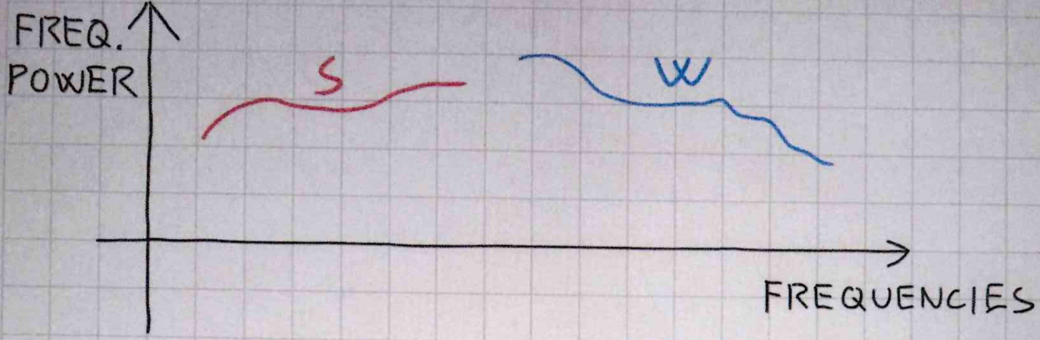
ESTIMATION AND FILTERING: A BRIEF INTRODUCTION

SIGNAL: A FUNCTION THAT DEPENDS ON TIME t, $s(t)$. IT CAN CARRY CRUCIAL INFORMATION FOR MODELING AND CONTROL PURPOSES.

PROBLEM: THEY CAN BE NON DIRECTLY MEASURABLE AND/OR CORRUPTED BY NOISE. CONSIDER THIS LAST CASE.

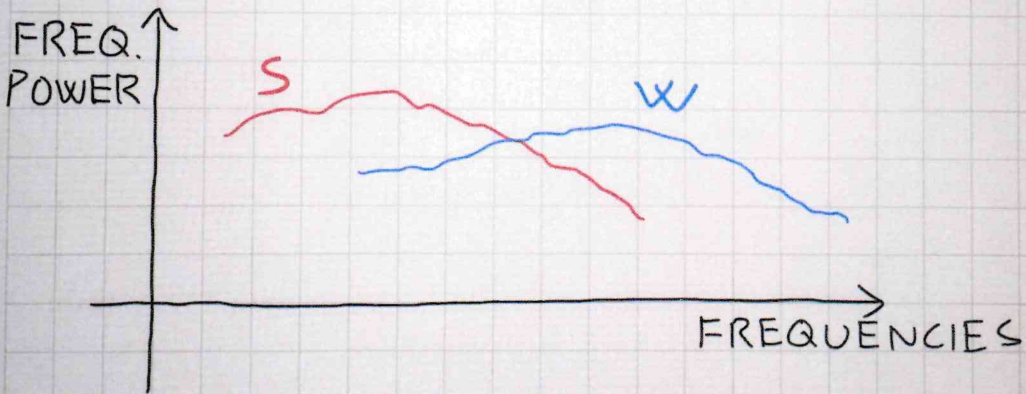
$$\begin{pmatrix}\text{MEASUREMENTS}\\ y\end{pmatrix} = \begin{pmatrix}\text{SIGNAL}\\ s\end{pmatrix} + \begin{pmatrix}\text{NOISE}\\ w\end{pmatrix}$$

ONE SIMPLE SCENARIO IS WHEN s AND w ARE KNOWN TO "LIVE" OVER DIFFERENT FREQUENCIES



A SIMPLE LOW-PASS FILTER CAN BE USED TO RECOVER S FROM Y

BUT A MUCH MORE REALISTIC SCENARIO INVOLVES OVERLAP



HOW TO DIVIDE S FROM W ?

WIENER (1940)

ROM DETERMINISTIC TO STATISTICAL SIGNAL PROCESSING

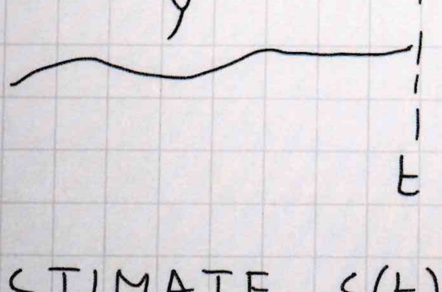
WIENER (1940)

