↓ stable & causal outside of 1/3 right-sided unit circle included.

Note that $H(z) = \dfrac{2z-1}{3z-1}$ has all poles and all zeros inside the unit circle, such filters are called <u>minimum phase filters</u>.

min. phase filters are <u>causal</u>, <u>stable</u> and <u>causally invertible</u>.

$H_{inv}(z) = \dfrac{1}{H(z)} = \dfrac{3z-1}{2z-1}$ → inverse filter has a pole at 1/2, zero at 1/3.

$$\xrightarrow{x[n]} \boxed{H(z)} \xrightarrow{y[n]} \boxed{1/H(z)} \xrightarrow{} z[n] = x[n]$$ } lossless operation, always coming back.

So, $H(z) = \dfrac{2z-1}{3z-1}$ is a possible choice for synthesis filter (min. phase filter). $H(z) = \dfrac{2z^{-1}-1}{3z-1}$ is also a valid choice but its inverse is not stable & causal.

Whitening filter: $\xrightarrow{x[n]} \boxed{G(z)} \xrightarrow{w[n]}$ } $P_x(z) = \dfrac{2z-1}{3z-1} \cdot \dfrac{2z^{-1}-1}{3z^{-1}-1} \longrightarrow G(z) = \dfrac{3z-1}{2z-1}$ min. phase whitening filter.

Up to now, we have worked on finding synthesis filters to realize a given $S_x(e^{jw})$ → Stochastic modeling.

Next, we will try to find filters s.t. the filter impulse response realizes a given sequence → deterministic modeling.

Deterministic Signal Modeling (Hayes)

Previously, (Stochastic signal modeling)

$$\xrightarrow[WN]{w[n]} \boxed{H(z)} \xrightarrow[P_x(e^{jw})]{x[n]}$$

Find $H(z)$ to realize / syntesize the r.p. $x[n]$ with $P_x(e^{jw})$.

Today, Given a sequence $x[n]$, $n \geq 0$ and $x[n]=0$ for $n<0$.

↑ measured impulse response of a causal system.

$$\xrightarrow[\delta[n]]{} \boxed{H(z)} \xrightarrow[h[n]]{}$$

$$H(z) = \frac{b_Q(0) + b_Q(1) z^{-1} + \dots + b_Q(Q) z^{-Q}}{1 + a_P(1) z^{-1} + \dots + a_P(P) z^{-P}}$$

Example:

$$H(z) = \frac{b_0}{1 - a_1 z^{-1}}$$

$\left.\begin{matrix} Q=0 \\ P=1 \end{matrix}\right\}$ $Q+P+1$ unknowns → 2 degrees of freedom.

(causality is always assumed) ⟶ $h[n] = b_0 a_1^n u[n]$

Let's equate first $P+Q+1$ samples of $x[n]$ to $h[n]$.

$x[0] = h[0] = b_0$ $\quad$ $x[1] = h[1] = b_0 a_1$

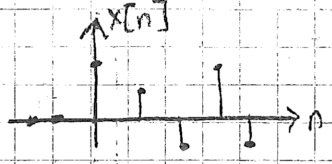
Example:

$$H(z) = \frac{b_0 + b_1 z^{-1}}{1 - a_1 z^{-1}} \overset{z}{\longleftrightarrow} h[n] = b_0 a_1^n u[n] + b_1 a_1^{n-1} u[n-1]$$

$x[0] = h[0] = b_0$ $\quad$ $x[1] = h[1] = b_0 a_1 + b_1$ $\quad$ $x[2] = h[2] = b_0 a_1^2 + b_1 a_1$

Pade's Approximation

$$H(z) = \frac{B_Q(z)}{A_P(z)} = \frac{\sum_{k=0}^{Q} b_Q(k) z^{-k}}{1 + \sum_{k=1}^{P} a_P(k) z^{-k}} \longrightarrow = X(z) \text{ if possible}$$



Assume $X(z) = H(z)$

$$X(z) \cdot A_p(z) = B_Q(z) \xrightarrow{z^{-1}\{\cdot\}} x[n] * \underbrace{a_p[n]}_{1 + a_p(1)\delta[n-1] + a_p(2)\delta[n-2] + \dots + a_p(p)\delta[n-p]} = b_Q[n]$$

$$\underbrace{\begin{array}{r|cccccc|}
 & x[n] & x[n-1] & x[n-2] & - & - & x[n-P] \\
=0\rightarrow & x[0] & 0 & 0 & - & - & 0 \\
=1\rightarrow & x[1] & x[0] & 0 & - & - & 0 \\
=2\rightarrow & x[2] & x[1] & x[0] & - & - & 0 \\
 & \vdots & & & & & \vdots \\
=Q\rightarrow & x[Q] & x[Q-1] & x[Q-2] & - & - & x[Q-P] \\
\hline
=Q{+}1\rightarrow & x[Q+1] & x[Q] & x[Q-1] & - & - & x[Q-P+1] \\
 & \vdots & & & & & \vdots \\
=N\rightarrow & x[N] & x[N-1] & x[N-2] & - & - & x[N-P] \\
 & \vdots & \vdots & \vdots & & & \vdots
\end{array}}_{\text{convolution matrix}}
\begin{bmatrix} 1 \\ a_p(1) \\ a_p(2) \\ \vdots \\ a_p(P) \end{bmatrix}
=
\begin{bmatrix} b_Q[0] \\ b_Q[1] \\ b_Q[2] \\ \vdots \\ b_Q[Q] \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Pade observes that the bottom part of the matrix equation below the dotted line contains only "$a_p(k)$" as unknown. Then, Pade uses the first P equations of the bottom part of the matrix to solve for $a_p(k)$'s. After finding $a_p(k)$'s, insert in the top part to find $b_Q(k)$'s.

$$\underset{P\times P}{\begin{bmatrix} x[Q] & x[Q-1] & - & - & x[Q-P+1] \\ x[Q+1] & x[Q] & & & x[Q-P+2] \\ \vdots & \vdots & & & \vdots \\ x[Q+P-1] & x[Q+P-2] & - & - & x[Q] \end{bmatrix}}
\underset{P\times 1}{\begin{bmatrix} a_p(1) \\ a_p(2) \\ \vdots \\ a_p(P) \end{bmatrix}}
= - \begin{bmatrix} x[Q+1] \\ x[Q+2] \\ \vdots \\ x[Q+P] \end{bmatrix}$$

Pade's method uses only $P+Q+1$ samples of $x[n]$ to set the parameters, but ignores all other $x[n]$ samples.

Example:

$$H(z) = \frac{b_0}{1 + a_1 z^{-1} + a_2 z^{-2}} \qquad \begin{array}{l} P=2 \\ Q=0 \end{array}$$

$$\begin{bmatrix} x[0] & 0 & 0 \\ x[1] & x[0] & 0 \\ x[2] & x[1] & x[0] \end{bmatrix} \begin{bmatrix} 1 \\ a_{p_1} \\ a_{p_2} \end{bmatrix} = \begin{bmatrix} b_0 \\ 0 \\ 0 \end{bmatrix}$$

$$a_p(1) = -\frac{x(1)}{x(0)}$$

$$a_p(2) = \frac{-x(2) + \frac{x(1)^2}{x(0)}}{x(0)}$$

$$b_0 = x(0)$$

Prony's method

Prony uses the bottom part of the same matrix but applies a least squares solution to the bottom part of the matrix equation system:

$$\underline{\underline{X}}_{bot}\begin{bmatrix}1\\ \underline{a_p}\end{bmatrix}=\underline{0}\longrightarrow\Big[\underbrace{\underline{X}^{1}_{bot}}_{\text{1st column of }\underline{\underline{X}}_{bot}}\ \underline{\underline{X}}^{rest}_{bot}\Big]\begin{bmatrix}1\\ \underline{a_p}\end{bmatrix}=\underline{0}$$

$\underline{\underline{X}}^{rest}_{bot}$ → 2nd and all other columns of $\underline{\underline{X}}_{bot}$

$$\underline{\underline{X}}^{rest}_{bot}\,\underline{a_p}=-\underline{X}^{1}_{bot}$$

$$\boxed{\underline{a_p}^{LS}=-\left({\underline{\underline{X}}^{rest}_{bot}}^{T}\,\underline{\underline{X}}^{rest}_{bot}\right)^{-1}{\underline{\underline{X}}^{rest}_{bot}}^{T}\,\underline{X}^{1}_{bot}}$$

→ Prony's estimate → $b_Q(k)$'s are found as in Pade's method.

Let's examine Prony's approach; i.e. LS solution of bottom part in more detail.

error in modeling → $e'[n]=x[n]-h[n]\longrightarrow E'(z)=X(z)-H(z)\rightsquigarrow\dfrac{B_Q(z)}{A_P(z)}$

$$\downarrow$$

$$\underbrace{A_P(z)\,E'(z)}_{E(z)}=X(z)\,A_P(z)-B_Q(z)$$

$E(z)$ → Prony's Error (usual error)

$$e[n]=x[n]*a_p[n]-b_Q[n]$$

$$=\begin{cases}x[n]+\sum\limits_{l=1}^{p}a_p(l)\,x[n-l], & n>Q\\ x[n]+\sum\limits_{l=1}^{p}a_p(l)\,x[n-l]-b_Q[n], & 0\leqslant n\leqslant Q\end{cases}$$

$$J^{Prony}(\underline{a_p})=\sum_{n=Q+1}^{\infty}|e[n]|^2\xrightarrow[\substack{\frac{\partial}{\partial a_p^*(k)}\\ (k\in\{1,\dots,P\})}]{}\frac{\partial J}{\partial a_p^*(k)}=\sum_{n=Q+1}^{\infty}\frac{\partial}{\partial a_p^*(k)}e[n]\,e^*[n]$$

treating $a_p[k]$ and $a_p^*[k]$ as independent variables (Re, Im → 2 degrees of freedom for $a_p[k]$) ⟶

$$=\sum_{n=Q+1}^{\infty}e[n]\overbrace{\frac{\partial}{\partial a_p^*[k]}e^*[n]}^{x^*[n-k]}$$

kth one is somewhere in the summation

$$=\sum_{n=Q+1}^{\infty}\Big(x[n]+\sum_{l=1}^{p}a_p(l)\,x[n-l]\Big)x^*[n-k]$$

$$\boxed{r_x(k,l)\triangleq\sum_{n=Q+1}^{\infty}x[n-l]\,x^*[n-k]}$$

deterministic auto-correlation (there's nothing random)

$$\rightsquigarrow\ =r_x(k,0)+\sum_{l=1}^{p}r_x(k,l)\,a_p(l)$$

Then, calculate $\dfrac{\partial J^{Prony}}{\partial a_p^*(k)}$ for $k=\{1,\dots,p\}$ and equate to zero.

L28

$$\begin{array}{c}\\ k=1\rightarrow \\ k=2\rightarrow \\ \\ \\ k=P\rightarrow\end{array}\begin{array}{cccc} l=1 & l=2 & & l=P \\ \left[r_x(1,1)\right. & r_x(1,2) & \cdots & \left.r_x(1,P)\right] \\ r_x(2,1) & r_x(2,2) & & r_x(2,P) \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ r_x(P,1) & r_x(P,2) & & r_x(P,P) \end{array}\begin{bmatrix} a_p(1) \\ a_p(2) \\ \vdots \\ \vdots \\ a_p(P)\end{bmatrix} = -\begin{bmatrix} r_x(1,0) \\ r_x(2,0) \\ \vdots \\ \vdots \\ r_x(P,0)\end{bmatrix}$$

$$\underline{a}_p^{LS} = -\Big(\underbrace{\underline{\underline{X}}_{bot}^{rest\,T}\,\underline{\underline{X}}_{bot}^{rest}}_{\text{identical}}\Big)^{-1}\underbrace{\underline{\underline{X}}_{bot}^{rest\,T}\,\underline{X}_{bot}^{1}}_{\text{identical}}$$

All Pole Modeling

14.12.2020

$$H(z) = \frac{b_0}{1 + a_p(1)z^{-1} + a_p(2)z^{-2} + \cdots + a_p(P)z^{-P}}$$

$Q=0$, $P=P$

$P+1$ degrees of freedom

$$J^{Prony}_{(\underline{a}_p)} = \sum_{\substack{n=Q+1 \\ \downarrow \\ Q=0}}^{\infty} |e[n]|^2 = \sum_{n=1}^{\infty} |e[n]|^2$$

note that $e[0] = x[0] - b_0$ is not a function of $a_p(k)$'s.

$$\tilde{J}^{Prony}_{(\underline{a}_p)} = \sum_{n=0}^{\infty} |e[n]|^2$$, then $\tilde{J}^{Prony}_{(\underline{a}_p)}$ has the same optimal $\underline{a}_p$ of $J^{Prony}_{(\underline{a}_p)}$

starting from 0, instead of 1

$$r_x(k,l) = \sum_{\substack{n=Q+1 \\ \hookrightarrow n=0}}^{\infty} x[n-l]\,x^*[n-k] \longrightarrow r_x(k+\Delta, l+\Delta) = \sum_{n=0}^{\infty} x[n-l-\Delta]\,x^*[n-k-\Delta]$$

$k \in \{1,\dots,P\}$

$l \in \{1,\dots,P\}$

Δ: integer, $\Delta > 0$

$$= \sum_{n=-\Delta}^{\infty} x[n-l]\,x^*[n-k] = \sum_{n=0}^{\infty} x[n-l]\,x^*[n-k] = r_x(k,l)$$

$x[n]=0$ for $n<0$

$r_x(k,l)$: function of $k-l$

$$\hookrightarrow r_x(k-l) = \sum_{n=0}^{\infty} x[n-l]\,x^*[n-k]$$

$$\begin{bmatrix} r_x(0) & r_x(-1) & \cdots & r_x(-P+1) \\ r_x(1) & r_x(0) & \cdots & r_x(-P+2) \\ \vdots & \vdots & & \vdots \\ r_x(P-1) & r_x(P-2) & \cdots & r_x(0)\end{bmatrix}\begin{bmatrix} a_p(1) \\ a_p(2) \\ \vdots \\ a_p(P)\end{bmatrix} = -\begin{bmatrix} r_x(1) \\ r_x(2) \\ \vdots \\ r_x(P)\end{bmatrix}$$

please compare this equation system with AR(P) Process Yule-Walker equations.

$$r_x[\gamma] = \sum_{n=0}^{\infty} x[n]\,x^*[n-\gamma]$$

(deterministic auto correlation)

In practice, AR(P) Process synthesis, generation of $H(z)$ filter generating AR(P) process with a desired $P_x(e^{jw})$, requires $r_x[k] = E\{x[n]\,x^*[n-k]\}$; but $r_x[k]$ is only estimated by

$$\hat{r}_x[k] = \sum_{n=0}^{\infty} x[n]x^*[n-k] \longrightarrow \text{deterministic auto-correlation}$$

$\hookrightarrow$ observed realization for the random process

Then, using $\hat{r}_x[k]$ instead of $r_x[k]$ in Yule-Walker equations results in a solution identical to the one for all pole modeling

All Pole Modeling With Finite Data Record

with Prony's method

Up to now, we have assumed that $x[n]$ is given for $0 \leq n < \infty$. In practice, $x[n]$ is known for $0 \leq n \leq N$. For finite data records: (finite N)

1). Auto Correlation method:

We use the matrix system as it is in Prony's method with no changes:

$$\begin{bmatrix} x[0] & 0 & 0 & - & - & 0 \\ x[1] & x[0] & 0 & & & \vdots \\ \vdots & \vdots & x[0] & & & \vdots \\ & & \vdots & & & \vdots \\ x[N] & x[N-1] & x[N-2] & & & x[N-p-1] \\ 0 & x[N] & x[N-1] & & & x[N-p] \\ 0 & 0 & x[N] & & & x[N-p+1] \end{bmatrix} \begin{bmatrix} 1 \\ a_p(1) \\ \vdots \\ \vdots \\ a_p(p) \end{bmatrix} = \begin{bmatrix} b_0 \\ 0 \\ 0 \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}$$

We fill all entries without measurement with a zero.

Then apply the usual solution (Prony's)

2) Covariance Method:

No assumptions for $x[n]$ values which are not observed.

$$\begin{array}{l} n=1 \rightarrow \\ n=2 \rightarrow \\ \vdots \\ n=p \rightarrow \\ \vdots \\ n=N \rightarrow \\ n=N+1 \rightarrow \end{array} \begin{bmatrix} x[0] & 0 & - & - & 0 \\ x[1] & x[0] & - & - & 0 \\ \vdots & \vdots & & & \vdots \\ x[p-1] & x[p-2] & - & - & x[-1] \\ & & & & \\ x[N-1] & x[N-2] & - & - & x[N-p-1] \\ x[N] & x[N-1] & & & x[N-p] \end{bmatrix} \begin{bmatrix} a_p(1) \\ a_p(2) \\ \vdots \\ a_p(p) \end{bmatrix} = - \begin{bmatrix} x[1] \\ x[2] \\ \vdots \\ x[p] \\ \vdots \\ x[N] \\ x[N+1] \end{bmatrix}$$

} Covariance Method Sub-Matrix

Covariance method applies LS solution to the sub-matrix system of equations between $n=p$ and $n=N$.

Comments

1) For finite data records, where N is small, covariance method can perform better.

2) Autocorrelation method is guaranteed to give a stable filter as H(z) system.

3) The matrix for autocorrelation method $\underline{X}\begin{bmatrix}1\\ \underline{a}_p\end{bmatrix}=\begin{bmatrix}b_0\\ \underline{0}\end{bmatrix}$ results in $\underline{X}^T\underline{X}$ being positive definite.

Optimization with complex Variables:

$J(z)$: function of a complex variable z. where $z = x+jy$

cost function $\to J(z): \mathbb{C}\to\mathbb{R}$

$J(z_1,z_2) = |z_1-(1+j)|^2 - |z_2|^2$, $J: \mathbb{C}^2\to\mathbb{R}$

↳ a real-valued function of 2 complex variables.

Note: Complex numbers can not be ordered / compared.

$\cancel{1+j > \frac{3}{5}j}$ $\quad |1+j| \geq |\frac{3}{5}j|$ ✓

$$J(z_1,z_2) = |z_1 - (\underbrace{1+j}_{w})|^2 + |z_2|^2 = (z_1-w)(z_1-w)^* + z_2 z_2^*$$

$$\frac{\partial J}{\partial z_1^*}=0 \to \frac{\partial}{\partial z_1^*}(z_1-w)z_1^* = \underbrace{z_1-w}_{\textcircled{I}}=0 \to z_1=w$$

$$\frac{\partial J}{\partial z_2^*}=0\to\frac{\partial}{\partial z_2^*}z_2z_2^* = \underbrace{z_2}_{\textcircled{II}}=0$$

$z_1, z_2 \to$ 4 degrees of freedom

but we're taking derivatives for 2 of them.

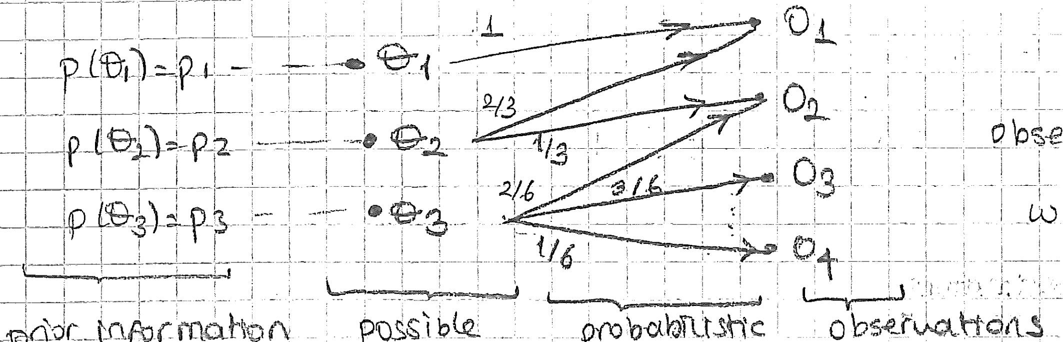
$$\frac{\partial J}{\partial z_1}=0\to \frac{\partial}{\partial z_1}(z_1-w)^*z_1 = (z_1-w)^*=0=\textcircled{I}^*$$

$$\frac{\partial J}{\partial z_2}=0\to\frac{\partial}{\partial z_2}z_2z_2^*=z_2^*=0=\textcircled{II}^*$$

already satisfied.

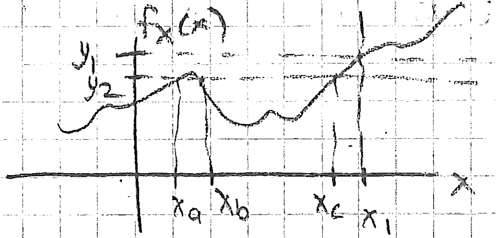
"Properly Handling Complex Differentiation in Optimization and Approximation Problems", IEEE Signal Processing Magazine. Lecture Notes, March 2019.

Estimation



observation : O_3

what is θ ?



Classes of Estimation Problems

non-random parameter estimation

θ: non-random

no prior information

↓

maximum likelihood (approach)

no optimality guarantee

- likelihood : $f(\text{obs}|\theta)$

$$\hat{\theta}_{ML} = \underset{\theta}{\text{argmax}}\ f(\underset{\underset{\text{fixed}}{\downarrow}}{\text{obs}}|\theta)$$

random parameter estimation

θ: random variable

i.e. assumes a prior information

(distribution) for $\underset{\sim}{\theta}$ r.v.

Once you see the observation, you can update your priori information to posterior information

The goal is, in general, to get the posterior distribution.

$f_{\underset{\sim}{\theta}}(\theta)$: prior $f_{\underset{\sim}{\theta}|\text{obs}}(\theta|\text{obs})$: posterior.

Example (non-random par. est.)

$x[n] = c + w[n]$, c: non-random parameter

$n = \{1, \dots, N\}$ $w[n]$: iid $N(0, \sigma_n^2)$

$x[n]$ is provided as observations. Find maximum likelihood estimate of c.

$$\underbrace{f(x_1, x_2, \dots, x_n}_{\underset{\sim}{x}}; c) = \prod_{k=1}^{N} f(x_k; c) = \prod_{k=1}^{N} \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left\{-\frac{(x_k-c)^2}{2\sigma_n^2}\right\}$$

use ; instead of | because RHS contains non-random variables.

$$\hat{c}_{ML} = \underset{c}{\text{argmax}}\ \underbrace{f(\underset{\sim}{x}; c)}_{\text{likelihood}} = \underset{c}{\text{argmax}}\ \underbrace{\log f(x; c)}_{\text{log likelihood}}$$

$$\hat{c}_{ML} = \underset{c}{\text{argmax}} \left(\text{constant} - \underbrace{\sum_{k=1}^{N} \frac{(x_k-c)^2}{2\sigma_n^2}}_{J(c)}\right)$$

$$\hat{c}_{ML} = \underset{c}{\text{argmax}}\ -J(c) = \underset{c}{\text{argmin}}\ J(c)$$

$$\frac{\partial}{\partial c} J(c) = 0 \rightarrow \frac{d}{dc}\sum_{k=1}^{N}\frac{(x_k-c)^2}{2\sigma_n^2} = -\sum_{k=1}^{N}\frac{2(x_k-c)}{2\sigma_n^2} = 0 = \sum_{k=1}^{N} x_k - c = 0$$

$$N\cdot c = \sum_{k=1}^{N} x_k \qquad \boxed{\hat{c}_{ML} = \frac{1}{N}\sum_{k=1}^{N} x_k} \rightarrow \text{sample mean}$$

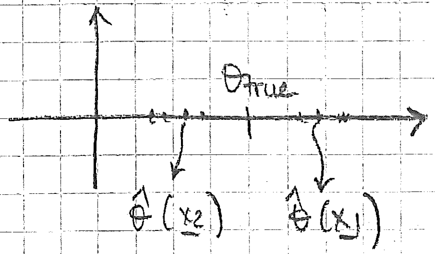
Properties of Estimators

1) Bias An estimator is called unbiased if $E_{\underline{x}}\{\hat{\theta}(\underline{x})\} = E_\theta\{\theta\}$ where $\hat{\theta}(x)$: estimator for θ and $\underline{x}$: observation vector.