FROM DETERMINISTIC TO STATISTICAL SIGNAL PROCESSING

S, W SEEN AS STOCHASTIC

PROCESSES → COLLECTION OF RANDOM VARIABLES INDEXED BY TIME

- STATISTICAL PROPERTIES OF S, W ARE EMBEDDED IN THEIR PROBABILITY DENSITY FUNCTIONS
- RECOVERING S FROM W BECOMES A STATISTICAL ESTIMATION PROBLEM
- WIENER THEORY → ALLOWS ON-LINE STATISTICAL SIGNAL PROCESSING



ESTIMATE $s(t)$ USING ONLY THE

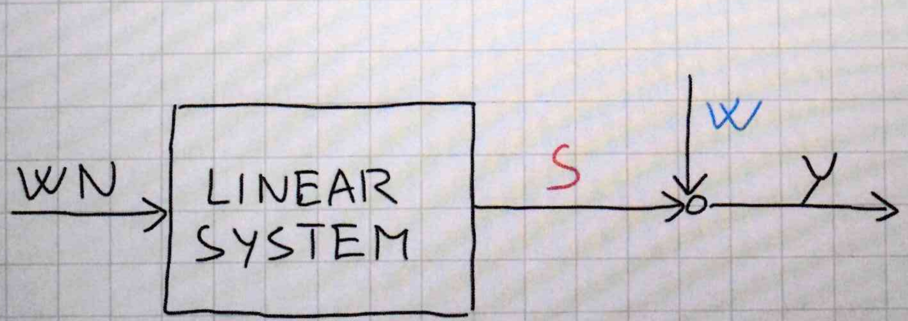
$\{y(\tau)\}_{\tau \leq t}$ (CAUSAL ESTIMATOR)

AND UPDATE EFFICIENTLY THE

ESTIMATE $s(t)$ USING ONLY THE $\{y(\tau)\}_{\tau \le t}$ (CAUSAL ESTIMATOR) AND UPDATE EFFICIENTLY THE ESTIMATE AS t GOES ON AND NEW DATA ARRIVE. THIS IS THE FILTERING PROBLEM

ONE FEATURE AND ONE KEY LIMITATION OF WIENER FILTER

- SIGNALS SEEN AS OUTPUTS OF LINEAR SYSTEMS DRIVEN BY WHITE NOISES. INPUT-OUTPUT VIEW



- ALL THE SIGNALS NEED TO BE

- ALL THE SIGNALS NEED TO BE ASSUMED STATIONARY STOCHASTIC PROCESSES, IMPLYING E.G.

$$E\, s(t) = \mu \quad \forall t$$

$$VAR\, s(t) = \sigma_s^2 \quad \forall t$$

$$VAR\, w(t) = \sigma^2 \quad \forall t$$

FROM WIENER TO KALMAN (1959):

FROM INPUT-OUTPUT TO STATE-SPACE MODELS

WE STILL USE LINEAR SYSTEMS BUT IN STATE-SPACE FORM

$$x(t+1) = A\,x(t) + B\,v(t) + B_u\,u(t)$$

$$y(t) = C\,x(t) + w(t)$$

v, w ARE NOISES (STOCHASTIC PROCESSES)

u CAN BE A DETERMINISTIC INPUT

x, y ARE THUS STOCHASTIC PROCESSES

WE CAN MEASURE THE OUTPUT y BUT NOT THE STATE x WHICH CONTAINS ALL THE INFORMATION ON THE SYSTEM. THE SIGNAL OF INTEREST s CAN BE SEEN AS "CONTAINED" IN x (ONE OR MORE COMPONENTS OF x, OR LINEAR COMBINATIONS OF THE STATES).

EXAMPLE

WE WANT TO MODEL A VEHICLE GOING IN A STRAIGHT LINE.

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \text{POSITION} \\ \text{VELOCITY} \end{bmatrix}$$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \text{POSITION} \\ \text{VELOCITY} \end{bmatrix}$$

u = COMMANDED ACCELERATION

y = MEASURED POSITION

SAY WE CAN CHANGE THE ACCEL. AND MEASURE THE POSITION EVERY T SECONDS

ELEMENTARY PHYSICS LAWS SAY

$$x_2(t+1) = x_2(t) + T\,u(t)$$ VELOCITY EVOLUTION

BUT THIS IS NOT REALISTIC!

VELOCITY WILL BE PERTURBED BY NOISE DUE TO WIND, POT-HOLES, OTHER UNFORTUNATE REALITIES.

THE $x_2(t)$ IS BETTER DESCRIBED AS A RANDOM VARIABLE SUBJECT TO TRANSITION NOISE v:

$$x_2(t+1) = x_2(t) + T\,u(t) + v_2(t)$$

< Note

TO TRANSITION NOISE v:

$$x_2(t+1) = x_2(t) + Tu(t) + v_2(t)$$

SIMILARLY, FOR THE POSITION

$$x_1(t+1) = x_1(t) + Tx_2(t) + \frac{1}{2}T^2u(t) + v_1(t)$$

SINCE THE MEASURED OUTPUT IS THE POSITION CORRUPTED BY SOME MEASUREMENT NOISE $w(t)$,

ONE HAS

$$x(t+1) = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} x(t) + \begin{bmatrix} T^2/2 \\ T \end{bmatrix} u(t) + v(t)$$

$$y(t) = [1 \quad 0]\, x(t) + w(t)$$

IF WE WANT TO CONTROL THE VEHICLE WITH SOME SORT OF FEEDBACK, WE NEED ESTIMATES OF x_1 AND x_2, A STATE ESTIMATOR!

THIS IS WHERE THE KALMAN

$$x(t+1) = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} x(t) + \begin{bmatrix} T^2/2 \\ T \end{bmatrix} u(t) + v(t)$$

$$y(t) = [1 \;\; 0]\, x(t) + w(t)$$

IF WE WANT TO CONTROL THE VEHICLE WITH SOME SORT OF FEEDBACK, WE NEED ESTIMATES OF x_1 AND x_2, A STATE ESTIMATOR!

THIS IS WHERE THE KALMAN FILTER COMES IN: DEEP CONSEQUENCES DUE TO THE STATE-SPACE VIEWPOINT.

LET US HAVE A LOOK AT THE INTRO OF THE ORIGINAL KALMAN PAPER

A New Approach to Linear Filtering and Prediction Problems[1]

R. E. KALMAN
Research Institute for Advanced Study,[2]
Baltimore, Md.

The classical filtering and prediction problem is re-examined using the Bode-Shannon representation of random processes and the "state transition" method of analysis of dynamic systems. New results are:

(1) The formulation and methods of solution of the problem apply without modification to stationary and nonstationary statistics and to growing-memory and infinite-memory filters.

(2) A nonlinear difference (or differential) equation is derived for the covariance matrix of the optimal estimation error. From the solution of this equation the coefficients of the difference (or differential) equation of the optimal linear filter are obtained without further calculations.

(3) The filtering problem is shown to be the dual of the noise-free regulator problem.

The new method developed here is applied to two well-known problems, confirming and extending earlier results.

The discussion is largely self-contained and proceeds from first principles; basic concepts of the theory of random processes are reviewed in the Appendix.

Introduction

AN IMPORTANT class of theoretical and practical problems in communication and control is of a statistical nature. Such problems are: (i) Prediction of random signals; (ii) separation of random signals from random noise; (iii) detection of signals of known form (pulses, sinusoids) in the presence of random noise.

In his pioneering work, Wiener [1][3] showed that problems (i) and (ii) lead to the so-called Wiener-Hopf integral equation; he also gave a method (spectral factorization) for the solution of this integral equation in the practically important special case of stationary statistics and rational spectra.

Many extensions and generalizations followed Wiener's basic work. Zadeh and Ragazzini solved the finite-memory case [2]. Concurrently and independently of Bode and Shannon [3], they also gave a simplified method [2] of solution. Booton discussed the nonstationary Wiener-Hopf equation [4]. These results are now in standard texts [5-6]. A somewhat different approach along these main lines has been given recently by Darlington [7]. For extensions to sampled signals, see, e.g., Franklin [8], Lees [9]. Another approach based on the eigenfunctions of the Wiener-Hopf equation (which applies also to nonstationary problems whereas the preceding methods in general don't), has been pioneered by Davis [10] and applied by many others, e.g., Shinbrot [11], Blum [12], Pugachev [13], Solodovnikov [14].

In all these works, the objective is to obtain the specification of a linear dynamic system (Wiener filter) which accomplishes the prediction, separation, or detection of a random signal.[4]

---

[1] This research was supported in part by the U. S. Air Force Office of Scientific Research under Contract AF 49 (638)-382.

[2] 7212 Bellona Ave.

[3] Numbers in brackets designate References at end of paper.

[4] Of course, in general these tasks may be done better by nonlinear filters. At present, however, little or nothing is known about how to obtain (both theoretically and practically) these nonlinear filters.

Contributed by the Instruments and Regulators Division and presented at the Instruments and Regulators Conference, March 29– April 2, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, February 24, 1959. Paper No. 59—IRD-11.

Present methods for solving the Wiener problem are subject to a number of limitations which seriously curtail their practical usefulness:

(1) The optimal filter is specified by its impulse response. It is not a simple task to synthesize the filter from such data.