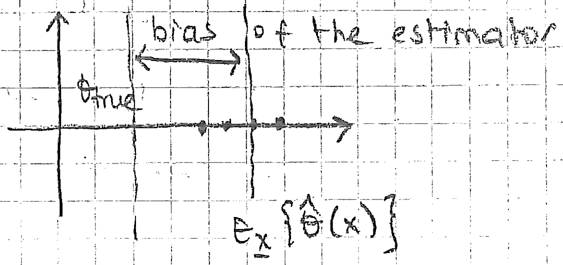
In the previous ML example, c is non random. Then, $E\{\hat{c}(\underline{x})\} = c$ for an unbiased estimator. Let's check whether the sample mean is an unbiased estimator or not.

$$E\{\hat{c}(\underline{x})\} = E\left\{\sum_{k=1}^{N}\frac{x_k}{N}\right\} = E\left\{\sum_{k=1}^{N}\frac{(c+w_k)}{N}\right\} = \frac{c\cdot N}{N} = c$$

We see that sample mean is an unbiased estimator.



on the average, estimation values matches the true value.



In some problems, bias can be a fixed quantity (independent of θ_{true}). Then, you can subtract bias from your estimates. (i.e. you generate a new estimator $\hat{\theta}_2(x) = \hat{\theta}_1(x) - \text{bias}(\hat{\theta}_1)$) Then, the new estimates are unbiased.

2) Consistency An estimator is consistent if $E\{(\theta - \hat{\theta}(\underline{x}))^2\} \rightarrow 0$ as the number of observations (N) goes to infinity ($N \rightarrow \infty$) where $\underline{x}$ is $N\times 1$. (if error is zero mean, estimator is unbiased, so this is the error variance. If error variance goes to zero as N goes to infinity → estimator is consistent).

Let's check consistency of sample mean example.

estimation error $\rightarrow \theta - \hat{\theta}(\underline{x}) = c - \frac{\sum_k x_k}{N} \rightarrow E_{\underline{x}}\left\{\left(c - \frac{\sum_k x_k}{N}\right)^2\right\}$

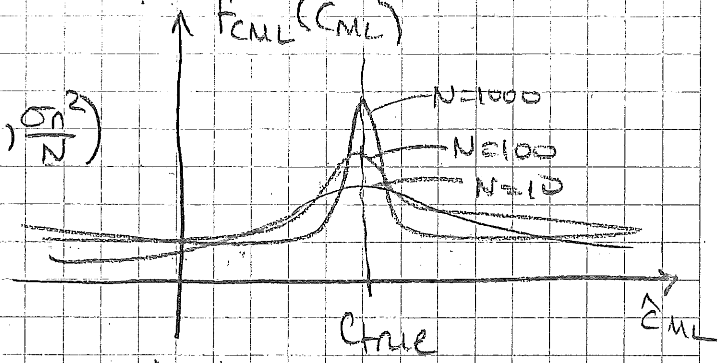
$$= E_{\substack{w_1 \\ w_2 \\ \vdots \\ w_N}}\left\{\left(c - \frac{\sum_k (c+w_k)}{N}\right)^2\right\} = E_{\underline{w}}\left\{\left(-\frac{\sum_k w_k}{N}\right)^2\right\} \overset{w_k:\text{iid}}{=} \frac{\sum_k E\{w_k^2\}}{N^2} = \frac{\sigma_n^2}{N}$$

So, $E\{(\theta - \hat{\theta}(\underline{x}))^2\} = \frac{\sigma_n^2}{N} \rightarrow 0$ as $N \rightarrow \infty$. Sample mean is a consistent estimator.

Sample mean estimator:

$$\hat{c}_{ML}(\underline{x}) = \frac{\sum_{k=1}^{N} x_k}{N} \sim \mathcal{N}\left(c_{true}, \frac{\sigma_n^2}{N}\right)$$

$$x_k \sim (c_{true}, \sigma_n^2)$$



3) Efficiency An estimator is said to be efficient (statistically efficient) if it is unbiased and its error variance is equal to the Cramer-Rao Bound (CRB).

$$\text{MSE} = \underbrace{E\{\underbrace{(\theta - \hat{\theta}(x))^2}_{\text{error}^2}\}}_{\text{mean(error}^2)} \geq \text{CRB}(\theta)$$

$$\text{CRB}(\theta) = \frac{1}{E_{\underline{x}}\left\{\left(\underbrace{\frac{\partial}{\partial\theta} \ln f(x;\theta)}_{\text{log likelihood}}\right)^2\right\}}$$

Let's check efficiency of sample mean estimator.

$$\log(f(x;\theta)) = \text{constant} - \frac{\sum_k (x_k - c)^2}{2\sigma_n^2}$$

$\frac{d}{d\theta} = \frac{d}{dc}$

$$+ \frac{\cancel{2}}{\cancel{2}\sigma_n^2} \sum_{k=1}^{N} (x_k - c) \xrightarrow{E\{(.)^2\}} E_{\underline{x}}\left\{\left(\frac{d}{dc} \log f(x;c)\right)^2\right\} = E_{\underline{x}}\left\{\frac{(\sum \overset{w_k}{(x_k - c)})^2}{\sigma_n^4}\right\}$$

$$= \frac{E_{\underline{w}}\{(\sum_k w_k)^2\}}{\sigma_n^4} \overset{\text{iid}}{=} \frac{N \cdot \sigma_n^2}{\sigma_n^4} = \frac{N}{\sigma_n^2} \qquad \text{So, CRB}(c) = \frac{\sigma_n^2}{N}$$

So, comparing MSE of sample mean estimator $\left(\frac{\sigma_n^2}{N}\right)$ and CRB of the problem, we conclude that sample mean is an efficient estimator.

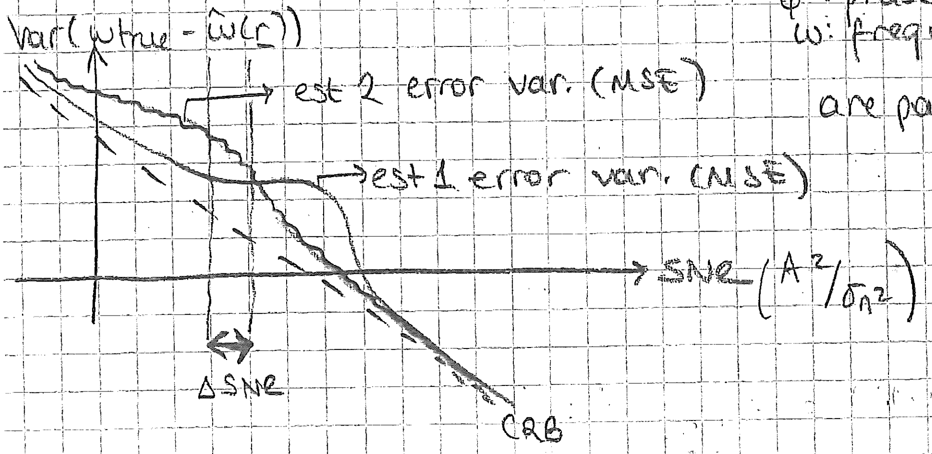
L30

n general, CRB depends on unknown variable ~~θ~~ and for other problems there can be several parameters and the joint estimation of several parameters of interest can be an issue.

Example

$$\underline{r} = Ae^{j\phi}e^{jwn} + \underset{\sim N(0,\sigma_n^2 I)}{\underline{w}}$$

A: amplitude
φ: phase
ω: frequency
} of complex exponential are parameters of interest.



at high SNR's, estimators meet at the CRB.

If an estimator meets CRB at high SNR (SNR→∞) or for large number of observations (N→∞), then such an estimator is called an asymptotically efficient estimator.

(Sample mean is always on the bound)

Signal Processing task: increasing SNR such that 2nd estimator will be performing better.

Folk Theorem: ML estimate is an asymptotically unbiased and efficient estimator.

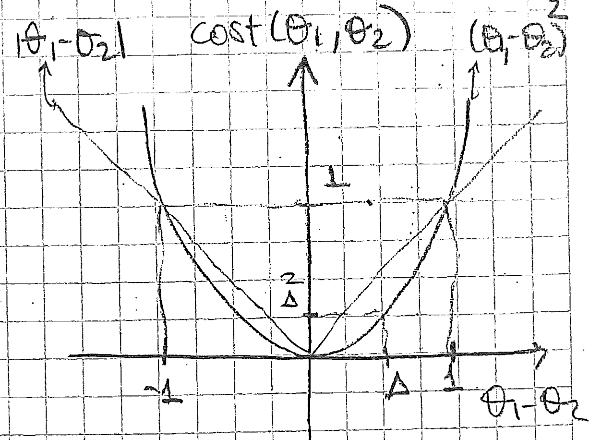
Random Parameter Estimation

$\underset{\sim}{\theta}$: r.v. $f_{\underset{\sim}{\theta}}(\theta) \rightarrow$ prior distribution

$\underline{x}$: observation vector.

The goal is to minimize

$E\{cost(\theta, \hat{\theta}(\underline{x}))\} = R =$ risk



Let's assume square error is selected as the cost function.

$$R = \underset{\theta, \underline{x}}{E}\{(\theta - \hat{\theta}(\underline{x}))^2\}$$

The goal is finding $\hat{\theta}(x)$ s.t. risk is minimized.

$$R = \underset{\underline{x}}{E}\left\{ \underset{\underset{\sim}{\theta}|\underline{x}}{E}\left\{ (\theta-\hat{\theta}(\underline{x}))^2 \right\}\right\} = \int_{\underline{x}} f_{\underline{x}}(\underline{x}) \left[\int_{\theta} (\theta-\hat{\theta}(x))^2 \underbrace{f_{\underset{\sim}{\theta}|\underline{x}}(\underline{\theta}|\underset{\sim}{\underline{x}})}_{\text{posterior density}} d\theta \right] d\underline{x}$$

x is constant in this paranthesis.

Let's minimize $I(\theta,\hat{\theta})$ for a given $\underline{x}$.

$\hat{\theta}(\underline{x}) = c$

$$I(\theta,\hat{\theta}) = \int (\theta - \overbrace{\hat{\theta}(\underline{x})}^{c})^2 f_{\theta|\underline{x}}(\theta|\underline{x})\, d\theta$$

$I(\theta,\hat{\theta}) \rightarrow I(\theta,\hat{\theta}(\underline{x}))$

final risk is the mean value of this function.

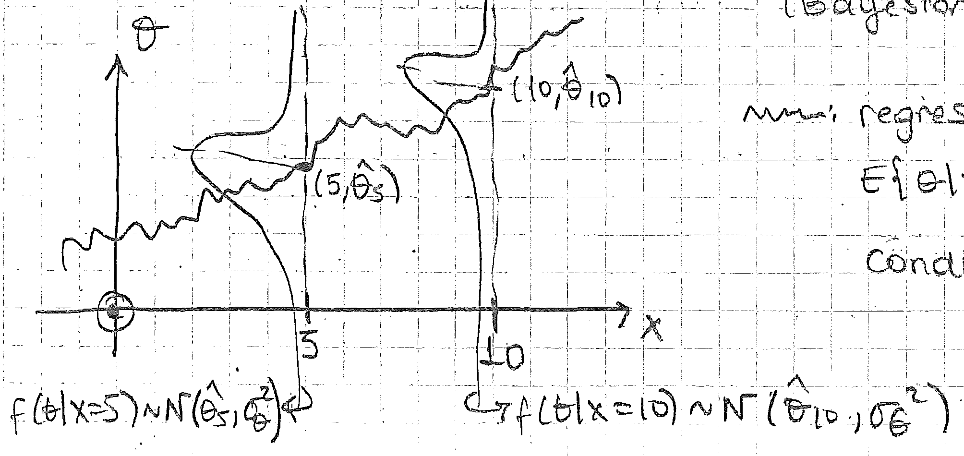
$\frac{d}{dc}$

$$-2\int (\theta-c) f_{\theta|x}(\theta|x)\, d\theta = 0$$

$$c = \int_{-\infty}^{+\infty} \theta f_{\theta|x}(\theta|x)\, d\theta = \underbrace{\underset{\theta|\underline{X}=x}{E}\{\theta \mid \underline{X}=x\}}_{\text{conditional mean.}}$$

$$\boxed{\hat{\theta}(\underline{x}) = E_{\theta|\underline{X}=x}\{\theta \mid \underline{X}=x\}} \rightarrow \text{optimal estimator minimizing mse.}$$

(Bayesian estimator)



∿∿ : regression line

$E\{\theta|x\} = \varphi(x) = \hat{\theta}(x)$

conditional mean estimator.

Properties of Conditional Mean Estimator

① If we want to estimate θ, β, γ, etc. (more than one r.v) from $\underline{x}$, then individually minimize mse for θ, β, γ (estimation error).

$$\hat{\theta}(\underline{x}) = E_{\theta|\underline{x}}\{\theta|\underline{x}\} \quad \hat{\beta}(\underline{x}) = E_{\beta|\underline{x}}\{\beta|\underline{x}\} \quad \hat{\gamma}(\underline{x}) = E_{\gamma|\underline{x}}\{\gamma|\underline{x}\}$$

If we define $\underline{\theta} \triangleq \begin{bmatrix}\theta\\ \beta\\ \gamma\end{bmatrix}$, $\boxed{E_{\underline{\theta}|\underline{x}}\{\underline{\theta}|\underline{X}=\underline{x}\} = \hat{\underline{\theta}}}$ → conditional mean vector for desired random vector $\underline{\theta}$.

Clearly, $\underset{\underline{\theta},\underline{x}}{E}\{\|\underline{\theta}-\hat{\underline{\theta}}(\underline{x})\|^2\}$ is minimized with the conditional mean vector.

$\hookrightarrow E\{(\theta-\hat{\theta})^2\} + E\{(\beta-\hat{\beta})^2\} + E\{(\gamma-\hat{\gamma})^2\} \rightarrow$ total mse.

② Orthogonality $\quad \hat{\underline{\theta}} = E\{\underline{\theta}|\underline{x}\}$: conditional mean vector estimator

Let $\underline{g}(\underline{x})$: vector valued function of observation vector $\underline{x}$.

For any $\underline{g}(\underline{x})$ function, we have

$$\boxed{E\{(\underline{\theta}-\hat{\underline{\theta}}(\underline{x}))(\underline{g}(\underline{x}))^T\} = \underline{\underline{0}}}$$

$$\left.\begin{matrix} \underline{\theta}, \hat{\underline{\theta}}(\underline{x}) : \begin{bmatrix} \; \\ \; \end{bmatrix}_{n\times 1} \\ \underline{g}(\underline{x}) : \begin{bmatrix} \; \\ \; \end{bmatrix}_{m\times 1} \end{matrix}\right\} n\times m$$

So, estimation error of optimal estimator is uncorrelated with any possible linear / non-linear processing of observations.

Proof
$$E\{\hat{\underline{\theta}}(\underline{x})\, \underline{g}^T(\underline{x})\} \overset{?}{=} E\{\underline{\theta}\, \underline{g}^T(\underline{x})\}$$
$$= E_{\underline{x}}\{E_{\underline{\theta}|\underline{x}}\{\underline{\theta}(\underline{x})\, \underline{g}^T(\underline{x})\}\}$$
$$= E_{\underline{x}}\{\underbrace{E_{\underline{\theta}|\underline{x}}\{\underline{\theta}(\underline{x})\}}_{\hat{\underline{\theta}}(x)}\, \underline{g}^T(\underline{x})\}$$
$$= E_{\underline{x}}\{\hat{\underline{\theta}}(x)\, g^T(x)\} \quad ✓$$

Also, reverse: if $E\{(\underline{\theta}-\hat{\underline{\theta}}(x))\, \underline{g}^T(\underline{x})\} = \underline{\underline{0}} \quad \forall\, \underline{g}(\underline{x})$, then $\hat{\underline{\theta}}(\underline{x})$ should be the conditional mean vector estimator.

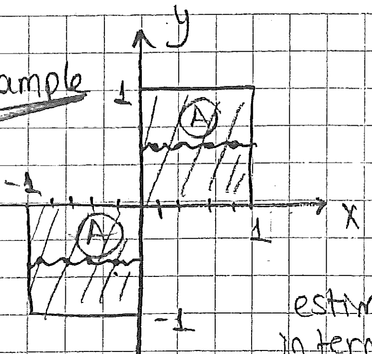
Proof
$$E_{\underline{\theta},\underline{x}}\{\underline{\theta}\, \underline{g}^T(\underline{x})\} - E_{\underline{x}}\{\hat{\underline{\theta}}(\underline{x})\, \underline{g}^T(\underline{x})\} = \underline{\underline{0}}$$
$$E_{\underline{x}}\{E_{\underline{\theta}|\underline{x}}\{\underline{\theta}\, \underline{g}^T(\underline{x})\} - \hat{\underline{\theta}}(\underline{x})\, \underline{g}^T(\underline{x})\} = \underline{\underline{0}}$$
$$E_{\underline{x}}\{(E_{\underline{\theta}|\underline{x}}\{\underline{\theta}\} - \hat{\underline{\theta}}(\underline{x}))\, \underline{g}^T(\underline{x})\} = \underline{\underline{0}}$$

Set $\underline{g}(\underline{x}) = E_{\underline{\theta}|\underline{x}}\{\underline{\theta}\} - \hat{\underline{\theta}}(\underline{x}) \rightarrow E\{\underline{g}(\underline{x})\, \underline{g}^T(\underline{x})\} = \underline{\underline{0}}$

$$\underline{g}(\underline{x}) = \begin{bmatrix} g_1(\underline{x}) \\ g_2(\underline{x}) \\ g_3(\underline{x}) \end{bmatrix} \qquad E\left\{\begin{bmatrix} g_1^2(\underline{x}) & g_1(\underline{x})\, g_2(\underline{x}) & \\ & g_2^2(\underline{x}) & \\ & & g_3^2(\underline{x}) \end{bmatrix}\right\} = \underline{\underline{0}}$$

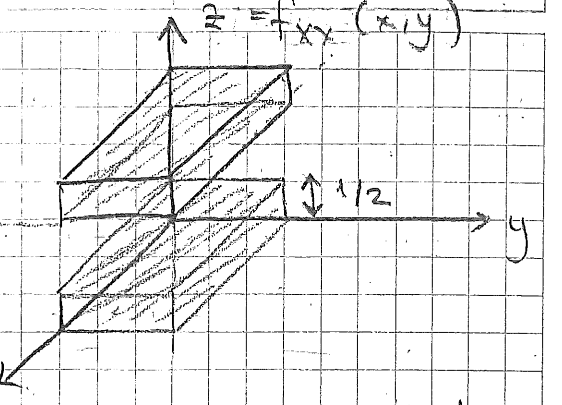
$$\hat{\underline{\theta}}(\underline{x}) = E_{\underline{\theta}|\underline{x}}\{\underline{\theta}\} \quad ✓$$

Example



$f_{XY}(x,y) = \frac{1}{2}$ over shaded area

find the optimal estimator for $\underset{\sim}{y}$ given $\underset{\sim}{x}$ in terms of MSE.



$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \begin{cases} \text{unif}[0,1] & \text{if } 0 \leq x \leq 1 \\ \text{unif}[-1,0] & \text{if } -1 \leq x < 0 \end{cases}$$

$$\hat{y}(x) = E_{Y|X}\{y|x\} = \frac{1}{2}\,\text{sgn}(x)$$

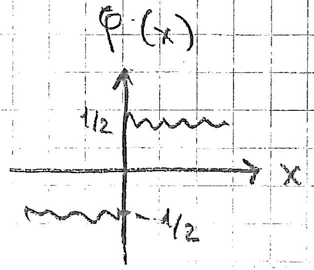
Let's calculate the MSE achieved by $\hat{y}(x) = \frac{1}{2}\,\text{sgn}(x)$ estimator.

$$R = E_{x,y}\{(y-\hat{y}(x))^2\} = E_x\{E_{y|x}\{(y-\hat{y}(x))^2 | x\}\}$$

$$= E_{x|x\geq 0}\{\overbrace{E_{y|x,x\geq 0}\{(y-\underbrace{\hat{y}(x)}_{1/2})^2|x\}}^{\text{var (unif }[0,1])}\}\, P\underbrace{\{x \geq 0\}}_{1/2} +$$

$$E_{x|x<0}\{\underbrace{E_{y|x,x<0}\{(y-\underbrace{\hat{y}(x)}_{-1/2})^2|x\}}_{\text{var (unif }[0,1])}\}\, \underbrace{P\{x<0\}}_{1/2}$$

$$= \frac{1}{12}\cdot\frac{1}{2} + \frac{1}{12}\cdot\frac{1}{2} = \frac{1}{12}$$



Minimum MSE estimator $= \hat{y}(x) = \varphi(x) = \frac{1}{2}\,\text{sgn}(x)$

Minimum MSE value $= E\{(y-\hat{y}(x))^2\}\Big|_{\hat{y}(x)=\frac{\text{sgn}(x)}{2}} = 1/12$

Posterior density: $f_{Y|X}(y|X=x)$ $\longrightarrow$ mean value: min. MSE estimate for y.
$\longrightarrow$ median value: min. MAE estimate for y.
$(E\{|y-\hat{y}(x)|\})$ (mean absolute error)

Another proof for min MSE value for the previous example: (orthogonality)

$\boxed{21.12.2020}$

$$\text{min MSE} = E_{Y,X}\{\underbrace{(y-\varphi(x))}_{\epsilon}{}^2\} = E\{(y-\varphi(x))y\} - \overbrace{E\{\underbrace{(y-\varphi(x))}_{\text{error of optimal MSE est.}}\underbrace{\varphi(x)}_{g(x)\text{, arbitrary function of observations.}}\}}^{0}$$

$$= E_y\{y^2\} - E\{\tfrac{1}{2}\,\text{sgn}(x)\,y\}$$

$$= \sigma_y^2 - \frac{1}{2}\iint_{x\in Ⓐ} \text{sgn}(x)\,y\,f_{XY}(x,y)\,dx\,dy$$

$$= \sigma_y^2 - \frac{1}{4}\left(\int_0^1\int_0^1 \underbrace{\text{sgn}(x)}_{1}\,y\,dx\,dy + \int_{-1}^0\int_{-1}^0 \underbrace{\text{sgn}(x)}_{-1}\,y\,dx\,dy\right)$$

$$= \frac{4}{12} - \frac{1}{4} = \frac{1}{12}//$$

Remember

$0 = E\{(y-E\{y(x)\}g(x)\}$

(orthogonality) $\forall g(x)$

if $\varphi(x)$ is optimal estimator.