(2) Numerical determination of the optimal impulse response is often quite involved and poorly suited to machine computation. The situation gets rapidly worse with increasing complexity of the problem.

(3) Important generalizations (e.g., growing-memory filters, nonstationary prediction) require new derivations, frequently of considerable difficulty to the nonspecialist.

(4) The mathematics of the derivations are not transparent. Fundamental assumptions and their consequences tend to be obscured.

This paper introduces a new look at this whole assemblage of problems, sidestepping the difficulties just mentioned. The following are the highlights of the paper:

(5) *Optimal Estimates and Orthogonal Projections.* The Wiener problem is approached from the point of view of conditional distributions and expectations. In this way, basic facts of the Wiener theory are quickly obtained; the scope of the results and the fundamental assumptions appear clearly. It is seen that all statistical calculations and results are based on first and second order averages; no other statistical data are needed. Thus difficulty (4) is eliminated. This method is well known in probability theory (see pp. 75–78 and 148–155 of Doob [15] and pp. 455–464 of Loève [16]) but has not yet been used extensively in engineering.

(6) *Models for Random Processes.* Following, in particular, Bode and Shannon [3], arbitrary random signals are represented (up to second order average statistical properties) as the output of a linear dynamic system excited by independent or uncorrelated random signals ("white noise"). This is a standard trick in the engineering applications of the Wiener theory [2–7]. The approach taken here differs from the conventional one only in the way in which linear dynamic systems are described. We shall emphasize the concepts of *state* and *state transition;* in other words, linear systems will be specified by systems of first-order difference (or differential) equations. This point of view is

AND NONLINEAR
SYSTEMS?

LAST PART OF THE
COURSE

BASED ON STOCHASTIC
SIMULATION TECHNIQUES
RECENTLY DEVELOPED

COURSE OUTLINE

- Overview of probability theory
- Static Bayesian estimation
- On-line estimation: Wiener filter
- On-line estimation: Kalman filter
- Static Bayesian estimation using rejection sampling and MCMC
- Nonlinear filtering and prediction: particle filters

COURSE OUTLINE

- Overview of probability theory
- Static Bayesian estimation
- On-line estimation: Wiener filter
- On-line estimation: Kalman filter
- Static Bayesian estimation using rejection sampling and MCMC
- Nonlinear filtering and prediction: particle filters

EXAMINATION (two hours)

- Exercise on Static Bayesian Estimation (SBE) and on Kalman Filter (KF): 4 parts, **20 points** overall, minimum **11 points** are required
- Theoretical questions on SBE and KF (convergence theorem), MCMC and particle filters: 2 parts, **11 points** overall, minimum **7 points** are required
- It is mandatory for any student to register in the examination list within the deadline (typically one week before the examination date), no exception to this rule

Università degli Studi di Padova

Uniweb

MENU

» Appelli d'esame » Visualizza appelli

Visualizza appelli

Appelli di: ESTIMATION AND FILTERING [INQ0091318] visualizza dettagli >>

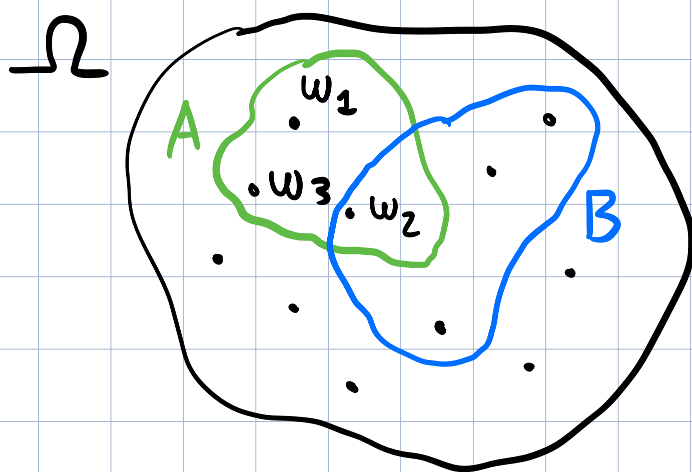
CONTROL SYSTEMS ENGINEERING [IN2546] (LM)...