On Orthogonality

Q: Let's check whether error r.v. of optimal estimator is uncorrelated with $g(x)=x^k$ or not. (k: positive integer)

A: $E\{(y-\frac{1}{2}\operatorname{sgn}(x))x^k\} \overset{?}{=} 0$

$$= \underbrace{E\{yx^k\}}_{A} - \frac{1}{2}\underbrace{E\{\operatorname{sgn}(x)\,x^k\}}_{B}$$

$$B = \begin{cases} E\{|x|^k\}, & x>0 \\ E\{|x|^k\}, & x<0,\ k: \text{odd} \\ E\{-|x|^k\}, & x<0,\ k: \text{even} \end{cases}$$

$$= \begin{cases} \frac{1}{2}\frac{1}{k+1} + \frac{1}{2}\left(\frac{+1}{k+1}\right), & k \text{ even} \\ \frac{1}{2}\frac{1}{k+1} + \frac{1}{2}\left(\frac{-1}{k+1}\right) = 0, & k \text{ odd} \end{cases}$$

$$A = \iint_{(x,y)\in Ⓐ} yx^k \frac{1}{2}\,dx\,dy$$

$$= \int_0^1\int_0^1 \frac{1}{2} yx^k\,dx\,dy + \int_{-1}^0\int_{-1}^0 \frac{1}{2} yx^k\,dx\,dy$$

$$= \frac{1}{2}\frac{1}{k+1}x^{k+1}\Big|_0^1 + \frac{1}{2}\frac{1}{k+1}x^{k+1}\Big|_{-1}^0 = \begin{cases} \frac{1}{2}\frac{1}{k+1} + \frac{1}{2}\left(\frac{-1}{k+1}\right), & k \text{ even} \\ \frac{1}{2}\frac{1}{k+1} + \frac{1}{2}\left(\frac{-1}{k+1}\right) = 0, & k \text{ odd} \end{cases}$$

$$E\{(y-\frac{1}{2}\operatorname{sgn}(x))\underbrace{\sin(x)}_{x-\frac{x^3}{3!}+\frac{x^5}{5!}-\cdots}\} = 0$$

Example

$\underset{\sim}{x} = \underset{\sim}{y} + \underset{\sim}{n}$ $\quad \underset{\sim}{x}, \underset{\sim}{n}$: jointly Gaussian r.v.

observation: x $\quad$ variable of interest: y

$\begin{bmatrix} \underset{\sim}{x} \\ \underset{\sim}{y} \end{bmatrix}$ jointly Gaussian distributed r.v.'s (Assume both $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are zero mean)

$$f_{xy}(x,y) = \frac{1}{2\pi|\underline{C}|^{1/2}} e^{-\frac{1}{2}[x\ y]\,\underline{C}^{-1}\begin{bmatrix} x \\ y \end{bmatrix}} \qquad \underline{C} = \begin{bmatrix} \sigma_x^2 & \rho_{xy}\sigma_x\sigma_y \\ \rho_{xy}\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$

$$= \frac{1}{2\pi|\underline{C}|^{1/2}} e^{\frac{-1}{2(1-\rho_{xy}^2)}\left(\frac{x^2}{\sigma_x^2} - \frac{2\rho_{xy}}{\sigma_x\sigma_y}xy + \frac{y^2}{\sigma_y^2}\right)}$$

$$\underline{C}^{-1} = \frac{1}{\sigma_x^2\sigma_y^2 - \rho_{xy}^2\sigma_x^2\sigma_y^2}\begin{bmatrix} \sigma_y^2 & -\rho_{xy}\sigma_x\sigma_y \\ -\rho_{xy}\sigma_x\sigma_y & \sigma_x^2 \end{bmatrix}$$

$$f_{y|x}(y|x) = \frac{f_{xy}(x,y)}{f_x(x)}$$

seems quadratic for y.

$$= \frac{c_1 \exp\left\{-\frac{(y-\rho_{xy}\frac{\sigma_y}{\sigma_x}x)^2}{2(1-\rho_{xy}^2)\sigma_y^2} \cancel{- \frac{x^2}{2\sigma_x^2}}\right\}}{c_2 \exp\left\{\cancel{-\frac{x^2}{2\sigma_x^2}}\right\}} = \frac{1}{\sqrt{2\pi(1-\rho_{xy}^2)\sigma_y^2}} e^{-\frac{(y-\rho_{xy}\frac{\sigma_y}{\sigma_x}x)^2}{2(1-\rho_{xy}^2)\sigma_y^2}}$$

Conditional density of y|x is a Gaussian $\longrightarrow N\left(\rho_{xy}\frac{\sigma_y}{\sigma_x}x,\ (1-\rho_{xy}^2)\sigma_y^2\right)$

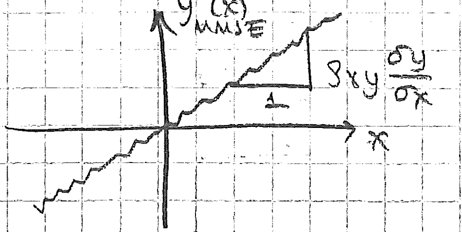
Then, the minimum MSE estimate for y given observation x is the mean value of the density $f_{Y|X}(y|X=x)$, which is $\hat{y}_{MMSE}(x) = \rho_{xy}\frac{\sigma_y}{\sigma_x}x$. The minimum MSE value is

$$E\{(y-\hat{y}_{MMSE}(x))^2\} = E\{(y-\hat{y}_{MMSE}(x))\,y\} - \cancelto{0}{E\{(y-\hat{y}_{MMSE}(x))\,\hat{y}(x)\}}$$

$$= E\{y^2\} - \underbrace{E\{\hat{y}_{MMSE}(x)\,y\}}_{} = \sigma_y^2 - \rho_{xy}^2\sigma_y^2 = \sigma_y^2(1-\rho_{xy}^2)$$

$$E\{\rho_{xy}\frac{\sigma_y}{\sigma_x}xy\} = \rho_{xy}\frac{\sigma_y}{\cancel{\sigma_x}}\rho_{xy}\cancel{\sigma_x}\sigma_y$$

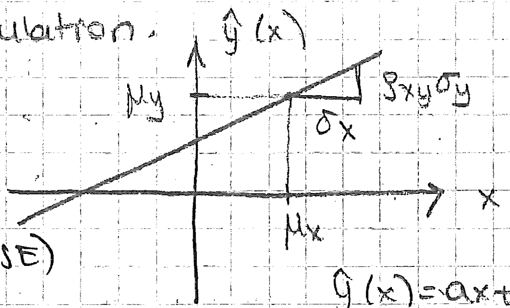
Note that,



$$\hat{y}_{MMSE}(x) = \rho_{xy}\frac{\sigma_y}{\sigma_x}x \rightarrow \text{linear estimator.}$$

So, for jointly Gaussian distributed observations and desired r.v's, the min. MSE estimator is a <u>linear</u> estimator. If $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are jointly Gaussian with non-zero means, then replace $\underset{\sim}{x}$ with $\underset{\sim}{x}-\mu_x$, $\underset{\sim}{y}$ with $\underset{\sim}{y}-\mu_y$ in $f_{y|x}(y|x)$ and do the same calculation.

$$y|x \sim N\left(\rho_{xy}\frac{\sigma_y}{\sigma_x}(x-\mu_x)+\mu_y ; (1-\rho_{xy}^2)\sigma_y^2\right)$$



Linear Minimum Mean Square Error (LMMSE) Estimation

$\hat{y}(x) = ax+b$

Previously, we have derived optimal estimators to minimize MSE. ($E\{(y-\hat{y})^2\}$) without any constraints on estimation function. Now, we introduce <u>parametric</u> estimators and optimize over parameters. We focus on affine/linear estimators.

$\hat{y}(x) = ax+b \rightarrow$ 2 unknown parameters of the estimator.

Then, the MSE for this estimator class becomes

$MSE = E\{(y-\hat{y}(x))^2\}\Big|_{\hat{y}(x)=ax+b}$. Let's find optimal a and b values for minimizing MSE (Linear MMSE estimation problem)

$$J(a,b) = E\{\overbrace{(y-(ax+b))}^{e}{}^2\} = E\{e^2\} \qquad (e = y-ax-b)$$

$\frac{\partial J}{\partial a}$

$$E\{\tfrac{\partial}{\partial a} e^2\} = E\{2e\tfrac{\partial e}{\partial a}\} = -2E\{ex\} = 0 \quad (1)$$

$\frac{J}{b}$

$$E\{\tfrac{\partial}{\partial b} e^2\} = E\{2e\tfrac{\partial e}{\partial b}\} = -2E\{e\} = 0 \quad (2)$$

$$(1) \quad E\{ex\} = E\{yx\} - a\underbrace{E\{x^2\}}_{\sigma_x^2+m_x^2} - b\underbrace{E\{x\}}_{m_x}$$

$$(2) \quad E\{e\} = \underbrace{E\{y\}}_{m_y} - a\underbrace{E\{x\}}_{m_x} - b$$

$$\begin{bmatrix} \sigma_x^2+m_x^2 & m_x \\ m_x & 1 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} E\{ex\} \\ m_y \end{bmatrix}$$

Solving for b from (2)

$\hookrightarrow \boxed{b_{OPT} = m_y - am_x}$

$\hookrightarrow \hat{y} = ax+b \downarrow b = m_y - am_x$

$\hookrightarrow \boxed{\hat{y} = a(x-m_x)+m_y}$

Insert b into (1)

$\hookrightarrow (\sigma_x^2+\cancel{m_x^2})a + m_x(m_y - \cancel{am_x}) = E\{xy\}$

$$a_{OPT} = \frac{E\{xy\}-m_xm_y}{\sigma_x^2} = \frac{Cov(x,y)}{\sigma_x^2}$$

$$\boxed{a_{OPT} = \frac{\rho_{xy}\sigma_x\sigma_y}{\sigma_x^2} = \frac{\rho_{xy}\sigma_y}{\sigma_x}}$$

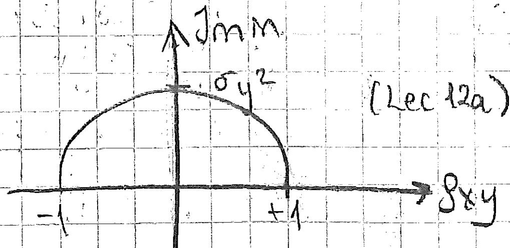
optimal LMMSE $\rightarrow \boxed{\hat{y}(x) = \frac{\rho_{xy}\sigma_y}{\sigma_x}(x-m_x)+m_y}$

Let's calculate the LMMSE value for optimal estimator.

$$J_{mm}(a_{OPT}, b_{OPT}) = E\{(y-(a_{OPT}x+b_{OPT}))^2\} = E\{((y-m_y)-a_{OPT}(x-m_x))^2\}$$

$$= \sigma_y^2 - 2\underbrace{a_{OPT}}_{\rho_{xy}\frac{\sigma_y}{\sigma_x}}\underbrace{Cov(x,y)}_{\rho_{xy}\sigma_x\sigma_y} + \underbrace{a_{OPT}^2}_{\frac{\rho_{xy}^2\sigma_y^2}{\sigma_x^2}}\sigma_x^2 = \sigma_y^2 - 2\sigma_y^2\rho_{xy}^2 + \rho_{xy}^2\sigma_y^2 = \sigma_y^2 - \rho_{xy}^2\sigma_y^2$$

$$\boxed{J_{mm}(a_{OPT}, b_{OPT}) = \sigma_y^2(1-\rho_{xy}^2)}$$



(Lec 12a)

uncorrelated $\rightarrow \rho_{xy}=0 \rightarrow a_{OPT}=0$

$\rightarrow$ observation is discarded $\rightarrow \hat{y} = m_y$

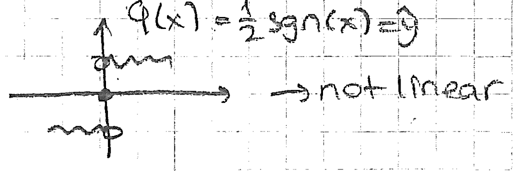
Comments

1) LMMSE estimators are parametric estimators, whose parameters can be optimally set by simply calculus. We do not need joint pdf information for optimization, but just the <u>moment</u> information

is sufficient to get the optimal parameters.

2) If $\underline{x}, \underline{y}$ are jointly Gaussian, then LMMSE estimator is the min MSE estimator.

3) In practice, we only estimate moments but not densities in general; hence, one can say that we assume jointly Gaussian distributed random variables in LMMSE calculations.



A simpler and highly recommended vector processing based calculation of LMMSE estimators:

$$MSE \rightarrow E\{(y-\underbrace{\hat{y}(x)}_{})^2\}$$

$$\underline{w}^T\underline{x} = [w_1\ w_2 \dots w_N]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \sum_{k=1}^{N} w_k x_k$$

How can we select linear combination coefficients (w_k) s.t. MSE is minimized?

$$J(\underline{w}) = E\{(y-\underline{w}^T\underline{x})^2\} = E\{\underline{e}^2\}$$

$$\nabla_{\underline{w}} J(\underline{w}) = 0 \rightarrow E\{\nabla_{\underline{w}} e^2\} = E\{2e\nabla_{\underline{w}} e\} = E\{2e\begin{bmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_N \end{bmatrix}\} = \underline{0}$$

$$\begin{bmatrix} \frac{\partial J}{\partial w_1} \\ \vdots \\ \frac{\partial J}{\partial w_N} \end{bmatrix}$$

$$E\{\underline{e}\,\underline{x}\} = 0 \rightarrow \text{orthogonality condition for LMMSE}$$

$$\downarrow$$

$$E\{(\underline{y}-\underline{w}^T\underline{x})\underline{x}\} \rightarrow E\{\underline{x}\,\overbrace{(\underline{w}^T\underline{x})}^{\text{scalar} = \underline{x}^T\underline{w}}\} = E\{y\,\underline{x}\}$$

$$(*)\ \boxed{E\{\underline{x}\,\underline{x}^T\}\,\underline{w} = E\{y\,\underline{x}\}}$$

$$\boxed{\underline{\underline{R}}_x\underline{w} = r_{y\underline{x}}}$$

$\underline{\underline{R}}_x$: autocorrelation matrix of observation vector entries

$$E\left\{\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}[x_1 \dots x_N]\right\} = \begin{bmatrix} E\{x_1,x_1\} & & E\{x_1,x_N\} \\ & & \\ E\{x_N,x_1\} & & E\{x_N,x_N\} \end{bmatrix}$$

$r_{y\underline{x}}$: cross-correlation vector of desired r.v. and observations

$$r_{y\underline{x}} = E\left\{y\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}\right\} = E\left\{\begin{bmatrix} yx_1 \\ \vdots \\ yx_N \end{bmatrix}\right\}$$

The min. LMMSE value:

$$J_{min}(\underline{w}) = E\{(\underbrace{y-\underline{w}^T\underline{x}}_{e})^2\}\Big\downarrow_{\underline{w}=\underline{w}_{OPT}} = E\{ey\} - E\{e\,\underline{w}_{OPT}^T\,\underline{x}\}$$

$$\hookrightarrow \underline{w}_{OPT}^T\,\underline{E\{e\underline{x}\} = 0}$$

orthogonality principle for LMMSE

$$= E\{(y-\underline{w}_{OPT}^T\underline{x})y\} = E\{y^2\} - \underline{w}_{OPT}^T E\{y\underline{x}\}$$

$$= \boxed{E\{y^2\} - \underline{w}_{OPT}^T\, r_{y\underline{x}}} = E\{y^2\} - r_{y\underline{x}}^T\,\underline{\underline{R}}_x^{-1}\,r_{y\underline{x}}$$

$$\hookrightarrow \underline{w}_{OPT} = \underline{\underline{R}}_x^{-1} r_{y\underline{x}}$$

e: error of optimal LMMSE estimator.

$\boxed{L33}$

Let's revisit the earlier problem with the vector notation:

$\hat{y}(\underline{x}) = ax+b = \underbrace{[a \ \ b]}_{\underline{\omega}^T} \underbrace{\begin{bmatrix} x \\ 1 \end{bmatrix}}_{\underline{x}}$ → the optimal weights (a, b) satisfy $\underline{\underline{R}}_x \, \underline{\omega} = r_{y\underline{x}}$

$$\underline{\underline{R}}_x = \left\{ \begin{bmatrix} x \\ 1 \end{bmatrix} [x \ \ 1] \right\} = E\left\{ \begin{bmatrix} x^2 & x \\ x & 1 \end{bmatrix} \right\} = \left[\begin{array}{c|c} \sigma_x^2 + m_x^2 & m_x \\ \hline m_x & 1 \end{array} \right]$$

$$r_{y\underline{x}} = E\left\{ y \begin{bmatrix} x \\ 1 \end{bmatrix} \right\} = \begin{bmatrix} E\{xy\} \\ m_y \end{bmatrix} \qquad \left[\begin{array}{c|c} \sigma_x^2 + m_x^2 & m_x \\ \hline m_x & 1 \end{array} \right] \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} E\{xy\} \\ m_y \end{bmatrix}$$

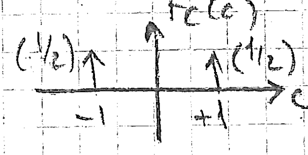
Example

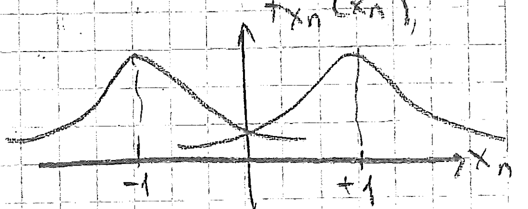
$\underset{\sim}{x_n} = \underset{\sim}{c} + \underset{\sim}{w_n}$ where $\underset{\sim}{c}$ is a zero mean r.v. and

$n = \{1, \ldots, N\}$ $\quad \underset{\sim}{w_n}$ is independent Gaussian r.v. $\mathcal{N}(0, \sigma_{w_n}^2)$

$\underset{\sim}{c}$ & $\underset{\sim}{w_n}$ → independent

Find LMMSE estimation for c. $\quad C = \{-1, +1\}$ with equal probability.





$$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \rightarrow \underline{R}_x = E\{\underline{x}\,\underline{x}^T\}$$

$$= \begin{bmatrix} E\{x_1^2\} & \cdots & E\{x_1 x_N\} \\ \vdots & & \vdots \\ E\{x_N x_1\} & & E\{x_N^2\} \end{bmatrix}$$

$f_{x_n}(x_n) = f_{\underset{\sim}{c}}(c) * f_{\underset{\sim}{w_n}}(w_n)$

$E\{(c+w_1)(c+w_N)\} = \sigma_c^2$

$\hookrightarrow E\{(c+w_N)(c+w_N)\} = \sigma_c^2 + \sigma_{w_N}^2$

$$r_{\underline{x}y} = E\left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} y \right\} = \begin{bmatrix} E\{x_1 y\} \\ \vdots \\ E\{x_N y\} \end{bmatrix} = \begin{bmatrix} E\{x_1 c\} \\ \vdots \\ E\{x_N c\} \end{bmatrix} = \begin{bmatrix} E\{c^2 + \overset{0}{\cancel{w_1 c}}\} \\ \vdots \\ E\{c^2 + \overset{0}{\cancel{w_N c}}\} \end{bmatrix} = \sigma_c^2 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} \sigma_c^2 + \sigma_{w_1}^2 & \sigma_c^2 & \sigma_c^2 \cdots \sigma_c^2 & \cdots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_{w_2}^2 & \sigma_c^2 & & \\ \sigma_c^2 & \cdots & \sigma_c^2 + \sigma_{w_k}^2 & \cdots & \\ \sigma_c^2 & & & & \sigma_c^2 + \sigma_{w_N}^2 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ \vdots \\ w_N \end{bmatrix} = \sigma_c^2 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

multiply both sides with $1/\sigma_c^2$

$$(SNR)_k = \frac{\text{signal power at } k^{th} \text{ observation}}{\text{noise power at } k^{th} \text{ observation}} = \frac{E\{c^2\}}{E\{w_k^2\}} = \frac{\sigma_c^2}{\sigma_{w_k}^2} \rightarrow \begin{bmatrix} 1+\frac{1}{SNR_1} & 1 & \cdots & 1 \\ 1 & 1+\frac{1}{SNR_2} & & \vdots \\ 1 & \cdots & 1+\frac{1}{SNR_k} & \vdots \\ 1 & \cdots & & 1+\frac{1}{SNR_N} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Let's check k^{th} equation $\rightarrow \sum_{k=1}^{N}(w_k) + \frac{w_k}{(SNR)_k} = 1$ Let $s = \sum_{k=1}^{N}(w_k)$

$$w_k = (1-s)SNR_k \rightarrow s = \sum_{k=1}^{N} w_k = \sum_{k=1}^{N}(1-s)SNR_k = (1-s)\sum_{k=1}^{N} SNR_k$$

$$s = \frac{\sum_k SNR_k}{1+\sum_k SNR_k} \longrightarrow w_k = (1-s)SNR_k \rightarrow \boxed{w_k = \frac{SNR_k}{1+\sum_k SNR_k}}$$ not completely linear.

$$\underline{w}^T = \left[\frac{SNR_1}{1+\sum_k SNR_k}, \frac{SNR_2}{1+\sum_k SNR_k}, \ldots, \frac{SNR_N}{1+\sum_k SNR_k}\right]$$

LMMSE value $= E\{(y-\hat{y})^2\} = E\{y^2\} - \underline{w}^T r_{\underline{x}y}$ where $\underline{w} = \underline{w}_{opt}$

$$\hookrightarrow \sigma_c^2 - \underline{w}^T(\sigma_c^2 \underline{1}) = \sigma_c^2(1-\underline{w}^T\underline{1}) = \sigma_c^2(1-s) = \boxed{\frac{\sigma_c^2}{1+\sum_k SNR_k}}$$

Example

$x_n = c + w_n$
$n = \{1, \ldots, N\}$

$c : N(\mu_c, \sigma_c^2)$
$w_n : N(0, \sigma_{w_n}^2)$ } uncorrelated r.v.'s.

Find LMMSE estimator for c.

$$E\{x_1^2\} = E\{c^2\} + E\{w_1^2\} = \sigma_c^2 + \mu_c^2 + \sigma_{w_1}^2 \rightarrow \underline{\underline{R}}_x = \begin{bmatrix} \sigma_c^2+\mu_c^2+\sigma_{w_1}^2 & \sigma_c^2+\mu_c^2 & \\ \sigma_c^2+\mu_c^2 & \sigma_c^2+\mu_c^2+\sigma_{w_2}^2 & \\ \vdots & \vdots & \vdots \end{bmatrix}$$

$$(\hat{SNR})_k = \frac{E\{c^2\}}{E\{w_k^2\}} = \frac{\sigma_c^2+\mu_c^2}{\sigma_{w_k}^2}$$

$\downarrow$

only SNR definition changes $\longrightarrow$

$$\begin{bmatrix} 1+\frac{1}{SNR_1} & 1 & \cdots & 1 \\ 1 & 1+\frac{1}{SNR_2} & & \vdots \\ \vdots & & & \\ & & & 1+\frac{1}{SNR_N} \end{bmatrix}\begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\hat{y} = \underline{w}^T\underline{x} = \frac{\sum_k (\hat{SNR})_k \cdot x_k}{1+\sum_k (\hat{SNR})_k}$$

Q: Does $\hat{y}$ for this problem result in an unbiased estimator?

A: $E\{\hat{y}\} = \underline{w}^T E\{\underline{x}\} = \underline{w}^T\begin{bmatrix}\mu_c \\ \vdots \\ \mu_c\end{bmatrix} = \mu_c(\underline{w}^T \cdot \underline{1}) \overset{?}{=} E\{c\} = \mu_c$

If $\underline{w}^T \cdot \underline{1} = 1$, then estimator is unbiased, but in this case

$\sum_k w_k = \frac{\sum_k SNR_k}{1+\sum_k SNR_k} \neq 1$. So, estimator is biased. If $\mu_c = 0$, then

LMMSE is unbiased. Let's try to remove the bias by including another (artificial) observation which is "1".

L34

$$\underline{x}_{new} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \\ 1 \end{bmatrix}_{(N+1)\times 1} \longrightarrow \hat{y} = \underline{w}_{new}^T \cdot \underline{x}_{new} = \sum_{k=1}^{N} w_k x_k + w_{N+1}^{new} \longrightarrow \underline{\text{affine}} \text{ estimator.}$$

$$\longrightarrow \hat{\underline{x}}_{new} = \begin{bmatrix} x_1 - \mu_c \\ x_2 - \mu_c \\ \vdots \\ x_N - \mu_c \\ 1 \end{bmatrix} \longrightarrow \text{subtracting a constant value from every observation.}$$

$$(x_k = \underbrace{c}_{\to N(\mu_c, \sigma_c^2)} + \underbrace{w_k}_{\to N(0, \sigma_{w_k}^2)})$$

$$\hat{x}_{new_k} = x_k - \mu_c = \underbrace{c - \mu_c}_{\hat{c} \sim N(0, \sigma_c^2)} + w_k$$

$$E\{\hat{\underline{x}}_{new} \hat{\underline{x}}_{new}^T\} \to \underbrace{\begin{bmatrix} \boxed{\text{same}, \underline{R}_x \quad \textcircled{N \times N}} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ 0 \cdots 0 & 1 \end{bmatrix}}_{(N+1)\times(N+1)} \underbrace{\begin{bmatrix} w_1^{new} \\ w_2^{new} \\ \vdots \\ w_N^{new} \\ w_{N+1}^{new} \end{bmatrix}}_{(N+1)\times 1} = \begin{bmatrix} \boxed{\begin{matrix}\text{same} \\ \sigma_c^2 \cdot \underline{1}\end{matrix}} \\ \mu_c \end{bmatrix}$$

$$(E\{\hat{x}_{new_k}\} = 0)$$

Then, we see that $w_1^{new}, w_2^{new}, \ldots, w_N^{new}$ are exactly as before. That is, $w_k^{new} = \dfrac{SNR_k}{1 + \sum_k SNR_k}$ and $w_{N+1}^{new} = \mu_c$.

$$\hat{y} = \underline{w}^{new^T} \cdot \underline{x}_{new} = \sum_{k=1}^{N} \frac{SNR_k}{1 + \sum_k SNR_k} (x_k - \mu_c) + 1 \cdot \mu_c \longrightarrow \text{unbiased affine MMSE estimator.}$$

Complex Valued Case for LMMSE Estimator: $\boxed{28.12.2020}$

$$J(\underline{w}) = E\{|y - \hat{y}|^2\} \qquad \hat{y} : \underline{w}^H \cdot \underline{x}$$

$$\nabla_{\underline{w}^*} J(\underline{w}) = \begin{bmatrix} \partial/\partial w_1^* \\ \\ \partial/\partial w_N^* \end{bmatrix} \underbrace{J(\underline{w})}_{E\{e \cdot e^*\}} = E\{\nabla_{\underline{w}^*}(e \cdot e^*)\} = E\{e^* \nabla_{\underline{w}^*}(e)\} = E\{e^*(-\underline{x})\} = \underline{0}$$

$$e^* \to y^* - \underline{w}^T \underline{x}^* \qquad e \to y - \underline{w}^H \underline{x} \quad (\text{here})$$

$$\boxed{E\{e^* \underline{x}\} = \underline{0} \longrightarrow \text{orthogonality principle}}$$

$$E\{\underline{x}(y^* - \underline{w}^T \underline{x}^*)\} = \underline{0} \longrightarrow E\{\underline{x}(\underbrace{\underline{w}^T \underline{x}^*}_{\text{scalar} = \underline{x}^H \underline{w}})\} = E\{\underline{x} y^*\} \qquad \to E\{\underline{x}\,\underline{x}^H \underline{w}\}$$

$$\boxed{\underbrace{\underline{\underline{R}}_x}_{E\{\underline{x}\,\underline{x}^H\}} \underbrace{\underline{w}}_{\underline{w}_{OPT}} = \underbrace{r_{\underline{x}y}}_{E\{\underline{x} y^*\}}}$$

Let's also calculate LMMSE relation:

$$J_{min}(\underline{w}_{OPT}) = E\{|e|^2\} = E\{e^*(y - \underline{w}^H \underline{x})\}$$

$$= E\{e^* y\} - \underline{w}^H \underbrace{\cancel{E\{e^* \underline{x}\}}}_{0} = E\{(y^* - \underline{w}^T \underline{x}^*) y\} = E\{|y|^2\} - \underline{w}^T E\{\underline{x}^* y\}$$

$$= E\{|y|^2\} - \underbrace{\underline{w}^T r_{\underline{x}y}^*}_{\text{scalar}} = E\{|y|^2\} - r_{\underline{x}y}^H \underline{w} = \boxed{E\{|y|^2\} - r_{\underline{x}y}^H \underline{\underline{R}}_x^{-1} r_{\underline{x}y}}$$

example: $\underset{\sim}{x} = \underset{\downarrow \text{ non-random.}}{\underline{p}}\, \underset{\sim}{c} + \underset{\sim}{n}$

$\underline{p}$: known vector
$\underset{\sim}{c}$: desired r.v. $E\{c\}=0$
$\underset{\sim}{n}$: noise , $CN(\underline{0}, \underline{\underline{C}}_n)$
↳ uncorrelated with c
} complex valued.

Estimate $\underset{\sim}{c}$ with LMMSE estimator.

Special Case $\underline{\underline{C}}_n = \underline{\underline{I}}$

$$\underline{\underline{R}}_x = E\{(\underline{p}c+\underline{n})(\underline{p}c+\underline{n})^H\} = \underline{p}\underline{p}^H\sigma_c^2 + \underline{\underline{C}}_n$$

$$r_{\underline{x}y} = E\{\underline{x}c^*\} = E\{(\underline{p}c+\underline{n})c^*\} = \underline{p}\sigma_c^2$$

$$\underline{\underline{R}}_x \underline{w} = r_{\underline{x}y}$$

$$(\underline{p}\underline{p}^H\sigma_c^2 + \underline{\underline{I}})\underline{w} = \underline{p}\sigma_c^2 \quad \text{(*)}$$

Matrix Inversion Lemma:

$$\underline{\underline{R}}_x = \sigma_c^2 \underline{p}\underline{p}^H + \underline{\underline{I}}$$

$$\underline{\underline{R}}_x \underline{p} = \sigma_c^2 \underline{p}\underbrace{\underline{p}^H\underline{p}}_{\|\underline{p}\|^2} + \underbrace{\underline{\underline{I}}\underline{p}}_{\underline{p}}$$

$$= (\sigma_c^2\|\underline{p}\|^2 + 1)\underline{p}$$

$$\underline{q}_k \perp \underline{p} \xrightarrow{} \underline{\underline{R}}_x\underline{q}_2 = \sigma_c^2 \cancel{\underline{p}\underbrace{\underline{p}^H\underline{q}_2}_{0}} + \underline{\underline{I}}\underline{q}_2$$

$k = \{2, \dots, N\}$

$\underline{e}_1 = \underline{p}$, $\lambda_1 = \|\underline{p}\|^2\sigma_c^2 + 1$
$\underline{e}_2 = \underline{q}_2$, $\lambda_2 = 1$
$\underline{e}_3 = \underline{q}_3$, $\lambda_3 = 1$
$\underline{e}_N = \underline{q}_N$, $\lambda_N = 1$

$$\underline{\underline{R}}_x \underline{p} = \lambda_1 \cdot \underline{p}$$
$$\downarrow \underline{\underline{R}}_x^{-1}$$
$$\underline{p} = \lambda_1 \underline{\underline{R}}_x^{-1}\underline{p}$$
$$\underline{\underline{R}}_x^{-1}\underline{p} = \frac{1}{\lambda_1}\underline{p}$$

$$\text{(*)} \rightarrow \underline{w} = \underline{\underline{R}}_x^{-1}\underline{p}\sigma_c^2 = \frac{1}{\lambda_1}\underline{p}\cdot\sigma_c^2 = \frac{\sigma_c^2}{\sigma_c^2\|\underline{p}\|^2+1}\cdot\underline{p} = \underline{w}$$

Then, $\hat{c} = \underline{w}^H\underline{x} = \dfrac{\sigma_c^2}{\sigma_c^2\|\underline{p}\|^2+1}\cdot\underline{p}^H\underline{x} \xrightarrow[\text{analyze further}]{\text{insert } \underline{x} \text{ to}} \hat{c} = \dfrac{\sigma_c^2}{\sigma_c^2\|\underline{p}\|^2+1}(\|\underline{p}\|^2c + \underline{p}^H\underline{n})$

if noise is small → $\hat{c} = \text{cons.} \cdot c$ → scaled version of c.
↳ if = 1 → it's exactly correct.

if zero mean → unbiased.

General Case

$$\underline{\underline{C}}_n \neq \underline{\underline{I}} \longrightarrow \underset{\downarrow \text{ processed observation}}{\hat{\hat{\underline{x}}}} = \overset{\longrightarrow \text{ whitening operator}}{\underline{\underline{C}}_n^{-1/2}}\underline{x} = \underbrace{\underline{\underline{C}}_n^{-1/2}\underline{p}}_{\hat{\hat{\underline{p}}}}c + \underbrace{\underline{\underline{C}}_n^{-1/2}\underline{n}}_{\hat{\hat{\underline{n}}}}$$

$$\hat{\hat{\underline{x}}} = \hat{\hat{\underline{p}}}c + \hat{\hat{\underline{n}}} \quad \hat{\hat{\underline{n}}} \hookrightarrow CN(\underline{0}, \underline{\underline{I}})$$

So, after whitening operation, we can use the previous results for uncorrelated noise case.

$$\hat{\hat{\underline{w}}} = \frac{\sigma_c^2}{\sigma_c^2\|\hat{\hat{\underline{p}}}\|^2+1}\hat{\hat{\underline{p}}} = \frac{\sigma_c^2\underline{\underline{C}}_n^{-1/2}\underline{p}}{\sigma_c^2\underline{p}^H\underline{\underline{C}}_n^{-1}\underline{p}+1}$$

$$\downarrow$$

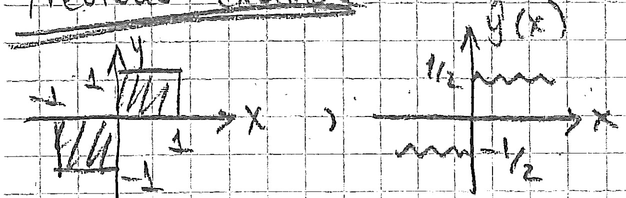
$$\hat{\hat{\underline{p}}}^H\hat{\hat{\underline{p}}} = \underline{p}^H\underline{\underline{C}}_n^{-1/2}\underline{\underline{C}}_n^{-1/2}\underline{p}$$

$$\hat{c} = \hat{\hat{\underline{w}}}^H\hat{\hat{\underline{x}}} = \frac{\sigma_c^2\underline{p}^H\underline{\underline{C}}_n^{-1/2}}{1+\sigma_c^2\underline{p}^H\underline{\underline{C}}_n^{-1}\underline{p}}\cdot\hat{\hat{\underline{x}}} \quad \hat{\hat{\underline{x}}} \nearrow \underline{\underline{C}}_n^{-\frac{1}{2}}\underline{x}$$

$$\hat{c} = \frac{\sigma_c^2}{1+\sigma_c^2\underline{p}^H\underline{\underline{C}}_n^{-1}\underline{p}}\cdot\underline{p}^H\underline{\underline{C}}_n^{-1}\underline{x} \quad \text{(**)}$$

mini assignment: set $\underline{p}=\underline{1}$ and $\underline{\underline{C}}_n = \mathrm{diag}(\sigma_{w_1}^2, \dots, \sigma_{w_N}^2)$ and compare the result (★★) with $\{\underset{\sim}{x}_n = \underset{\sim}{c} + \underset{\sim}{w}_n , n = 1, \dots, N\}$ and also compare the MSE values.

Previous example



min MSE estimator → $\hat{y}_{mmse} = \frac{1}{2}\mathrm{sgn}(x)$
(non-linear)

$E\{y \mid \underline{x} = x\}$ ↑
(conditional mean)

min MSE → $E\{(y-\hat{y}_{mmse})^2\} = \frac{1}{12} = 0.08\overline{3}$

Let's derive affine MMSE estimator for the same problem and compare the results.

$$\underline{x} = \begin{bmatrix} 1 \\ x \end{bmatrix} \rightarrow \underset{\substack{A \\ \text{affine}}}{\hat{y}}(x) = \underline{w}_A^T \underline{x} = \underbrace{w_A(1)\, 1}_{\text{constant term} \rightarrow \text{ not linear, but affine}} + w_A(2)\, x$$

$$E\{(y-\hat{y}_A(x))^2\} \longrightarrow \underline{\underline{R}}_x \underline{w}_A = \underline{r}_{xy} \rightarrow E\{\underline{x}\,\underline{x}^T\}\underline{w}_A = E\{\underline{x} y\}$$

$$\begin{bmatrix} E\{1\} & E\{x\} \\ E\{x\} & E\{x^2\} \end{bmatrix} \begin{bmatrix} w_A(1) \\ w_A(2) \end{bmatrix} = \begin{bmatrix} E\{y\} \\ E\{xy\} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1/3 \end{bmatrix} \begin{bmatrix} w_A(1) \\ w_A(2) \end{bmatrix} = \begin{bmatrix} 0 \\ 1/4 \end{bmatrix} \rightarrow \underline{w}_A = \begin{bmatrix} 0 \\ 3/4 \end{bmatrix}$$

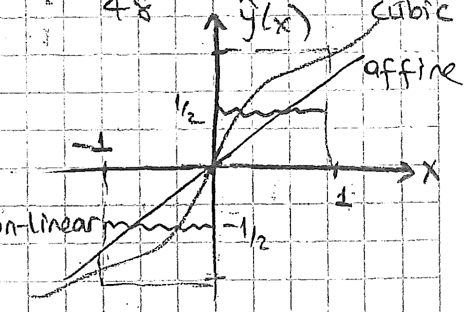
$$E\{yx^k\} = \iint_A y x^k f_{xy}(x,y)\, dx\, dy = \begin{cases} 1/2(k+1) & , k: \text{odd} \\ 0 & , k: \text{even} \end{cases}$$

$$E\{x^k\} = \begin{cases} 0 & , k: \text{odd} \\ 1/k+1 & , k: \text{even} \end{cases}$$

$$\boxed{\hat{y}_A = \frac{3}{4} x}$$

MSE of affine estimator → $E\{(y-\hat{y}_A(x))^2\}$

$$= E\{y^2\} - \underline{w}_A^T \underline{r}_{xy} = \frac{1}{3} - \frac{3}{4} \cdot \frac{1}{4} = \frac{7}{48} \approx 0.1458$$



Since we are constructing parametric estimators with optimized parameters ($\underline{w}$ vector); we can also take the following as the observation vector and apply the formalism for optimal parameter finding.

estimator	min MSE
non-linear	$0.08\overline{3}$
affine	≈ 0.1458
cubic	≈ 0.1185
5th order	≈ 0.1075

$$\underline{x} = \begin{bmatrix} 1 \\ x \\ x^2 \\ x^3 \end{bmatrix} \begin{matrix} - x_1 \\ - x_2 \\ - x_3 \\ - x_4 \end{matrix} \qquad \underset{\substack{\text{cubic} \\ \text{estimator}}}{\hat{y}_c(x)} = w_0^c\, 1 + w_1^c x + w_2^c x^2 + w_3^c x^3$$

$$\underline{\underline{R}}_x \underline{w}^c = r_{\underline{x}y} \quad \begin{array}{c|cccc} & 1 & x & x^2 & x^3 \\ \hline 1 & 1 & 0 & 1/3 & 0 \\ x & 0 & 1/3 & 0 & 1/5 \\ x^2 & 1/3 & 0 & 1/5 & 0 \\ x^3 & 0 & 1/5 & 0 & 1/7 \end{array} \begin{bmatrix} w_0{}^c \\ w_1{}^c \\ w_2{}^c \\ w_3{}^c \end{bmatrix} = \begin{array}{c|c} & y \\ \hline 1 & 0 \\ x & 1/4 \\ x^2 & 0 \\ x^3 & 1/8 \end{array}$$

$$w_0{}^c = w_2{}^c = 0$$

$$\begin{bmatrix} 1/3 & 1/5 \\ 1/5 & 1/7 \end{bmatrix} \begin{bmatrix} w_1{}^c \\ w_3{}^c \end{bmatrix} = \begin{bmatrix} 1/4 \\ 1/8 \end{bmatrix} \rightarrow \begin{array}{l} w_3{}^c = -35/32 \\ w_1{}^c = 45/32 \end{array}$$

MSE of cubic estimator:

$$E\{y^2\} - (\underline{w}^c)^T r_{\underline{x}y} = \frac{91}{256 \times 3} \approx 0.1185$$

Properties of LMMSE Estimators

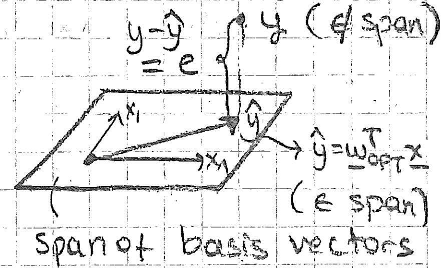
① Geometric (Vector space) interpretation:

$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}$: observation vector y : desired r.v. $\hat{y} = \underline{w}^T \cdot \underline{x}$: LMMSE estimator provided that $E\{\underbrace{(y-\hat{y})}_{e}\underline{x}\} = 0$

$\downarrow$ orthogonality.

$$E\{e\underline{x}\} = \underline{0}_{N\times 1} \rightarrow \underline{\underline{R}}_x \underline{w} = r_{\underline{x}y}$$

$\llcorner$ estimation error of LMMSE estimator.



say x_i: abstract basis vectors.

$\underline{w}$: adds x_i's

$\hat{y}$: linear combination of basis vectors.

$$\underset{\text{inner product}}{\langle x, y\rangle} = E\{xy\} \rightarrow \underset{\text{norm induced by inner product}}{\|\underline{x}\|^2} = \langle \underline{x}, \underline{x} \rangle = E\{\underline{x}^2\}$$

$\underline{\underline{R}}_x$ = matrix of inner products $\rightarrow$ $\begin{bmatrix} E\{x_1^2\} & E\{x_1x_2\} & - & - & E\{x_1x_n\} \\ \langle x_2, x_1\rangle & - & - & & \langle x_2, x_n\rangle \\ & & & & E\{x_n^2\} \end{bmatrix}$

② Multiple random variable estimation from $\underline{X}$ vector:

$$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \begin{array}{l} \rightarrow E\{|y_1 - \underline{w}_1{}^H \underline{x}|^2\} \rightarrow \underline{\underline{R}}_x \underline{w}_1 = r_{\underline{x}y_1} \rightarrow E\{\underline{x}y_1{}^*\} \\ \qquad\qquad\qquad\qquad\qquad\qquad \nearrow E\{\underline{x}\underline{x}^H\} \\ \rightarrow E\{|y_2 - \underline{w}_2{}^H \underline{x}|^2\} \rightarrow \underline{\underline{R}}_x \underline{w}_2 = r_{\underline{x}y_2} \rightarrow E\{\underline{x}y_2{}^*\} \end{array}$$

$$\underline{\underline{R}}_x \underbrace{[\underline{w}_1 \mid \underline{w}_2]}_{\underline{\underline{W}}} = [r_{\underline{x}y_1} \mid r_{\underline{x}y_2}] \qquad \hat{\underline{y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix} = \begin{bmatrix} \underline{w}_1{}^H \\ \underline{w}_2{}^H \end{bmatrix} \underline{x} = \underline{\underline{W}}^H \underline{x}$$

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \qquad E\{\|\underline{y} - \underbrace{\hat{\underline{y}}}_{\underline{\underline{W}}^H \underline{x}}\|^2\} = E\left\{\left\|\begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \end{bmatrix}\right\|^2\right\} = E\{\|y_1 - \overset{\underline{w}_1{}^H \underline{x}}{\hat{y}_1}\|^2\} + E\{\|y_2 - \overset{\underline{w}_2{}^H \underline{x}}{\hat{y}_2}\|^2\}$$

$\llcorner$ desired random vector

Total MSE → MSE for $\hat{y}_1$ and $\hat{y}_2$ estimators

L36a

So, total MSE is minimized by $\underline{w}_1$ and $\underline{w}_2$ vectors if

$$\underline{W} \triangleq [\underline{w}_1 \mid \underline{w}_2] \quad \text{and} \quad \underline{\underline{R}}_x \underline{\underline{W}} = [r_{\underline{x}y_1} \mid r_{\underline{x}y_2}] = E\{\underline{x}\,\underbrace{\underline{y}^H}_{[y_1^* \; y_2^*]}\}$$

$$\underline{\underline{R}}_x \underline{\underline{W}} = r_{\underline{x}\underline{y}} \rightarrow \boxed{\hat{\underline{y}} = \underline{\underline{W}}^H \underline{x}} \rightarrow \text{optimal estimator minimizing total MSE.}$$

$$\hat{\underline{y}} = \underline{\underline{W}}^H \underline{x} = (\underline{\underline{R}}_x^{-1} r_{\underline{x}\underline{y}})^H \underline{x} = r_{\underline{x}\underline{y}}^H \underline{\underline{R}}_x^{-1} \underline{x} \longrightarrow \boxed{\hat{\underline{y}} = r_{\underline{y}\underline{x}} \underline{\underline{R}}_x^{-1} \underline{x}}$$

decorrelating $\underline{x}$ and estimating using x-y relation.

Example

$$\underset{\text{observation vector}}{\underline{x}} = \underset{\text{channel}}{\underline{\underline{H}}}\,\overset{\text{desired vector}}{\underline{y}} + \underset{\text{noise vector}}{\underline{n}}$$

noise and desired vector, are uncorrelated. Find LMMSE for $\hat{\underline{y}}$. (all zero mean).

$$\underline{\underline{R}}_x = E\{\underline{x}\,\underline{x}^H\} = \underline{\underline{H}}\,\underline{\underline{R}}_y \underline{\underline{H}}^H + \underline{\underline{R}}_n \qquad r_{\underline{y}\underline{x}} = E\{\underline{y}\,\underline{x}^H\} = \underline{\underline{R}}_y \underline{\underline{H}}^H$$

$$\hat{\underline{y}} = r_{\underline{y}\underline{x}} \underline{\underline{R}}_x^{-1} \underline{x} = \underline{\underline{R}}_y \underline{\underline{H}}^H (\underline{\underline{H}}\,\underline{\underline{R}}_y \underline{\underline{H}}^H + \underline{\underline{R}}_n)^{-1} \underline{x} \rightarrow \text{LMMSE estimator for } \underline{y}.$$

③ Linear Combination of observations ($\underline{\underline{M}}\,\underline{x}$) as a new observation vector:

$$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \longrightarrow \underline{x}_m = \underset{\substack{\downarrow \\ \text{invertible matrix}}}{\underline{\underline{M}}}\,\underline{x}$$

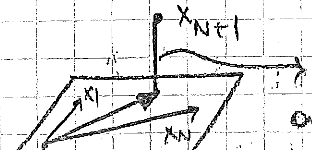
$$\hat{\underline{y}}_m = \underset{\substack{\downarrow \\ E\{\underline{y}(\underline{\underline{M}}\,\underline{x})^H\}}}{r_{\underline{y}\underline{x}_m}}\;\underset{\substack{\downarrow \\ (\underline{\underline{M}}\,\underline{\underline{R}}_x \underline{\underline{M}}^H)^{-1}}}{\underline{\underline{R}}_{x_m}^{-1}}\;\underline{x}_m = E\{\underline{y}\,\underline{x}^H\}\underbrace{\underline{\underline{M}}^H \underline{\underline{M}}^{H^{-1}}}_{\underline{\underline{I}}}\underline{\underline{R}}_x^{-1}\underbrace{\underline{\underline{M}}^{-1}\underline{\underline{M}}}_{\underline{\underline{I}}}\underline{x}$$

$$= r_{\underline{y}\underline{x}} \underline{\underline{R}}_x^{-1} \underline{x}$$

Recursive Estimators

$$\underline{x}_N = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \rightarrow N \text{ observations} \qquad \underline{\underline{R}}_{x_N} \underline{w}^{(N)} = r_{\underline{x}_N y} \qquad \underline{x}_{N+1} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \\ x_{N+1} \end{bmatrix} \qquad \underline{\underline{R}}_{x_{N+1}} \underline{w}^{(N+1)} = r_{x_{N+1}y}$$

desired scalar vector



→ newly added dimension

If x_{N+1} belongs to space spanned by $x_1, \dots, x_N$, result does not change.

$x_{N+1} \notin$ span ↳

$$\begin{array}{c|ccc|c} & x_1 & \cdots & x_N & x_{N+1} \\ x_1 & & & & \\ \vdots & & \underline{\underline{R}}_{x_N} & & \\ x_N & & & (N\times N) & \\ x_{N+1} & & - - - - & & E\{x_{N+1}^2\} \end{array} \begin{bmatrix} w_1^{(N+1)} \\ \vdots \\ w_N^{(N+1)} \\ w_{N+1}^{(N+1)} \end{bmatrix} = \begin{array}{c} x_1 \\ \vdots \\ x_N \\ x_{N+1} \end{array} \begin{bmatrix} E\{x_1 y\} \\ \vdots \\ \\ \\ \end{bmatrix}$$

A

$$\hat{\underline{x}}_{N+1} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \\ \hline x_{N+1} - (\alpha_1 x_1 + \dots + \alpha_N x_N) \end{bmatrix}$$

$$\underbrace{e_{x_{N+1}} = x_{N+1} - \hat{x}_{N+1}}_{\text{estimation error}} \rightarrow \text{freedom in selection of } \alpha_i\text{'s.}$$

try to estimate $\hat{x}_{N+1}(x_N)$ to make "A" terms 0.

$$\underline{\underline{R}}_{\hat{x}_{N+1}} \hat{\underline{w}}^{(N+1)} = r_{\hat{\underline{x}}_{N+1} y}$$

$$\begin{array}{cc} & \begin{array}{ccccc} x_1 & x_2 & \cdots & x_N & x_{N+1}\end{array} \\ \begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_N \\ x_{N+1}\end{array} & \left[\begin{array}{cc} \boxed{\quad (N \times N) \quad} & E\{e_{x_{N+1}} x_1\} \\ \underbrace{E\{e_{x_{N+1}} x_1\} \qquad}_{0} & E\{e^2_{x_{N+1}}\}\end{array}\right] \end{array} \left[\begin{array}{c} \hat{w}_1^{(N+1)} \\ \vdots \\ \hat{w}_N^{(N+1)} \\ \hat{w}_{N+1}^{(N+1)} \end{array}\right] = \left[\begin{array}{c} r_{\underline{x}_N y} \\ \hline E\{e_{x_{N+1}} y\}\end{array}\right]$$

? e var m-1

$$\xrightarrow{\quad} \left[\begin{array}{c} \boxed{\underline{w}^{(N)}} \\ \hline K \end{array}\right]$$

$$\hat{y}_{N+1} = \left(\hat{\underline{w}}^{(N+1)}\right)^H \hat{\underline{x}}_{N+1}$$

$$K = \frac{E\{e_{x_{N+1}} y\}}{E\{e^2_{x_{N+1}}\}}$$

$$\hat{y}_{N+1} = \left(\underline{w}^{(N)}\right)^H \underline{x}_N + K(e_{x_{N+1}})$$

$$\boxed{\hat{y}_{N+1} = \underset{\substack{\downarrow \\ \text{estimate with N observations} \\ \text{(earlier estimate)}}}{\hat{y}_N} + K\Big(\underbrace{x_{N+1} - \hat{x}_{N+1}(\underline{x}_N)}_{\text{innovation}}\Big)}$$

④ Estimation of Linear Combination of desired r.v's from $\underline{x}$

$\underline{x}$: observation vector. $\hat{\underline{y}} = r_{\underline{y}\underline{x}} \underline{R}_x^{-1} \underline{x}$

Assume that we want to estimate $\underline{z} = \underline{N}\,\underline{y}$ from $\underline{x}$.

Q: Is $\hat{\underline{z}} = \underline{N}\,\hat{\underline{y}}$ a good estimator or not?

$$\hat{\underline{z}} = \underbrace{r_{\underline{z}\underline{x}}}_{E\{(\underline{N}\,\underline{y})x^H\} = \underline{N}\, r_{\underline{y}x}} \underline{R}_x^{-1} \underline{x} = \underline{N} \cdot r_{\underline{y}\underline{x}} \cdot \underline{R}_x^{-1} \cdot x = \underline{N} \cdot \hat{\underline{y}}.$$

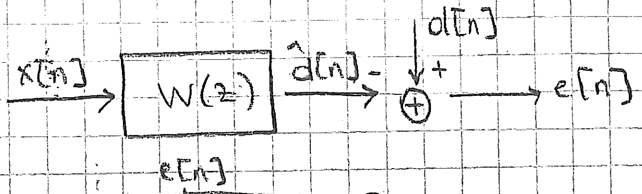
So, indeed, $\hat{\underline{z}} = \underline{N}\,\hat{\underline{y}}$ is the optimal LMMSE estimator.

L36b

136

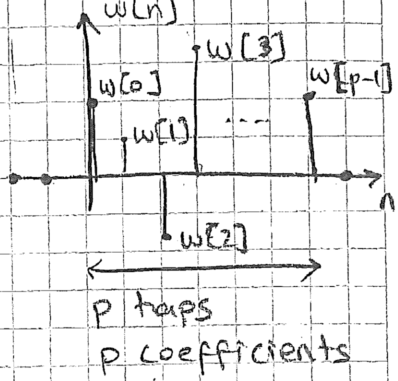
Wiener Filtering

x[n]: observed process } jointly WSS
d[n]: desired process }



The goal is to design W(z) s.t. $\underset{x[n],d[n]}{E}\{\overbrace{(d[n]-\hat{d}[n])}^{e[n]}{}^2\}$ is minimized.

FIR Wiener Filter



$W(z) = w_0 + w_1 z^{-1} + w_2 z^{-2} + \dots + w_{p-1} z^{p-1} \rightarrow$ p tap

Method 1

$\hat{d}[n] = \sum_{k=0}^{p-1} w_k x[n-k]$, $J(\underline{w}) = E\{(d[n] - \sum_{k=0}^{p-1} w_k x[n-k])^2\}$

$\nabla_{\underline{w}} J(\underline{w}) = \underline{0} \rightarrow$ get optimal coefficient equation.

Method 2

$$\hat{d}[n] = [w_0 \ w_1 \ \dots \ w_{p-1}] \begin{bmatrix} x[n] \\ x[n-1] \\ \vdots \\ x[n-(p-1)] \end{bmatrix}$$

$\hookrightarrow (\underline{w}^T \underline{x}[n])$

$\hookrightarrow$ last p samples of x[n] starting from x[n] to x[n-(p-1)]

$J(\underline{w}) = E\{(\underbrace{d[n]}_{y} - \underline{w}^T \underline{x}[n])^2\}$ ← LMMSE problem.

$\underline{\underline{R}}_x \underline{w} = r_{\underline{x}y}$

$$\underline{\underline{R}}_x = E\left\{\begin{bmatrix} x[n] \\ x[n-1] \\ \vdots \\ x[n-(p-1)] \end{bmatrix} [x[n] \ x[n-1] \dots x[n-(p-1)]]\right\} = \begin{array}{c|cccc} & x_n & & & x_{n-(p-1)} \\ \hline x_n & r_x[0] & r_x[1] & \dots & r_x[p-1] \\ & r_x[-1] & r_x[0] & & r_x[p-2] \\ & \vdots & & & \\ x_{n-(p-1)} & r_x[-(p-1)] & & & r_x[0] \end{array}_{p \times p}$$

$$r_{\underline{x}y} = E\left\{\begin{bmatrix} x[n] \\ x[n-1] \\ \vdots \\ x[n-(p-1)] \end{bmatrix} d[n]\right\} = \begin{bmatrix} r_{dx}[0] \\ r_{dx}[1] \\ \vdots \\ r_{dx}[p-1] \end{bmatrix}$$

$r_{dx}[p-1] = r_{xd}[-(p-1)]$

(For the rest of this course, we focus on real valued processes)

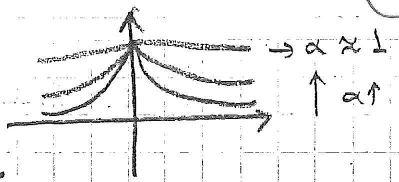
Then, by solving $\underline{\underline{R}}_x \underline{w} = r_{d\underline{x}}$, we get optimal filter coefficients for FIR case.

The MSE value for optimal filtering → $MSE = E\{(\underbrace{d[n] - \underline{w}^T \underline{x}[n]}_{e[n]})^2\}$ ↓ $\underline{w} = \underline{\underline{R}}_x^{-1} r_{d\underline{x}}$

$= E\{e[n]\,d[n]\} - \underline{w}^T \overset{0}{\cancel{E\{e[n]\,\underline{x}[n]\}}} = E\{(d[n])^2\} - \underline{w}^T E\{d[n]\,\underline{x}[n]\}$

$= r_d[0] - \underline{w}^T r_{d\underline{x}} = r_d[0] - r_{d\underline{x}}^T \underline{\underline{R}}_x^{-1} r_{d\underline{x}}$

Example (Hayes)



$x[n] = \underset{\text{desired}}{d[n]} + \underset{\text{noise}}{v[n]}$

$r_d[k] = \alpha^{|k|}$, $0 < \alpha < 1$

$r_v[k] = \sigma_v^2 \delta[k] \Rightarrow$ white noise

d, v : uncorrelated , $E\{d[n]\} = E\{v[n]\} = 0$

Find $\hat{d}[n] = w_0 x[n] + w_1 x[n-1]$ estimator s.t. MSE is minimized.

$\hat{d}[n] = \underline{w}^T \underline{x}[n]$

$r_x[k] = E\{x[n]x[n-k]\} = E\{d[n]d[n-k]\} + E\{v[n]v[n-k]\}$

$\underline{R}_x \underline{w} = r_{d\underline{x}}$

$= r_d[k] + r_v[k] = \alpha^{|k|} + \sigma_v^2 \delta[k]$

$r_{d\underline{x}} = E\{d[n] \underbrace{x[n-k]}_{d[n-k]+v[n-k]}\} = r_d[k] = \alpha^{|k|}$

$$\begin{bmatrix} 1+\sigma_v^2 & \alpha \\ \alpha & 1+\sigma_v^2 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \alpha \end{bmatrix} \longrightarrow \begin{matrix} \alpha = 0.8 \\ \sigma_v^2 = 1 \end{matrix} \longrightarrow \begin{bmatrix} 2 & 0.8 \\ 0.8 & 2 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.8 \end{bmatrix} \to \begin{matrix} w_0 = 0.4048 \\ w_1 = 0.2381 \end{matrix}$$

LMMSE $\to r_d(0) - r_{d\underline{x}}^T \underline{\underline{R}}_x^{-1} r_{d\underline{x}}$ 2-Tap LMMSE $\to 1 - [1 \ 0.8] \begin{bmatrix} 0.4048 \\ 0.2381 \end{bmatrix} = 0.4048$

Let's also compute 1-Tap Wiener Filter and calculate it's MSE.

$\hat{d}[n] = \gamma_0 x[n]$, set γ_0 s.t. $E\{(d[n] - \hat{d}[n])^2\}$ is minimized.

$\underline{R}_x \underline{\gamma} = r_{d\underline{x}} \to r_x(0)\gamma_0 = r_{dx}(0) \to (1+\sigma_v^2)\gamma_0 = 1 \to \gamma_0 = \frac{1}{1+\sigma_v^2} \Big\downarrow_{\substack{\alpha=0.8 \\ \sigma_v^2=1}} = \frac{1}{2}$

1-Tap LMMSE $\to r_d(0) - \underline{w}^T \cdot r_{d\underline{x}} = 1 - \frac{1}{2} = \frac{1}{2}$

Let's also discuss SNR before and after noise removal from observations by Wiener Filtering. $x[n] = \underset{\text{signal}}{d[n]} + \underset{\text{noise}}{v[n]}$

$$(SNR)_{input} = \frac{E\{\text{signal}^2\}}{E\{\text{noise}^2\}} = \frac{r_d(0)}{r_v(0)} = \frac{\alpha^{|k|}}{\sigma_v^2 \delta[k]} \Big\downarrow_{k=0} = \frac{1}{1} = 1 \to 0 \text{ dB}$$

$(SNR)_{dB} = 10 \log_{10} (SNR)_{linear}$

1-Tap Filter ⟶ no improvement in SNR!

$\hat{d}[n] = \gamma x[n] \Big\downarrow_{\gamma=\frac{1}{2}} = \underbrace{\gamma d[n] + \gamma v[n]}_{\text{signal \& noise scaled by same factor}} \longrightarrow (SNR)_{output}^{1\text{-Tap}} = \frac{E\{\cancel{\gamma^2} d^2[n]\}}{E\{\cancel{\gamma^2} v^2[n]\}} = 1 \to 0\text{dB}$

2-Tap Filter

$\hat{d}[n] = \underbrace{[0.4048 \ 0.2381]}_{(\underline{w}^{2\text{-Tap}})^T} \underbrace{\begin{bmatrix} x[n] \\ x[n-1] \end{bmatrix}}_{\underline{x}[n] = \underline{d}[n] + \underline{v}[n]} = \underbrace{(\underline{w}^{2\text{-Tap}})^T \underline{d}[n]}_{\text{Signal}: S} + \underbrace{(\underline{w}^{2\text{-Tap}})^T \underline{v}[n]}_{\text{noise}: \beta}$

$$(SNR)^{2-Tap}_{output} = \frac{E\{\overbrace{(\underline{w}^{2-Tap})^T \underline{d}[n]}^{s} \overbrace{\underline{d}[n]^T \underline{w}^{2-Tap}}^{s^T = s}\}}{E\{\underbrace{(\underline{w}^{2-Tap})^T \underline{v}[n]}_{\beta} \underbrace{\underline{v}[n]^T \underline{w}^{2-Tap}}_{\beta^T=\beta}\}} = \frac{(\underline{w}^{2-Tap})^T \underline{\underline{R}}_d \, \underline{w}^{2-Tap}}{(\underline{w}^{2-Tap})^T \underline{\underline{R}}_v \, \underline{w}^{2-Tap}} \approx 2\,dB$$

Q. What is the maximum SNR for this problem with a 2-Tap filter?

$$\underline{w}_{max\text{-}SNR} = \underset{\underline{w}}{\operatorname{argmax}} \frac{\underline{w}^T \underline{\underline{R}}_d \underline{w}}{\underbrace{\underline{w}^T \underline{\underline{I}}\, \underline{w}}} \rightarrow \|\underline{w}\|^2 = \underset{\substack{\underline{w} \\ \|\underline{w}\|=1}}{\operatorname{argmax}} \; \underline{w}^T \underline{\underline{R}}_d \underline{w}$$

Given an optimal w, I can always scale it to unit norm while SNR stays constant.

maximizing $\underline{w}^T \underline{\underline{R}}_d \underline{w} \rightarrow$ Let's try an eigenvector of $\underline{\underline{R}}_d$

$$\underset{\|\underline{e}_1\|=1}{\underline{\underline{R}}_d \underline{e}_1 = \lambda_1 \underline{e}_1} \rightarrow \underline{w} = \underline{e}_1 \rightarrow \underline{w}^T \underline{\underline{R}}_d \underline{w} \Big|_{\underline{w}=\underline{e}_1} = \underline{w}^T \lambda_1 \underline{e}_1 = \lambda_1 \|\underline{e}_1\|^2 = \lambda_1$$

Let's assume $\underline{w}$ is a combination of two eigenvectors. $\underline{w} = \alpha \underline{e}_1 + \beta \underline{e}_2$ $(\|\underline{w}\|^2 = 1)$

$$\|\underline{w}\|^2 = \underline{w}^T \underline{w} = \alpha^2 + \beta^2 = 1 \rightarrow \underline{w} = \alpha \underline{e}_1 + \sqrt{1-\alpha^2}\, \underline{e}_2$$

$$\underline{w}^T \underline{\underline{R}}_d \underline{w} \Big|_{\underline{w} = \alpha \underline{e}_1 + \sqrt{1-\alpha^2} \underline{e}_2} = \underbrace{\alpha^2 \lambda_1 + (1-\alpha)^2 \lambda_2}_{\rightarrow \text{ weighted average of } \lambda_1 \text{ and } \lambda_2} \leq \max(\lambda_1, \lambda_2)$$

Conclusion The eigenvector of $\underline{\underline{R}}_d$ with maximum eigenvalue is the SNR maximizing solution/filter coefficient.

$$\underline{\underline{R}}_d \rightarrow \begin{array}{cccc} \boxed{\lambda_1} > & \lambda_2 > & \cdots > & \lambda_N \\ \downarrow & \downarrow & & \downarrow \\ \underline{e}_1, & \underline{e}_2, & \cdots, & \underline{e}_N \\ \lambda_{max} & & & \end{array} \qquad \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix} \begin{array}{l} \rightarrow \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \underline{e}_1, \; \lambda_1 = 1.8 \\ \rightarrow \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \underline{e}_2, \; \lambda_2 = 0.2 \end{array}$$

Then, SNR maximizing 2-Tap filter is $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ filter and max

$$(SNR)_{output} = \frac{[1 \;\; 1] \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}}{[1 \;\; 1] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}} = 1.8 = \lambda_1 \approx 2.55\,dB$$

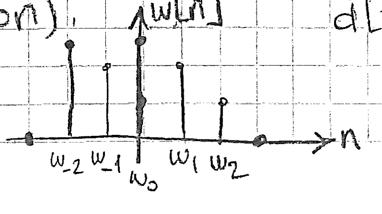
04.01.2021

A Categorization of Signal Processing Operations

1) **Filtering:** A linear combination of input samples not utilizing the input samples in the future (causal operation)

2) **Smoothing:** A linear combination of input samples which are both in the history (past) and future of current sample (anticausal operation).

5-Tap Filter →



$$\hat{d}[n] = [w_{-2} \; w_{-1} \; w_0 \; w_1 \; w_2] \begin{bmatrix} x[n+2] \\ x[n+1] \\ x[n] \\ x[n-1] \\ x[n-2] \end{bmatrix} = \underline{w}^T \underline{x}[n]$$

$$\hookrightarrow w[n] * x[n] = \sum_{k=-2}^{+2} w_k x[n-k]$$

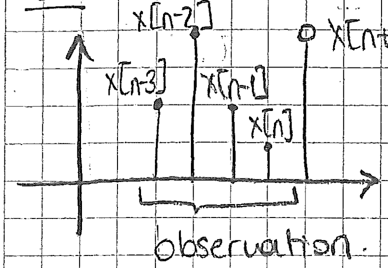
Normal Equations become $\longrightarrow \underline{\underline{R}}_x \underline{w} = r_{d\underline{x}}$

$\nearrow \underline{\underline{R}}_x = E\left\{\begin{bmatrix} x[n+2] \\ \vdots \\ x[n-2] \end{bmatrix} \begin{bmatrix} x[n+2] & \cdots & x[n-2] \end{bmatrix}\right\}$

$\searrow r_{d\underline{x}} = E\left\{ d[n] \begin{bmatrix} x[n+2] \\ \vdots \\ x[n-2] \end{bmatrix} \right\} = \begin{bmatrix} r_{dx}[-2] \\ \vdots \\ r_{dx}[2] \end{bmatrix}$

3) Prediction:

3.1 Forward Prediction



$\hat{x}[n+1] = [w_0 \; w_1 \; w_2 \; w_3] \begin{bmatrix} x[n] \\ x[n-1] \\ x[n-2] \\ x[n-3] \end{bmatrix}$

$d[n] \triangleq x[n+1]$

$\downarrow$

$\underline{\underline{R}}_x \underline{w} = r_{d\underline{x}} \longrightarrow E\left\{ \underbrace{d[n]}_{x[n+1]} \begin{bmatrix} x[n] \\ x[n-1] \\ x[n-2] \\ x[n-3] \end{bmatrix} \right\} = \begin{bmatrix} r_x[1] \\ r_x[2] \\ r_x[3] \\ r_x[4] \end{bmatrix}$

$\text{LmmsE} = r_x(0) - \underline{w}^T r_{d\underline{x}}$

Example (Hayes) Find 2-Tap predictor for $x[n+1]$ in the sense of optimal filtering, i.e. minimizing MSE.

$r_x[k] = \alpha^{|k|}$

$\hat{x}[n+1] = [w_0 \; w_1] \begin{bmatrix} x[n] \\ x[n-1] \end{bmatrix} \quad \underline{\underline{R}}_x \underline{w} = r_{d\underline{x}} \rightarrow \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} \alpha \\ \alpha^2 \end{bmatrix} \rightarrow \begin{matrix} w_0 = \alpha \\ w_1 = 0 \end{matrix}$

Let's interpret this result.

$r_x(k) = \alpha^{|k|} \rightarrow$ AR(1) process $\quad x[n] = \alpha x[n-1] + v[n]$ white noise (σ_v^2)

$r_x[k] = \frac{\sigma_v^2}{1-\alpha^2} \alpha^{|k|}$ So, the $x[n]$ process corresponds to AR(1) process with $\sigma_v^2 = 1-\alpha^2$. (Lec 25 → 27:09)

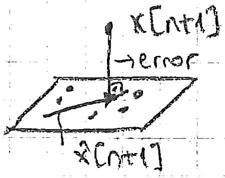
Then, $\hat{x}[n+1] = w_0 x[n] + w_1 x[n-1] = \alpha x[n]$ makes a lot of sense since the process is generated by a similar recursion.

Let's also calculate LmmsE for $\hat{x}[n+1]$.

$\text{LmmsE} \rightarrow \underset{k=0}{\alpha^{|k|}\downarrow} - [\alpha \; 0] \begin{bmatrix} \alpha \\ \alpha^2 \end{bmatrix} = 1 - \alpha^2 \rightarrow$ variance of noise.

Comments Prediction is an important operation, since $x[n+1]$ may have some dependance on earlier samples (correlation of $x[n]$ samples).

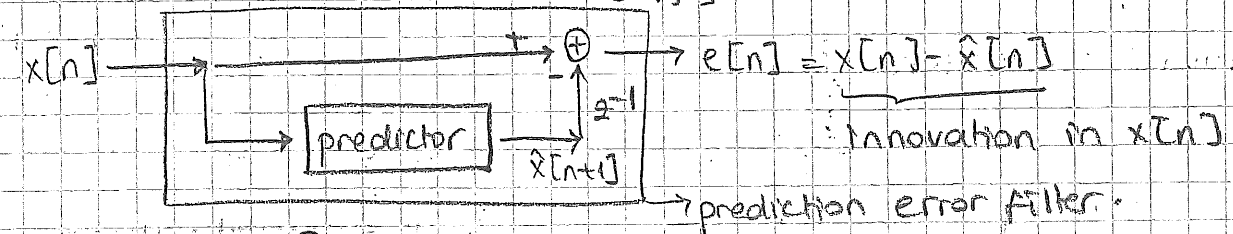
We know that $\rightarrow e[n] = x[n+1] - \underbrace{\hat{x}[n+1]}_{\underline{w}^T \begin{bmatrix} x[n] \\ \vdots \\ x[n-p] \end{bmatrix}} \longrightarrow$



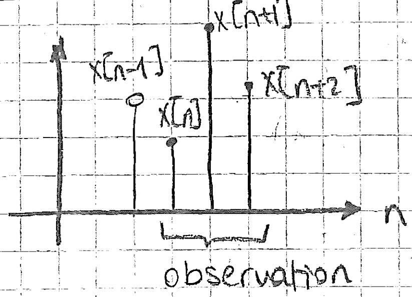
$$\underbrace{E\left\{ e[n] \begin{bmatrix} x[n] \\ \vdots \\ x[n-p] \end{bmatrix}_{(p+1)\times 1} \right\} = \underline{0}}_{\text{Orthogonality principle for LmmsE estimators.}}$$

Then, the prediction error filter $e[n] = x[n+1] - \hat{x}[n+1]$

$$= [1\ \ -w_0\ \ -w_1\ \ \cdots\ \ -w_p] \begin{bmatrix} x[n+1] \\ x[n] \\ \vdots \\ x[n-p] \end{bmatrix}$$



3.2. Backward Prediction



$$\hat{x}[n-1] = \underbrace{[\alpha\ \ \beta\ \ \gamma]}_{\underline{w}^{Backward}} \begin{bmatrix} x[n] \\ x[n+1] \\ x[n+2] \end{bmatrix}$$

$$\underline{\underline{R}}_x\, \underline{w}^{Backward} = \underset{x[n-1]}{r_{d\underline{x}}} = \begin{bmatrix} r_x[1] \\ r_x[2] \\ r_x[3] \end{bmatrix}$$

Backward prediction weights and forward prediction weights are identical. Assume, x[n] process has auto-correlation $r_x[k]$.

reversed sequence → $x^r[n] = x[-n]$ → autocorrelation of $r_{x^r}[k]$?

$r_{x^r}[k] = E\{x^r[n]\, x^r[n-k]\} = E\{x[-n]\, x[-(n-k)]\} = r_x[k]$?

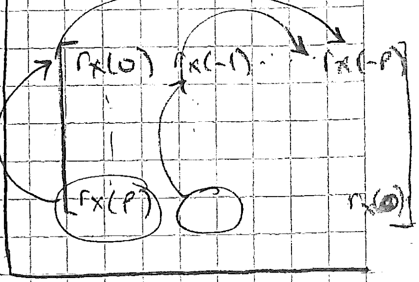
So, forward or backward process has the same autocorrelation.

Remark $\underline{\underline{R}} = E\{\underline{x}\,\underline{x}^T\} \rightarrow \hat{\underline{\underline{R}}} = \frac{1}{N}\sum_{k=1}^{N} \underline{x}_k \underline{x}_k^T \rightarrow$ sample covariance matrix estimator.

$$\hat{\underline{\underline{R}}} = \frac{1}{2N}\left(\sum_{k=1}^{N} \underline{x}_k \underline{x}_k^T + \sum_{k=1}^{N} \underbrace{\underline{x}_k^r}_{\underline{\underline{J}}\underline{x}_k} \underline{x}_k^{r\,T}\right)$$

If $\underline{\underline{R}}$ is Toeplitz and symmetric, i.e. corresponds to an auto-correlation matrix of a WSS process, then $\underline{\underline{J}}\,\underline{\underline{R}}\,\underline{\underline{J}}^T = \underline{\underline{R}}$ where

$$\underline{\underline{J}} = \begin{bmatrix} & 0 & 1 \\ & 1 & 0 \\ 1 & & \end{bmatrix}$$ is the reversal matrix and $\underline{\underline{J}} = \underline{\underline{J}}^T$.



Since $\underline{x}[n]$ vector contains samples of a WSS process, its reversed vector has the same autocorrelation. $\underline{x}^r = \underline{\underline{J}}\,\underline{x} \rightarrow E\{\underline{x}^r \underline{x}^{r\,T}\} = \underline{\underline{J}}\, \underline{\underline{R}}_x\, \underline{\underline{J}}^T = \underline{\underline{R}}_x$

(reversible Markov Chains are related to this.)

IIR Wiener Filtering

$$\hat{d}[n] = \sum_{k'=-\infty}^{+\infty} h[k']\,x[n-k'] \qquad \min_{\substack{h[k] \\ (-\infty<k<\infty)}} \underbrace{E\{(\overbrace{d[n]-\hat{d}[n]}^{e[n]})^2\}}_{J(\underline{h})} \qquad \underline{h} = \begin{bmatrix} h(-\infty) \\ \vdots \\ h(0) \\ \vdots \\ h(+\infty) \end{bmatrix}$$

$$\frac{\partial J(\underline{h})}{\partial h[k]} = E\{-2e[n]\,x[n-k]\} = 0 \longrightarrow \underset{\text{orthogonality condition}}{\underset{\downarrow}{E\{e[n]\,x[n-k]\} = 0}} \quad \forall k$$

$$r_{dx}[k] = \sum_{k'=-\infty}^{+\infty} h[k']\,r_x[k-k'] \quad \forall k$$

→ infinitely many equations with infinitely many unknowns.

DTFT

$$S_{dx}(e^{j\omega}) = H(e^{j\omega})\,S_x(e^{j\omega}) \longrightarrow H^{\overset{\to \text{non-causal}}{IIR-NC}}(e^{j\omega}) = \frac{S_{dx}(e^{j\omega})}{S_x(e^{j\omega})}$$

The achieved mmsE value with $H^{IIR-NC}(e^{j\omega})$:

$$E\{(e[n])^2\}\Big\downarrow_{h[n]=h[n]^{IIR-NC}} = E\{e[n](d[n]-\hat{d}[n])\} = E\{e[n]d[n]\} - \cancel{E\{e[n]\underset{\underset{\sum_{k'} h[k']x[n-k']}{\downarrow}}{\hat{d}[n]}\}}^{\,0}$$

$$r_d(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_d(e^{j\omega})e^{j\omega k}\,d\omega \longrightarrow J_{min}^{IIR-NC} = r_d(0) - \sum_{k'} h[k']\,r_{dx}[k']$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\Big(S_d(e^{j\omega}) - H^{IIR-NC}(e^{j\omega})\underbrace{S_{dx}(e^{j\omega})}_{\text{conjugated, but here: real.}}\Big)d\omega$$

Parseval's Relation:

$$\sum_n x[n]\,y^*[n] = \frac{1}{2\pi}\int_{-\pi}^{+\pi} X(e^{j\omega})\,X^*(e^{j\omega})\,d\omega$$

$\langle \underline{x}, \underline{y} \rangle$ (inner product) $\qquad \langle \underline{\underline{F}}\,\underline{x}, \underline{\underline{F}}\,\underline{y} \rangle = \langle \underline{x}, \underbrace{\underline{\underline{F}}^H\underline{\underline{F}}}_{I}\,\underline{y} \rangle = \langle \underline{x}, \underline{y} \rangle$

Filtering Application for IIR-NC Wiener Filter:

$$\underset{\text{observation}}{x[n]} = \underset{\text{desired}}{d[n]} + \underset{\text{noise}}{v[n]} \qquad r_x[k] = r_d[k] + r_v[k] \xleftrightarrow{DTFT} S_x(e^{j\omega}) = S_d(e^{j\omega}) + S_v(e^{j\omega})$$

uncorrelated

$$r_{dx}[k] = r_d[k] \xleftrightarrow{DTFT} S_{dx}(e^{j\omega}) = S_d(e^{j\omega})$$

$r_x[k]$, $r_{dx}[k]$.

$$H^{IIR-NC}(e^{j\omega}) = \frac{S_{dx}(e^{j\omega})}{S_x(e^{j\omega})} = \frac{S_d(e^{j\omega})}{S_d(e^{j\omega}) + S_v(e^{j\omega})}$$

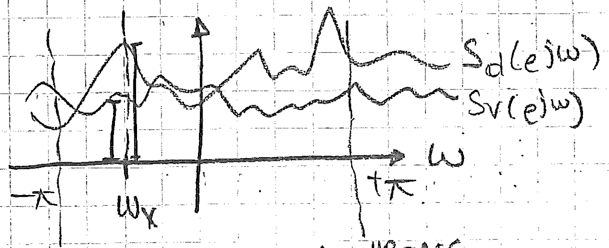
Comments

① $H^{IIR-NC}(e^{j\omega})$ is real valued → $h^{IIR-NC}[n]$ is an even sequence.

② The optimal filter (IIR-NC) processes each frequency "ω_x" independent of other frequencies, i.e. optimal filter design is a decoupled problem in frequency.

On Comment ①, Lec 37 → WSS processes have no forward or backward time, i.e. processes can be reversed without any change in their autocorrelation properties.

On comment ②, Lec 23b → WSS processes are decorrelated by F.T.



$$\frac{S_d(e^{j\omega})}{S_d(e^{j\omega})+S_v(e^{j\omega})} \rightarrow \text{weighted average of powers}$$

Let's calculate J_{min}^{IIR-NC} for the filtering application.

$$J_{min}^{IIR-NC} = \frac{1}{2\pi}\int_{-\pi}^{+\pi}\Big(S_d(e^{j\omega}) - H^{IIR-NC}(e^{j\omega})\underbrace{S_{dx}^{*}(e^{j\omega})}_{S_d(e^{j\omega})}\Big)d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{+\pi} S_d(e^{j\omega})\left[1-\frac{S_d(e^{j\omega})}{S_d(e^{j\omega})+S_v(e^{j\omega})}\right]d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{+\pi}\underbrace{\frac{S_d(e^{j\omega})\,S_v(e^{j\omega})}{S_d(e^{j\omega})+S_v(e^{j\omega})}}_{H^{IIR-NC}(e^{j\omega})} = \frac{1}{2\pi}\int_{-\pi}^{+\pi} H^{IIR-NC}(e^{j\omega})\,S_v(e^{j\omega})\,d\omega$$

If $r_v[k]=\sigma_v^2\delta[k]$ (white noise) $\longleftrightarrow S_v(e^{j\omega})=\sigma_v^2$

$\hookrightarrow J_{min}^{IIR-NC} = \sigma_v^2\, h^{IIR-NC}[0]$

Example

$x[n]=d[n]+v[n]$ $r_d[k]=0.8^{|k|}$ $r_v[k]=\overset{\to 1}{\sigma_v^2}\delta[k]$

Construct IIR-NC filter and compare its MSE value with earlier cases (2-Tap: 0.4048, 1-Tap: 0.5)

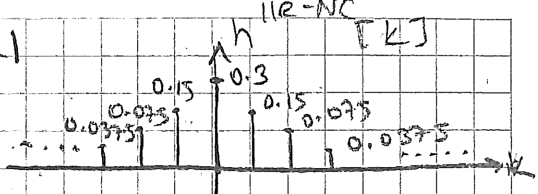
$$\mathcal{Z}\{\alpha^{|k|}\} = \frac{1-\alpha^2}{(1-\alpha z^{-1})(1-\alpha z)}, \quad S_d(z)=\mathcal{Z}\{0.8^{|k|}\}, \quad S_v(z)=\mathcal{Z}\{r_v[k]\}=\sigma_v^2=1$$

$$H^{IIR-NC}(z)=\frac{S_{dx}(z)}{S_x(z)}=\frac{S_d(z)}{S_d(z)+S_v(z)}=\frac{\dfrac{1-\alpha^2}{(1-\alpha z^{-1})(1-\alpha z)}}{\dfrac{1-\alpha^2}{(1-\alpha z^{-1})(1-\alpha z)}+1} \quad \alpha=0.8$$

$$=\frac{0.36}{\underbrace{2(1-0.4z^{-1}-0.4z)}_{-0.8z^{-1}(z-0.5)(z-2)}} = 0.3\,\frac{\left(1-\frac{1}{4}\right)}{\left(1-\frac{1}{2}z^{-1}\right)\left(1-\frac{1}{2}z\right)}$$

$$=1.6\left(1-\tfrac{1}{2}z^{-1}\right)\left(1-\tfrac{1}{2}z\right)$$

$h^{IIR-NC}[n] = \mathcal{Z}^{-1}\{H^{IIR-NC}(z)\} = 0.3 \cdot \frac{1}{2}^{|k|}$



$J_{min}^{IIR-NC} = \sigma_v^2 h^{IIR-NC}[0] \rightarrow$ for white noise

$= 0.3$

Let's calculate J_{min}^{IIR-NC} using time-domain results and compare with $\sigma_v^2 h^{IIR-NC}[0] = 0.3$ value.

$$J_{min}^{IIR-NC} = r_d[0] - \sum_k h^{IIR-NC}[k]\, r_{dx}[k] = 1 - 0.3 \sum_k \frac{1}{2}^{|k|} 0.8^{|k|}$$

$$= 1 - 0.3\left(2\left(\sum_{k=0}^{\infty} 0.4^{|k|}\right) - 1\right) = 1 - 0.3\left(2 \cdot \frac{1}{1-0.4} - 1\right)$$

$$= 1 - 0.3\left(\frac{10}{3} - 1\right) = 0.3_{//}$$

IIR Causal Wiener Filter

$x[n], d[n] \rightarrow$ jointly WSS , $\hat{d}[n] = \sum_{l=0}^{\infty} h[l]\, x[n-l]$, $h[k] = 0$ for $k \leq 0$

$$\frac{\partial}{\partial h[k]} E\{(d[n] - \hat{d}[n])^2\} = E\{e[n]\, x[n-k]\} = 0 , \quad k = \{0, 1, \dots\}$$

$$\downarrow$$

$$r_{dx}[k] = \sum_{l=0}^{\infty} h[l]\, r_x[k-l] , \quad k = \{0, 1, \dots\}$$

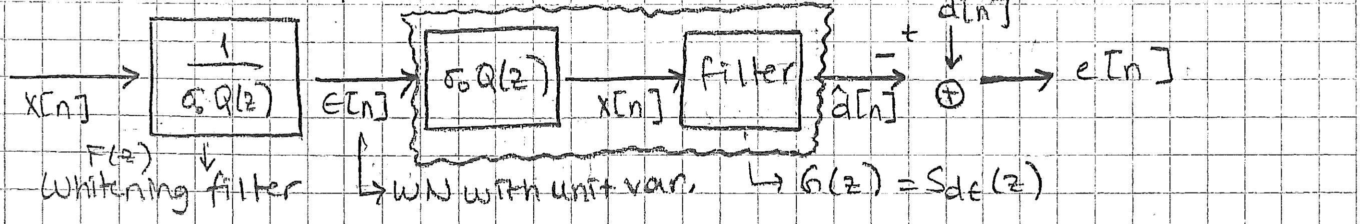
Let's look at the special case of $r_x[k] = \delta[k]$, i.e. $x[n]$ is white noise.

$$r_{dx}[k] = \sum_{l=0}^{\infty} h[l] \underbrace{r_x[k-l]}_{\delta[k-l]} = h[k] , \quad k = \{0, 1, \dots\}$$

So, for the white noise input $\epsilon[n]$, the optimal causal IIR Wiener filter is $h[n] = r_{d\epsilon}[n]$, $n = \{0, 1, \dots\}$. ($\epsilon[n]$: WN input ($x[n]$)) $r_\epsilon[k] = \delta[k]$

$S_x(z) = \sigma_0^2 \underset{\sim}{Q(z)}\, Q^*(1/z^*) \rightarrow$ spectral factorization.

$\hookrightarrow$ min-phase (causal) filter which is a monic polynomial.



Note that, the filter $G(z)$ has $\epsilon[n]$ as input (which is WN) ; hence $g[n]$ can be set to $r_{d\epsilon}[n]$, $n = \{0, 1, \dots\}$ to get IIR causal Wiener filter from $\epsilon[n]$ input for the estimation of $d[n]$.

Let's calculate $r_{d\epsilon}[k]$. $\left(\epsilon[n] = \sum_{\ell=0}^{\infty} f(\ell)\, x[n-\ell]\right)$

$$r_{d\epsilon}[k] = E\{d[n]\,\epsilon^*[n-k]\} = \sum_{\ell=0}^{\infty} f^*(\ell)\, r_{dx}[k+\ell]$$

$$\ell' = -\ell \longrightarrow \; = \sum_{\ell'=-\infty}^{0} f^*(-\ell')\, r_{dx}[k-\ell'] = f^*(-k) * r_{dx}[k]$$

$$S_{d\epsilon}(z) = F^*(1/z^*)\,.\,S_{dx}(z) \longrightarrow G(z) = \lfloor S_{d\epsilon}(z) \rfloor_+$$ - causal part of the argument

$\left(\lfloor 3z + 3 + 4z^{-1} \rfloor_+ = 5 + 4z^{-1} \right)$

Then, $H^{IIR\text{-}causal}(z) = F(z)\,.\,G(z)$

$$= \frac{1}{\sigma_0 Q(z)} \left\lfloor F^*(1/z^*)\, S_{dx}(z) \right\rfloor_+ = \frac{1}{\sigma_0^2 Q(z)} \left\lfloor \frac{S_{dx}(z)}{Q^*(1/z^*)} \right\rfloor_+$$

Let's compare this result with $H^{IIR\text{-}NC}(z)$.

$$H^{IIR\text{-}NC}(z) = \frac{S_{dx}(z)}{S_x(z)} = \frac{S_{dx}(z)}{\sigma_0^2 Q(z)\, Q^*(1/z^*)}$$ non-causal part of the filter.

Let's also write the expression for min MSE

$$J_{min}^{IIR\text{-}causal} = E\{e^2[n]\} = E\{e(n)\,(d(n) - \sum_{\ell=0}^{\infty} h^{IIR\text{-}causal}[\ell]\, x[n-\ell])\} = E\{e[n]\,d[n]\}$$

$$= r_d(0) - \sum_{\ell=0}^{\infty} h^{IIR\text{-}causal}[\ell]\, r_{dx}[\ell]$$

example: $x[n] = d[n] + v[n]$ $\quad r_d[k] = 0.8^{|k|}$ $\quad$ find optimall IIR-causal estimator for $d[n]$.

$r_v[k] = \sigma_v^2 \delta[k]$, $\sigma_v^2 \to 1$

$$S_x(z) = S_d(z) + S_v(z)$$

$$= \mathcal{Z}\{0.8^{|k|}\} + \mathcal{Z}\{\delta[k]\} = \underset{\sigma_0^2}{1.6} \underbrace{\left(\frac{1-0.5z^{-1}}{1-0.8z^{-1}}\right)}_{Q(z)} \underbrace{\left(\frac{1-0.5z}{1-0.8z}\right)}_{Q(1/z)} = \sigma_0^2\, Q(z)\, Q^*(1/z^*)$$

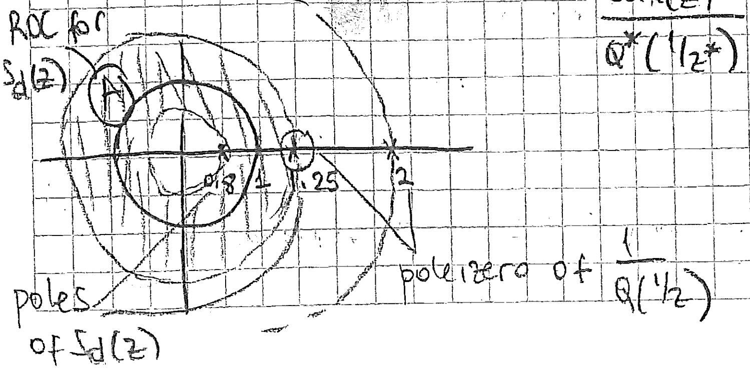
$Q(z)$ → min-phase, monic polynomial.

$$\begin{array}{r|l} 1-0.5z^{-1} & 1-0.8z^{-1} \\ \hline 1-0.8z^{-1} & 1+0.3z^{-1}+0.24z^{-2}+\ldots \\ \hline 0.3z^{-1} & \\ 0.3z^{-1}-0.24z^{-2} & \\ \hline 0.24z^{-2} & \end{array}$$

$1+0.3z^{-1}+0.24z^{-2}+\ldots$ → causal, monic polynomial

$$S_{dx}(z) = S_d(z) = \frac{0.36}{(1-0.8z^{-1})(1-0.8z)}$$

$$\frac{S_{dx}(z)}{Q^*(1/z^*)} = \frac{0.36}{(1-0.8z^{-1})\cancel{(1-0.8z)}} \cdot \frac{1-\cancel{0.8z}}{1-0.5z} = \frac{0.36}{(1-0.8z^{-1})(1-0.5z)}$$



$z \in$ Ⓐ → find the causal part

$$\left\lfloor \frac{0.36}{(1-0.8z^{-1})(1-0.5z)} \right\rfloor_+ = ?$$

$$\frac{0.36\,z^{-1}}{(1-0.8z^{-1})(1-0.5z)z^{-1}} \overset{z\,inv}{=} \frac{0.36\,\boxed{z^{-1}}}{(1-0.8z^{-1})(z^{-1}-0.5)}$$

$$z\{\alpha^n u[n]\} = \frac{1}{1-\alpha z^{-1}}$$

$$z\{-\alpha^n u[-n-1]\} = \frac{1}{1-\alpha z^{-1}}$$

$$= \frac{\frac{0.36\,(1.25)}{1.25-0.5}}{1-0.8z^{-1}} + \frac{\frac{0.36\ (0.5)}{1-0.4}}{z^{-1}-0.5} = \frac{0.6}{1-0.8z^{-1}} + \frac{0.3\ (-2)}{(z^{-1}-0.5)(-2)}$$

$$= \underbrace{\frac{0.6}{1-0.8z^{-1}}}_{\text{causal}} - \underbrace{\frac{0.6}{1-2z^{-1}}}_{\text{anti causal}} \longrightarrow \left[\frac{0.36}{(1-0.8z^{-1})(1-0.5z)}\right]_+ = \frac{0.6}{1-0.8z^{-1}}$$

$$H^{IIR-causal}(z) = \frac{1}{1.6}\,\frac{(1-0.8\cancel{z^{-1}})}{(1-0.5z^{-1})}\cdot\frac{0.6}{(1-0.\cancel{8z^{-1}})} = \frac{3}{8}\,\frac{1}{(1-0.5z^{-1})}$$

$$h^{IIR-causal}[n] = \frac{3}{8}\left(\frac{1}{2}\right)^n u[n]$$

$$J_{min}^{IIR-causal} = r_d(0) - \sum_{l=0}^{\infty} h^{IIR-causal}[l]\,.\,\underbrace{r_{dx}[l]}_{r_d[l]=0.8^{|l|}} = 1 - \sum_{l=0}^{\infty}\frac{3}{8}\,\frac{1}{2}^{l}\,.\,0.8^{l}$$

$$= 1-\frac{3}{8}\sum_{l=0}^{\infty}(0.4)^l = 1-\frac{3}{8}\,\frac{1}{1-0.4} = \frac{3}{8}$$

Let's summarize our findings on this example.

	$h[n]$	J_{min}
1-Tap	$0.5\delta[n]$	0.5
2-Tap	$0.4048\delta[n]+0.2381\delta[n-1]$	0.4048
IIR-Causal	$0.375\left(\frac{1}{2}\right)^n u[n]$	0.375
IIR-Noncausal	$0.3\left(\frac{1}{2}\right)^{\|n\|}$	0.3

Ergodicity

11.01.2021

In our earlier discussions, we have made use of moment information of random processes. In practice, mean, auto-correlation and possibly other moment properties of a r.p. are not known a-priori, but they have to be estimated from data.

Mean Ergodicity (Ergodic in the mean)

$\underset{\sim}{x}[n]$: WSS $\rightarrow \mu_x[n] = E\{x[n]\} = \text{constant} = \mu_x$

An estimator for the mean μ_x can be $\hat{\mu}_x = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$ → sample mean.

Remember that, in ML estimation, we have derived sample mean as the optimal estimator for the parameter estimation of c parameter in the following setting.

$$\underset{\text{observations}}{x[n]} = \underset{\text{constant}}{c} + \underset{\text{noise} \sim N(0,\sigma_w^2),\ \text{iid}}{w[n]} \qquad n = \{0,1,\dots,N-1\} \rightarrow \underline{x} = \underline{1}\,c + \underline{w}$$

non-random

If $\hat{\mu}_x \rightarrow \mu_x$ as $N \rightarrow \infty$, then the process is said to be mean ergodic.

$\underset{\sim}{Q}$: How can we justify the convergence concept for the r.v. $\hat{\mu}_x$?

$\underset{\sim}{A}$: For mean ergodicity, convergence in mean-square sense is taken as the definition. So, for mean ergodicity, we need

1) $E\{\hat{\mu}_x\} \rightarrow \mu_x$ as $N \rightarrow \infty$ (unbiased estimator)
2) $Var\{\hat{\mu}_x - \mu_x\} \rightarrow 0$ as $N \rightarrow \infty$ (consistent estimator)

Let's study the mean ergodicity of arbitrary WSS x[n] processes:

① Unbiasedness condition $\quad E\{\hat{\mu}_x\} \rightarrow \mu_x$ as $N \rightarrow \infty$

$$E\{\hat{\mu}_x\} = E\left\{\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right\} = \frac{1}{N}\sum_{n=0}^{N-1}\mu_x = \mu_x \quad \forall N$$

② Consistency condition $\quad Var\{\hat{\mu}_x - \mu_x\} = ?$

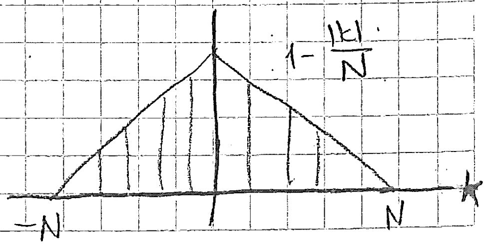
$$\underbrace{\frac{1}{N}\sum_{n=0}^{N-1} x[n]}_{\hat{\mu}_x = \frac{1}{N}\sum_{n=0}^{N-1} x[n]} - \mu_x = \frac{1}{N}\sum_{n=0}^{N-1}(x[n]-\mu_x) = \frac{1}{N}\,\underline{1}^T \underbrace{\begin{bmatrix} x[0]-\mu_x \\ \vdots \\ x[N-1]-\mu_x \end{bmatrix}}_{\underline{v}}$$

$$Var\{\hat{\mu}_x - \mu_x\} = \frac{1}{N^2}\,\underline{1}^T E\{\underline{v}\,\underline{v}^T\}\,\underline{1}$$

$\rightarrow C_x(k) = E\{(x[n]-\mu_x)(x[n-k]-\mu_x)\}$
$= R_x(k) - \mu_x^2$

$$= \frac{1}{N^2}[1\ 1 \dots 1]\begin{bmatrix} C_x(0) & C_x(-1) & \dots & C_x(-N+1) \\ C_x(1) & & & \\ \vdots & & & \\ C_x(N-1) & \dots & \dots & C_x(0) \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$= \frac{1}{N^2}\sum_{k=-N+1}^{N-1}(N-|k|)\,C_x(k) = \frac{1}{N}\sum_{k=-N+1}^{N-1}\left(1-\frac{|k|}{N}\right)C_x(k)$$



Bartlet Window

Then, if $\frac{1}{N}\sum_{k=-(N-1)}^{N-1}\left(1-\frac{|k|}{N}\right)C_x(k) \to 0$ as $N \to \infty$,

then the r.p. $x[n]$ is said to be mean ergodic.

✓ $\frac{1}{N}\sum_{k=-(N-1)}^{N-1}\left(1-\frac{|k|}{N}\right)C_x(k) \to 0 \xleftrightarrow{\text{iff}}$ mean ergodicity.
(as $N \to \infty$)
↳ triangle window
↳ necessary and sufficient condition.

✓ $\frac{1}{N}\sum_{k=0}^{N-1} C_x(k) \to 0 \xleftrightarrow{\text{iff}}$ mean ergodicity
(as $N \to \infty$)
↳ another necessary and sufficient condition.
(rectangle window
(if pos. half → 0, then neg half → 0)

✓ $\lim_{k\to\infty} C_x(k) \to 0 \longrightarrow$ mean ergodicity
↳ sufficient condition.

Example

1st Process

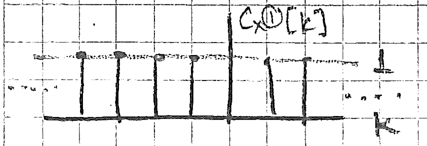
$x^{①}[n] = a \quad a \in \{-1, 1\}$ with 1/2 probability

$E\{x^{①}[n]\} = E\{a\} = 0$

$r_x^{①}[k] = E\{x^{①}[n]\, x^{①}[n-k]\}$

$= E\{a^2\} = 1$

$C_x^{①}[k] = r_x^{①}[k] - \cancel{m_x^2} = 1$



$x^{①}[n]$ is not mean ergodic

2nd Process

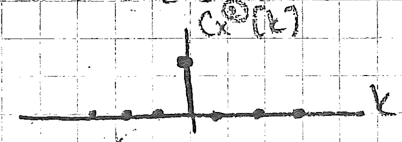
$x^{2}[n] = a_n \quad a_n \in \{-1, 1\}$ with 1/2 prob. and i.i.d.

$E\{x^{②}[n]\} = E\{a_n\} = 0$

$r_x^{②}[k] = E\{x^{②}[n]\, x^{②}[n-k]\} = E\{a_n a_{n-k}\}$

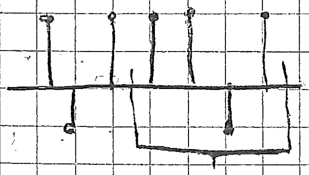
$C_x^{②}[k] = r_x^{②}[k] - \cancel{m_x^2} = \begin{cases} E\{\overset{0}{\cancel{a_n}}\}E\{\overset{0}{\cancel{a_{n-k}}}\}, & k \neq 0 \\ E\{a_n^2\} & , k = 0 \end{cases}$

$= \delta[k]$



$x^{②}[n]$ is mean ergodic.

Sample mean



$x[n] \to$ a realization of this process

→ for $x^{①}[n]$ we are always seeing 1's.

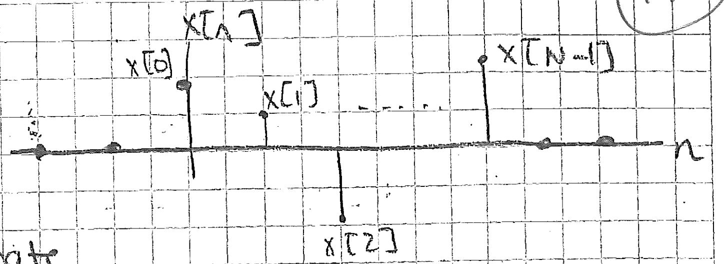
We are never aware of there's possibility of −1's.

ensemble average $\to E\{x[n]\} = \mu_x$ — average of all possible values

time average $\to \frac{\sum_{n=0}^{N-1} x[n]}{N} = \hat{\mu}_x$ — average of a single realization.

if time average → ensemble average as $N \to \infty$, then mean ergodic

Ergodicity in Auto-Correlation



$$\hat{r}_x[k] = \frac{1}{N} \sum_{n=0}^{N-1} x[n]\,x[n-k]$$

Let's check bias of $\hat{r}_x[k]$ estimate

$$E\{\hat{r}_x[k]\} = \frac{1}{N} E\{\sum_{n=0}^{N-1} x[n]\,x[n-k]\}$$

↳ n-k should be between 0 and N-1

$$\downarrow \sum_{n=k}^{N-1}$$

$$= \frac{1}{N} \sum_{n=k}^{N-1} r_x[k] = \frac{N-k-1+1}{N} \cdot r_x[k] = \frac{N-k}{N} r_x[k]$$

So, $\hat{r}_x[k]$ is a biased estimator for $k \neq 0$. To remove the bias, we can use the following: $\hat{r}_x[k] = \frac{1}{N-k} \sum_{n=0}^{N-1} x[n]\,x[n-k]$ → unbiased estimator for $r_x[k]$

In MATLAB: xcorr(x, 'unbiased');
xcorr(x, 'biased');

The biased estimator for $r_x[k]$ is guaranteed to be a valid autocorrelation estimate, while the unbiased one is not guaranteed to be a valid auto-correlation estimate.

In Lec 22, we've said that $\hat{r}_x[k]$ is a valid auto-correlation sequence $\longleftrightarrow$ DTFT$\{\hat{r}_x[k]\} \geq 0$ for all frequencies.

Then, let's check the DTFT of $\hat{r}_x[k]$ biased estimator.

$$\hat{r}_x[k] = \frac{1}{N} \sum_{n=0}^{N-1} x[n]\,x^*[n-k]$$

$\big\downarrow$ DTFT

$$\sum_{k=-\infty}^{+\infty} \hat{r}_x[k] e^{-j\omega k} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \underbrace{\sum_{k=-\infty}^{+\infty} x^*[n-k] e^{-j\omega k}}_{\substack{\text{DTFT}\{x^*[-k]\} \\ \text{shifted by } n \text{ samples} \\ = X^*(e^{j\omega}) e^{-j\omega n}}}$$

$$= \frac{1}{N} X^*(e^{j\omega}) \underbrace{\sum_{n=0}^{N-1} x[n] e^{-j\omega n}}_{X(e^{j\omega})}$$

$$= \frac{1}{N} |X(e^{j\omega})|^2 \geq 0$$

$x[n] \longleftrightarrow X(e^{j\omega})$

$x[-n] \longleftrightarrow X(e^{-j\omega})$

$x^*[-n] \longleftrightarrow X^*(e^{j\omega})$

$$((\sum_n x^*[-n] e^{-j\omega n})^*)^*$$

$$= (\sum_n x[-n] e^{j\omega n})^*$$

$$= (\sum_m x[m] e^{-j\omega m})^*$$

$$= X^*(e^{j\omega})$$

→ To discuss the consistency of $\hat{r_x}[k]$ estimates, we use earlier results on mean autocorrelation

$z_k[n] = x[n]\,x[n-k] \longrightarrow E\{z_k[n]\} = r_x[k]$

if $z_k[n]$ is mean ergodic, then $x[n]$ is auto-correlation ergodic, i.e. a consistent estimator for the auto-correlation.

sufficient condition $\longrightarrow \lim_{l \to \infty} C_{z_k}[l] = 0$

$$\underbrace{E\{z_k[n]\,z_k[n-l]\}}_{E\{x[n]\,x[n-k]\,x[n-l]\,x[n-l-k]\}} - E\{z_k[n]\}\,E\{z_k[n-l]\}$$

So, ergodicity in the auto-correlation requires knowledge of 4^{th} order moments and therefore may not be practical to use/check. For Gaussian processes, there's a simplification to the following:

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} C_x^2[k] = 0 \xleftrightarrow{\text{iff}} \text{ergodicity in auto-correlation}$$

$$E\{x_1 x_2 x_3 x_4\} = E\{x_1 x_2\}E\{x_3 x_4\} + E\{x_1 x_3\}E\{x_2 x_4\} + E\{x_1 x_4\}E\{x_2 x_3\}$$

↳ valid for jointly Normal r.v.

Best Linear Unbiased Estimator (BLUE)

c: desired unknown (non-random)

x_k: k^{th} observation on "c"

$$\hat{c} = [w_1\ w_2] \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\underline{x}} \longrightarrow \text{linear estimator}$$

$\left.\begin{matrix} x_1 = c + n_1 \\ x_2 = c + n_2 \end{matrix}\right\rangle$ zero mean noise with cov. mat. $\underline{R}_n$ $\Big\}$ 2 observations

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} c + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}$$

$$\underline{x} = \underline{1}\,c + \underline{n}$$

MSE: $E\{(c-\hat{c})^2\}$ is aimed to be minimized.

$\epsilon = c - \hat{c} = c - \underline{w}^T\underline{x} = c(1-\underline{w}^T\underline{1}) - \underline{w}^T\underline{n}$, $MSE = E\{\epsilon^2\}$

$E\{\epsilon\} = c(1-\underline{w}^T\underline{1}) - 0 = c(1-\underline{w}^T\underline{1})$ if $E\{\epsilon\} = 0 \longleftrightarrow$ unbiased estimator.

↑ Bias($\underline{w}$, c)

$$E\{\epsilon^2\} = [\text{Bias}(\underline{w},c)]^2 + E\{(\underline{w}^T\underline{n})^2\} - 2E\{\cancelto{0}{\text{Bias}(\underline{w},c)\,\underline{w}^T\underline{n}}\}$$

$$\hookrightarrow \underline{w}^T\underline{n}\,\underline{n}^T\underline{w}$$

$$= [\text{Bias}(\underline{w},c)]^2 + \underline{w}^T \underline{\underline{R_n}}\, \underline{w} = (E\{\epsilon\})^2 + \text{var}\{\epsilon\}$$

Since "c" is non-random, MSE contains the unknown "c" in the cost function, i.e. MSE expression to be minimized!

Then, "optimal" values for $\underline{w}$ MSE minimization depends on "c". This estimator is an <u>unrealizable</u> estimator.

<u>Well</u>, if $\text{Bias}(\underline{w},c)=0 \rightarrow$ MSE becomes a function of noise covariance matrix and $\underline{w}$. So, let's have the optimization problem as $\min_{\underline{w}} \underbrace{E\{(c-\hat{c})^2\}}_{\text{MSE}}$ s.t. $\underline{E\{\hat{c}\}=c}$ $\hookrightarrow$ unbiasedness condition.

$$E\{\hat{c}-c\}=0 \longleftrightarrow c(1-\underline{w}^T\underline{1})=0$$

Then, $\min_{\underline{w}}$ MSE s.t. $w_1+w_2=1 \longleftrightarrow \min_{\underline{w}} \underline{w}^T\underline{\underline{R_n}}\,\underline{w}$ s.t. $w_1+w_2=1$

Solution $\rightarrow$ $J(\underline{w}) = \underline{w}^T\underline{\underline{R_n}}\,\underline{w}$ s.t. $w_1+w_2=1$, minimize $J(\underline{w})$

Lagrange $\longrightarrow$ $L(\underline{w},\lambda) = J(\underline{w}) + \lambda(w_1+w_2-1)$

Let $\underline{\underline{R_n}} = \begin{bmatrix} \sigma_{n_1}^2 & 0 \\ 0 & \sigma_n^2 \end{bmatrix}$

$J(\underline{w}) = w_1^2\sigma_{n_1}^2 + w_2^2\sigma_{n_2}^2$

precision: $\tau_k = 1/\sigma_{n_k}^2$
(reciprocal of variance)

$$\left.\begin{aligned} \frac{\partial}{\partial w_1}L=0 &\longrightarrow 2w_1\sigma_{n_1}^2+\lambda=0 \\ \frac{\partial}{\partial w_2}L=0 &\longrightarrow 2w_2\sigma_{n_2}^2+\lambda=0 \\ \frac{\partial}{\partial \lambda}L=0 &\longrightarrow w_1+w_2=1 \end{aligned}\right] \quad w_1 = \frac{1/\sigma_{n_1}^2}{1/\sigma_{n_1}^2+1/\sigma_{n_2}^2} \qquad w_2 = \frac{1/\sigma_{n_2}^2}{1/\sigma_{n_1}^2+1/\sigma_{n_2}^2}$$

The estimator is $\hat{c}_{BLUE} = \underline{w}^T\underline{x} = \frac{\tau_1}{\tau_1+\tau_2}x_1 + \frac{\tau_2}{\tau_1+\tau_2}x_2 \rightarrow$ <u>**BLUE**</u>

$$E\{(c-\hat{c}_{BLUE})^2\} = \underline{w}^T\underline{\underline{R_n}}\,\underline{w} = w_1^2\sigma_{n_1}^2 + w_2^2\sigma_{n_2}^2 = \frac{\tau_1^2/\tau_1+\tau_2^2/\tau_2}{(\tau_1+\tau_2)^2} = \frac{\tau_1+\tau_2}{(\tau_1+\tau_2)^2}$$

$$\boxed{= \frac{1}{\tau_1+\tau_2} = \frac{1}{1/\sigma_{n_1}^2+1/\sigma_{n_2}^2}} \qquad (\sigma_{n_1}^2 // \sigma_{n_2}^2)$$

In Lec 33, we have solved a very similar problem where "c" is a r.v.

$$\hat{c}_{LMMSE} = \frac{SNR_1 \cdot x_1}{SNR_1+SNR_2+1} + \frac{SNR_2 \cdot x_2}{SNR_1+SNR_2+1} \quad \text{where } SNR_k \triangleq \frac{E\{c^2\}}{E\{n_k^2\}}$$

General Case:

$M > N$

$$\underline{x}_{M\times 1} = \underline{\underline{A}}_{M\times N}\,\underline{y}_{N\times 1} + \underline{n}_{M\times 1}$$

$$\begin{bmatrix} \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \\ \end{bmatrix}\begin{bmatrix} \\ \end{bmatrix} + \begin{bmatrix} \\ \\ \end{bmatrix}$$

$\underline{x}$: observation vector
$\underline{y}$: desired vector
$\underline{n}$: zero mean noise vector with cov. mat. $\underline{\underline{R}}_n$

$\hat{\underline{y}} = \underline{\underline{K}}\,\underline{x}$ Total MSE $\rightarrow E\{\|\underline{y}-\hat{\underline{y}}\|^2\} = E\{\sum_{k=1}^{N}(y_k-\hat{y}_k)^2\}$

$$\underline{e} = \underline{y}-\hat{\underline{y}} = \underbrace{(\underline{\underline{I}}-\underline{\underline{K}}\,\underline{\underline{A}})\underline{y}}_{\text{Bias}(\underline{\underline{K}},\underline{y})} - \underline{\underline{K}}\,\underline{n} \rightarrow E\{\underline{e}\} = (\underline{\underline{I}}-\underline{\underline{K}}\,\underline{\underline{A}})\underline{y}$$

if $(\underline{\underline{I}}-\underline{\underline{K}}\,\underline{\underline{A}}) = \underline{\underline{0}}$, then $\hat{\underline{y}}$ is an unbiased estimator.

$$E\{\|\underline{e}\|^2\} = E\{\underline{e}^T\underline{e}\} = \underline{\text{Bias}}^T(\underline{\underline{K}},\underline{y})\,\underline{\text{Bias}}(\underline{\underline{K}},\underline{y}) + \underbrace{E\{(\underline{\underline{K}}\,\underline{n})^T(\underline{\underline{K}}\,\underline{n})\}}$$

$$= \underline{\text{Bias}}^T\,\text{Bias} + \underbrace{tr(\underline{\underline{K}}\,\underline{\underline{R}}_n\,\underline{\underline{K}}^T)}_{\text{total error variance}}$$

$$= tr(E\{(\underline{\underline{K}}\,\underline{n})^T(\underline{\underline{K}}\,\underline{n})\})$$

$$= E\{tr[(\underline{\underline{K}}\,\underline{n})^T(\underline{\underline{K}}\,\underline{n})]\}$$

$$= E\{tr[\underline{\underline{K}}\,\underline{n}\,\underline{n}^T\underline{\underline{K}}^T]\}$$

$$= tr(\underline{\underline{K}}\,\underline{\underline{R}}_n\,\underline{\underline{K}}^T)$$

So, we see that total MSE depends on unknown $\underline{y}$ vector, unless $\underline{\text{Bias}}(\underline{\underline{K}},\underline{y}) = 0$. So, we focus on unbiased case and minimize the total error variance.

$$\min_{\underline{\underline{K}}} \underbrace{tr(\underline{\underline{K}}\,\underline{\underline{R}}_n\,\underline{\underline{K}}^T)}_{\substack{\text{MSE for unbiased}\\ \text{estimator}}} \quad \text{s.t.} \quad \underbrace{\underline{\underline{I}}-\underline{\underline{K}}\,\underline{\underline{A}} = \underline{\underline{0}}}_{\substack{\text{unbiasedness}\\ \text{condition}}} \quad \rightarrow \underline{\underline{K}}\,\underline{\underline{A}} = \underline{\underline{I}}$$

A: tall matrix
K: may not be unique.

Solution $\rightarrow$ $\underline{\underline{K}}_{BLUE} = (\underline{\underline{A}}^T\underline{\underline{R}}_n^{-1}\underline{\underline{A}})^{-1}\underline{\underline{A}}^T\,\underline{\underline{R}}_n^{-1}$ $\quad \underline{y}_{BLUE} = \underline{\underline{K}}_{BLUE}\cdot\underline{x}$

$$J_{BLUE}(\underline{\underline{K}}) = tr(\underline{\underline{K}}\,\underline{\underline{R}}_n\,\underline{\underline{K}}^T) = tr((\underline{\underline{A}}^T\underline{\underline{R}}_n^{-1}\underline{\underline{A}})^{-1})$$

MSE of BLUE

Special Cases

① $\underline{\underline{R}}_n = \sigma_n^2\,\underline{\underline{I}}$ (noise is white)

$\underline{\underline{A}}\,\underline{y} = \underline{x}$
$\hat{y}_{LS} = (\underline{\underline{A}}^T\underline{\underline{A}})^{-1}\underline{\underline{A}}^T\underline{x}$

$$\hat{\underline{y}}_{BLUE} = (\underline{\underline{A}}^T\underline{\underline{R}}_n^{-1}\underline{\underline{A}})^{-1}\underline{\underline{A}}^T\underline{\underline{R}}_n^{-1}\underline{x} = (\underline{\underline{A}}^T\frac{1}{\sigma_n^2}\underline{\underline{A}})^{-1}\underline{\underline{A}}^T\frac{1}{\sigma_n^2}\underline{x} = (\underline{\underline{A}}^T\underline{\underline{A}})^{-1}\underline{\underline{A}}^T\underline{x} = \hat{\underline{y}}_{LS}$$

If noise is white, then BLUE is the LS solution of $\underline{\underline{A}}\,\underline{y} = \underline{x}$ equation system. (independent of σ_n^2 in $\underline{\underline{R}}_n = \sigma_n^2\underline{\underline{I}}$)

② If $R_n \neq \sigma_n^2 I$ (not white noise)

$$x = A\,y + n \xrightarrow[\text{vector with } R_n^{-1/2}]{\text{Whitening of observation}} \underbrace{R_n^{-1/2}x}_{\hat{x}} = \underbrace{R_n^{-1/2}A\,y}_{\hat{A}\,y} + \underbrace{R_n^{-1/2}n}_{\hat{n}}$$

$R_n = E\{nn^T\}$

$\hat{x} = \hat{A}\,y + \hat{n}$

$\hookrightarrow R_{\hat{n}} = I$

Then, the LS solution on $\hat{x}$ should be the BLUE estimator on x.

$$\hat{y}_{LS}(\hat{x}) = (\hat{A}^T\hat{A})^{-1}\hat{A}^T\hat{x} = (A^T \underbrace{R_n^{-1/2}R_n^{-1/2}}_{R_n^{-1}} A)^{-1} A^T \underbrace{R_n^{-1/2}R_n^{-1/2}}_{R_n^{-1}} x = \hat{y}_{BLUE}(x)$$

So, indeed, the BLUE estimator is the LS estimator after whitening!

Example (earlier example)

$$x = \begin{bmatrix}1\\1\end{bmatrix}c + \begin{bmatrix}n_1\\n_2\end{bmatrix} \quad R_n = \begin{bmatrix}\sigma_{n_1}^2 & 0\\ 0 & \sigma_n^2\end{bmatrix} \rightarrow \text{whitening with } R_n^{-1/2} = \begin{bmatrix}1/\sigma_{n_1} & 0\\ 0 & 1/\sigma_{n_2}\end{bmatrix}$$

$$\begin{bmatrix}1/\sigma_{n_1} & 0\\ 0 & 1/\sigma_{n_2}\end{bmatrix} x = \begin{bmatrix}1/\sigma_{n_1}\\ 1/\sigma_{n_2}\end{bmatrix} c + \begin{bmatrix}n_1/\sigma_{n_1}\\ n_2/\sigma_{n_2}\end{bmatrix}$$

$$\hat{c} = (A^TA)^{-1}A^Tx = \left(\begin{bmatrix}1/\sigma_{n_1} & 1/\sigma_{n_2}\end{bmatrix}\begin{bmatrix}1/\sigma_{n_1}\\ 1/\sigma_{n_2}\end{bmatrix}\right)^{-1}\begin{bmatrix}1/\sigma_{n_1} & 1/\sigma_{n_2}\end{bmatrix}\begin{bmatrix}x_1/\sigma_{n_1}\\ x_2/\sigma_{n_2}\end{bmatrix} = \frac{\tau_1 x_1 + \tau_2 x_2}{\tau_1 + \tau_2}$$

same as before.

Comparison with LMMSE Estimator:

$x = A\,y + n$ assume y is a zero-mean vector with $E\{yy^T\} = R_y$

n is zero-mean noise with R_n

$\hat{y} = K\,x$, $\underset{x,y}{E}\{\|y - \hat{y}\|^2\}$

$\hat{y}_{LMMSE} = R_{yx}\,R_x^{-1}\,x$

$\nearrow R_x = A\,R_y\,A^T + R_n$

$\rightarrow R_{xy} = A\,R_y = R_{yx}^T$

$$\hat{y}_{LMMSE} = R_y A^T (A\,R_y A^T + R_n)^{-1} x$$

$\downarrow$ matrix inversion lemma

$$\hat{y}_{LMMSE} = (A^T R_n^{-1} A + R_y^{-1})^{-1} A^T R_n^{-1} x \quad \Bigg| \quad \hat{y}_{BLUE} = (A^T R_n^{-1} A)^{-1} + A^T R_n^{-1} x$$

If $R_y^{-1} \approx 0 \rightarrow R_y = \begin{bmatrix}\infty & \infty & \cdots\\ \infty & \ddots & \\ \vdots & & \end{bmatrix}$. So, as SNR $\rightarrow \infty$, i.e. desired vector component have much power than noise, $\hat{y}_{LMMSE} \rightarrow \hat{y}_{BLUE}$ as SNR $\rightarrow \infty$

(random par. est.) (non-random par. est.)

(Folk's Thm.)

Karhunen-Loeve Transform (KL Transform)

Case of finite dimensional vector:

$\underline{x}_{N\times 1}$: random vector of interest

The goal is to approximate $\underline{x}$ in a properly selected subspace,

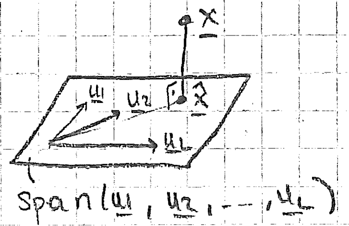
that is $\hat{\underline{x}} = \alpha_1 \underline{u_1} + \alpha_2 \underline{u_2} + \dots + \alpha_L \underline{u_L}$ $(L \leq N)$

The desired $\hat{\underline{x}}$ is

1) to have uncorrelated exponsion coefficients

2) to be a good approximation to $\underline{x}$, i.e. $E\{\|\underline{x}-\hat{\underline{x}}\|^2\}$ should be as small as possible.

In this problem, the unknowns of the problem are $\underline{u_k}$ vectors that is the subspace that $\hat{\underline{x}}$ lies in.



Let's start with 1-Dimensional case:

1-D Case

$\hat{\underline{x}} = \alpha_1 \underline{u_1} \rightarrow \min\limits_{\underline{u_1}} \underbrace{E\{\|\underline{x}-\hat{\underline{x}}\|^2\}}_{\text{MSE for representation error}}$

As discussed before, we may assume that $\underline{u_k}$ vectors form an orthonormal set without any loss of generality. $\|\underline{u_k}\| = 1$ $\forall k$

$\underline{u_k} \perp \underline{u_\ell}$, $k \neq \ell$

Then, $\hat{\underline{x}} = \underline{\underline{P}}_{u_1} \cdot \underline{x}$ → $\|\underline{x}-\hat{\underline{x}}\|^2$ is minimized for a given $\underline{x}$ vector with $\hat{\underline{x}} = \underline{\underline{P}}_{u_1} \cdot x$

$\underline{\underline{P}}_{u_1} = \underline{u_1} \cdot \underline{u_1}^T$

Then, let's write the cost function of the problem: $\uparrow = \underline{\underline{P}}_{u_1}^{c}$

$$MSE = E\{\|\underline{x}-\hat{\underline{x}}\|^2\} = E\{\|\underbrace{(\underline{\underline{I}}-\underline{\underline{P}}_{u_1})}_{\underline{\underline{P}}_{u_1}^{c}}\underline{x}\|^2\} = E\{\|\underline{\underline{P}}_{u_1}^{c} \cdot \underline{x}\|^2\} = E\{\underline{x}^T \underbrace{\underline{\underline{P}}_{u_1}^{c\,T} \underline{\underline{P}}_{u_1}^{c}}_{\underline{\underline{P}}_{u_1}^{c\,2} = \underline{\underline{P}}_{u_1}^{c}} \underline{x}\}$$

$$= E\{\underline{x}^T \underline{\underline{P}}_{u_1}^{c} \underline{x}\} = E\{\underline{x}^T (\underline{\underline{I}}-\underline{\underline{P}}_{u_1})\underline{x}\} = E\{\underline{x}^T \underline{x}\} - E\{\underline{x}^T \underline{\underline{P}}_{u_1} \underline{x}\}$$

$\hookrightarrow \underline{u_1}\,\underline{u_1}^T$

$$= E\{\|\underline{x}\|^2\} - E\{(\underset{a}{\underline{x}^T \underline{u_1}}) \cdot (\underset{b}{\underline{u_1}^T \underline{x}})\} = E\{\|\underline{x}\|^2\} - \underline{u_1}^T \underline{\underline{R}}_x \underline{u_1}$$

$= b \cdot a$

$\hookrightarrow \underline{x}^T \cdot \underline{x} = tr\{\underline{x}^T \underline{x}\} = tr\{\underline{x} \cdot \underline{x}^T\}$

$$= tr(\underline{\underline{R}}_x) - \underline{u_1}^T \underline{\underline{R}}_x \underline{u_1}$$

$\longrightarrow tr(\underline{\underline{R}}_x)$ is independent of $\underline{u_1}$

$$\underline{u_1} = \underset{\substack{\underline{u_1} \\ \|\underline{u_1}\|=1}}{\operatorname{argmin}}\ MSE = \underset{\substack{\underline{u_1} \\ \|\underline{u_1}\|=1}}{\operatorname{argmax}}\ \underline{u_1}^T \underline{\underline{R}}_x \underline{u_1} = \frac{\underline{e_1}}{\underbrace{\|\underline{e_1}\|}_{1}} = \underline{e_1}$$

Note: Eigenvectors of $\underline{\underline{R}}_x$:

$\underline{\underline{R}}_x$'s eigenvectors are ordered in decreasing value of eigenvalues.

$$\begin{matrix} \lambda_1 \geq & \lambda_2 \geq & \cdots \geq & \lambda_N \\ | & | & & | \\ \underline{e_1} & \underline{e_2} & \cdots & \underline{e_N} \end{matrix} \rightarrow \underline{e}_k\text{'s are orthonormal set of vectors.}$$

So, $\underline{e_1}$ is the eigenvector of $\underline{\underline{R}}_x$ with max eigenvalue minimizing MSE. Since, we are working for the 1-D approximation problem, there is only one exponsion coefficient (α_1) (remember $\hat{\underline{x}} = \alpha_1 \underline{u_1}$); there are no other expansion coefficients for α_1 to be uncorrelated. So, the problem of 1-D approximation is solved.

$$MSE_{min}^{1\text{-}D} = E_{\underline{x}}\{\|\underline{x} - \hat{\underline{x}}\|^2\} = \underbrace{tr\{\underline{\underline{R}}_x\}}_{\sum_{k=1}^{N} \lambda_k} - \underline{e_1}^T \underbrace{\underline{\underline{R}}_x \underline{e_1}}_{\lambda_1 \underline{e_1}} = \sum_{k=2}^{N} \lambda_k$$

2-D Case:

$\hat{\underline{x}} = \alpha_1 \underline{u_1} + \alpha_2 \underline{u_2} \rightarrow \|\underline{x} - \hat{\underline{x}}\|$ is minimized with

$$\underline{\underline{P}}_{u_1 u_2} \cdot \underline{x} = \hat{\underline{x}} = (\underline{u_1}\underline{u_1}^T + \underline{u_2}\underline{u_2}^T)\underline{x}$$

$$\|\underline{x} - \underline{\underline{P}}_{u_1 u_2} \cdot \underline{x}\|^2 = \|\underbrace{\underline{x} - \underline{u_1}\underline{u_1}^T\underline{x}}_{\overset{\Delta}{\underline{x}} \ (\underline{u_1} = \underline{e_1} \text{ is known})} - \underline{u_2}\underline{u_2}^T \cdot \underline{x}\|^2$$

With the $\overset{\Delta}{\underline{x}}$, the problem of MSE minimization is immediately solved, since we know from 1-D case, $\underline{u_2}$ should be the eigenvector of $\overset{\Delta}{\underline{\underline{R}}}_x$ with the maximum eigenvalue.

Then, auto-correlation matrix of $\overset{\Delta}{\underline{x}}$ is needed to finalize the solution.

$$\overset{\Delta}{\underline{x}} = \underbrace{(\underline{\underline{I}} - \underline{e_1}\underline{e_1}^T)}_{\underline{\underline{P}}_{e_1}^C}\underline{x} \qquad \overset{\Delta}{\underline{\underline{R}}}_x = \underline{\underline{P}}_{e_1}^C \underline{\underline{R}}_x \underline{\underline{P}}_{e_1}^{C^T} \xrightarrow{\;} = \underline{\underline{P}}_{e_1}^C \quad = (\underline{\underline{I}} - \underline{e_1}\underline{e_1}^T)\underline{\underline{R}}_x(\underline{\underline{I}} - \underline{e_1}\underline{e_1}^T)$$

Q: What are the eigenvectors of $\underline{\underline{\hat{R}_x}}$?

A: Claim:

Eigenvectors of $\underline{\underline{R_x}}$ are also the eigenvectors of $\underline{\underline{\hat{R}_x}}$.

$\underline{\underline{\hat{R}_x}} \cdot \underline{e_1} = 0 \cdot \underline{e_1} = \underline{0}$

$\underline{\underline{\hat{R}_x}} \cdot e_2 = \lambda_2 \underline{e_2} \longrightarrow \lambda_2$ is the max. eigenvalue of $\underline{\underline{\hat{R}_x}}$.

$\vdots$

$\underline{\underline{\hat{R}_x}} \cdot \underline{e_N} = \lambda_N \cdot \underline{e_N}$

So, $\underline{u_2}$ should be selected as $\underline{u_2}$. (eigenvector with second largest eigenvalue of $\underline{\underline{R_x}}$)

Let's also answer the question (Q) by simply expanding the product.

$$\underbrace{\underline{\underline{R_x}} = \sum_{k=1}^{N} \lambda_k \underline{e_k}\,\underline{e_k}^T}_{\text{eigendecomposition}} \longrightarrow \underline{\underline{\hat{R}_x}} = (\underline{\underline{R_x}}) - \underline{\underline{I}}(\underline{\underline{R_x}})\,\underline{e_1}\,\underline{e_1}^T - \underline{e_1}\,\underline{e_1}^T(\underline{\underline{R_x}})\underline{\underline{I}} + (\underline{e_1}\,\underline{e_1}^T)(\underline{\underline{R_x}})(\underline{e_1}\,\underline{e_1}^T)$$

$$= (\underline{\underline{R_x}}) - \underline{\underline{I}}\,\lambda_1 \underline{e_1}\,\underline{e_1}^T - \underline{e_1}\lambda_1 \cancel{\underline{e_1}^T} + \lambda_1 \cancel{\underline{e_1}}\,\underline{e_1}^T$$

$$= \sum_{k=1}^{N} \lambda_k \underline{e_k} \cdot \underline{e_k}^T - \lambda_1 \underline{e_1} \cdot \underline{e_1}^T = \sum_{k=2}^{N} \lambda_k \cdot \underline{e_k} \cdot \underline{e_k}^T$$

Then, in 2-D case $\left\{\begin{matrix} \underline{u_1} = \underline{e_1} \\ \underline{u_2} = \underline{e_2} \end{matrix}\right\}$ i.e. 2 eigenvectors with largest eigenvalues are the basis vectors for MSE minimization and the minimized MSE for 2-D case is:

$$J_{min}^{2\text{-}D} = E\{\|\underline{x} - \hat{\underline{x}}\|^2\} = \sum_{k=3}^{N} \lambda_k = \lambda_3 + \lambda_4 + \dots + \lambda_N$$

$\downarrow$ represent $\underline{x}$, $\hat{\underline{x}}$ with $e_1, \dots, \underline{e_N}$ basis.

$$\text{error} = \underline{x} - \hat{\underline{x}} = \underset{\substack{\downarrow \\ \underline{e_3}^T \underline{x}}}{\beta_3} \underline{e_3} + \beta_4 \underline{e_4} + \dots + \underset{\substack{\downarrow \\ \underline{e_N}^T \underline{x}}}{\beta_N} \underline{e_N}$$

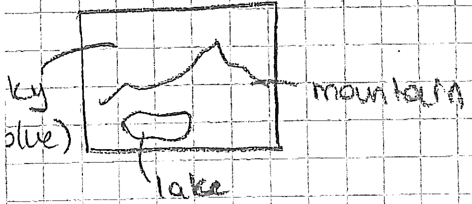
$$E\{\|\underline{x} - \hat{\underline{x}}\|^2\} = E\left\{\sum_{k=3}^{N} (\beta_k)^2\right\} = E\left\{\sum_{k=3}^{N} (\underline{e_k}^T \underbrace{\underbrace{\underline{x})(\underline{x}^T}_{R_x} \underline{e_k})}_{\lambda_k \underline{e_k}}\right\} = \sum_{k=3}^{N} \lambda_k$$

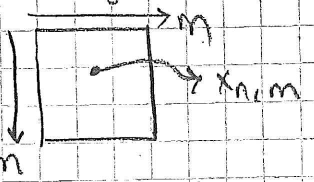
Note that, since $\underline{e_1}$ and $\underline{e_2}$ are eigenvectors of $\underline{\underline{R_x}}$, they form a decorrelating transformation for $\underline{x}$ vector and therefore the expansion coefficients of $\hat{\underline{x}} = \alpha_1 \underline{e_1} + \alpha_2 \underline{e_2}$ are automatically uncorrelated

Comments

① KL Transform selects the set of eigenvectors of $\underline{\underline{R}}_x$ with largest possible eigenvalues and transforms / projects to the space spanned by eigenvectors.

② Application example: Image compression





8x8 pixels

$$x_{n+1,m} = \overset{0.99}{\rho}\, x_{n,m} + \underset{WN}{\text{noise}} \quad \sim AR(1)$$

$$\underline{\underline{\hat{R}}}_x^{\,8\times 8} \rightarrow \begin{matrix} \underline{e}_1 \\ \underline{e}_2 \end{matrix} \Bigg\} \text{ MSE optimum construction}$$

neighboring pixels are highly correlated. (ensemble of blocks)

③ $\underline{\underline{R}}_x$ matrix for WSS processes:

If x[n] is a WSS process, the $\underline{\underline{R}}_x$ matrix is a Toeplitz matrix.

$$\underline{\underline{R}}_x = \begin{bmatrix} r_x(0) & r_x(-1) & \cdots & r_x(-P) \\ r_x(1) & r_x(0) & \ddots & \\ \vdots & \ddots & \ddots & r_x(-1) \\ r_x(P) & \cdots & r_x(1) & r_x(0) \end{bmatrix}_{(P+1)\times(P+1)} \qquad \underline{x} = \begin{bmatrix} x[n] \\ \vdots \\ x[n+P] \end{bmatrix}$$

Then, the DFT is <u>approximately</u> the KL transform of $\underline{x}_{(P+1)\times 1}$ vector, that is the subspace of KL transformation is approximately columns of $(P+1)\times(P+1)$ dimensional DFT matrix. And the eigenvalues of this matrix is approximately the samples of its power spectral density. Remember that, in Lec. 23b, we have discussed the decorrelation of WSS processes with Fourier Transform. Let's assume $\underline{x}$ vector is an infinite dimensional vector (instead of P+1 dimensions). Then,

$$X(e^{jw_1}) = \sum_{n=-\infty}^{+\infty} x[n]e^{-jw_1 n} \qquad x[n]: \text{WSS, zero mean.}$$

Let's calculate the correlation of $X(e^{j\omega_1})$ and $X(e^{j\omega_2})$, $\omega_1 \neq \omega_2$

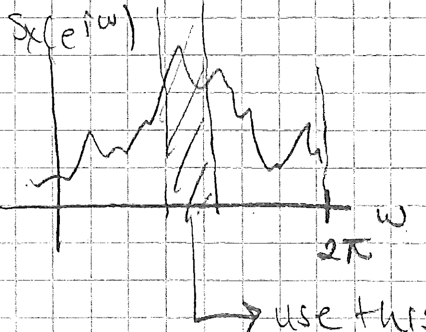
$$E\{X(e^{j\omega_1})X^*(e^{j\omega_2})\} = E\{\sum_{n_1} x[n_1]e^{-j\omega_1 n_1} \sum_{n_2} x^*[n_2]e^{j\omega_2 n_2}\}$$

$$= \sum_{n_1}\sum_{n_2} \underbrace{E\{x[n_1]x^*[n_2]\}}_{r_x[n_1-n_2]} e^{-j\omega_1 n_1} e^{j\omega_2 n_2} = \sum_{n_2} \underbrace{\sum_{n_1} r_x[n_1-n_2]e^{-j\omega_1 n_1}}_{S_x(e^{j\omega_1})e^{-j\omega_1 n_2}} e^{j\omega_2 n_2}$$

$$= S_x(e^{j\omega_1}) \underbrace{\sum_{n_2} 1.e^{-j(\omega_1-\omega_2)n_2}}_{2\pi\delta(\omega_1-\omega_2)} = S_x(e^{j\omega_1})2\pi\delta(\omega_1-\omega_2) = \begin{cases} 2\pi S_x(e^{j\omega_1}), & \omega_1 = \omega_2 \\ 0, & \omega_1 \neq \omega_2 \end{cases}$$

So, DTFT indeed decorrelates $x[n]$ (WSS process) and since DTFT is a unitary transform, then the space of KL Transform coincides with the space of DTFT, that is

$\underline{e}_\omega[n] = e^{+j\omega n}$ $-\infty < n < \infty$ is the infinite dimensional eigenvector of infinite dimensional $\underline{R}_x$ matrix.



→ use this band to capture max energy.

(max eigenvalues → max $S_x(e^{j\omega})$)