Elenco appelli d'esame

Nuova prova finale **Nuova prova parziale** Visualizza recenti

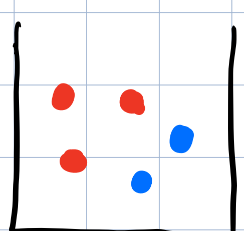
Descrizione		Data, ora e aula	Numero iscritti		Esiti		Verbali caricati		Azioni		
Estimation and Filtering		06/09/2024 13:00									
Estimation and Filtering		10/07/2024 09:00									
Estimation and Filtering		17/06/2024 09:00									

SAMPLE SPACE (SET),
EVENTS (SUBSETS),
PROBABILITY (SUBS → [0,1])



EXAMPLES

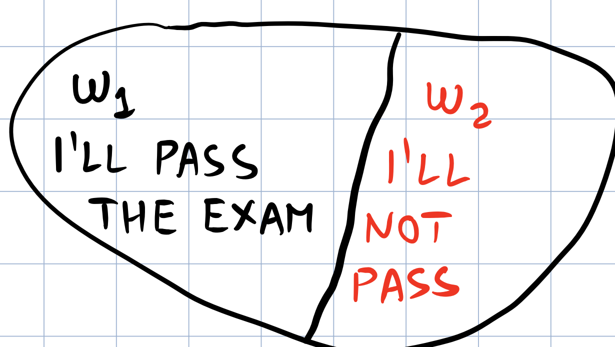
URN



$P(\bullet) = 3/5$

(OBJECTIVE)

STUDENT



$P(\omega_1) = \ldots\ldots$

(SUBJECTIVE)

AXIOMS AND PROPERTIES

$P(\emptyset) = 0, \quad P(\Omega) = 1,$

$P\left(\bigcup_i A_i\right) = \sum_i P(A_i), \quad A_i \cap A_j = \emptyset$
$i \neq j$

$P(A) = 1 - P(\bar{A}), \quad \bar{A} = \Omega \setminus A$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

A, B INDEPENDENT $\iff P(A \cap B) = P(A)P(B)$

(A, B, C IND. $\iff$ $P(A \cap B) = P(A)P(B)$,
$P(A \cap C) = P(A)P(C)$,
$P(B \cap C) = P(B)P(C)$,
$P(A \cap B \cap C) = P(A)P(B)P(C)$

A, B, C, D IND $\iff$ ....)

BAYES RULE

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

• $|B \Rightarrow B$ IS THE "NEW Ω"

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

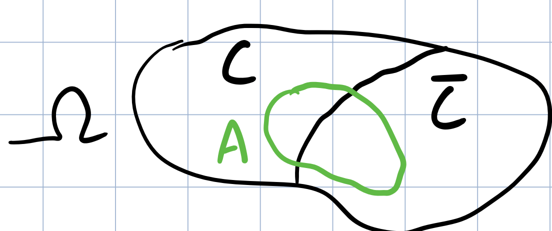
$$\Downarrow$$

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

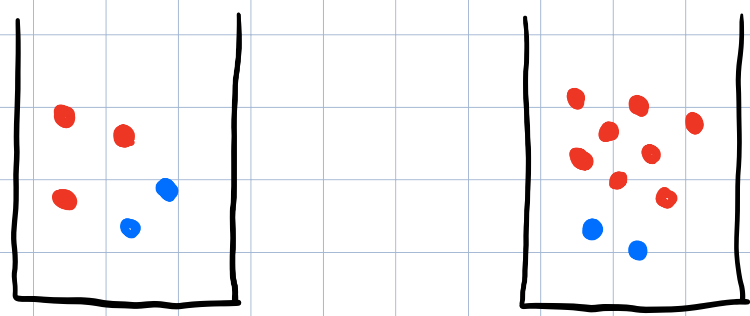
BAYES RULE

$$P(A) = P(A|C)P(C) + P(A|\bar{C})P(\bar{C})$$

TOTAL PROBABILITY



EXERCISE



URN A　　　　URN B

AT EXPERIMENT 1 I RANDOMLY DRAW A OR B, $P(A) = 0.5$

AT EXPERIMENT 2 AND 3 I RANDOMLY DRAW A BALL FROM THE URN SELECTED AT EXPERIMENT 1. THEN I REINSERT THE BALL.

THE "DATA"

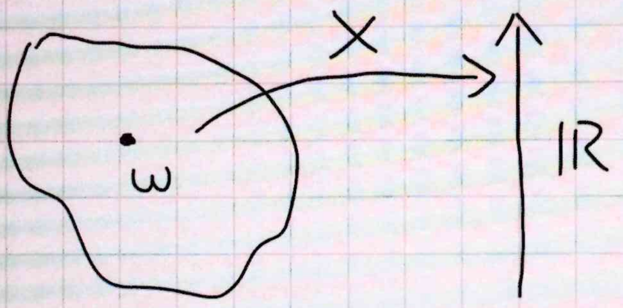
EXPERIMENT 2: IT IS KNOWN THAT A BLUE BALL WAS SELECTED

THE QUESTION

$P(\bullet)$ AT EXPERIMENT 3 = ?

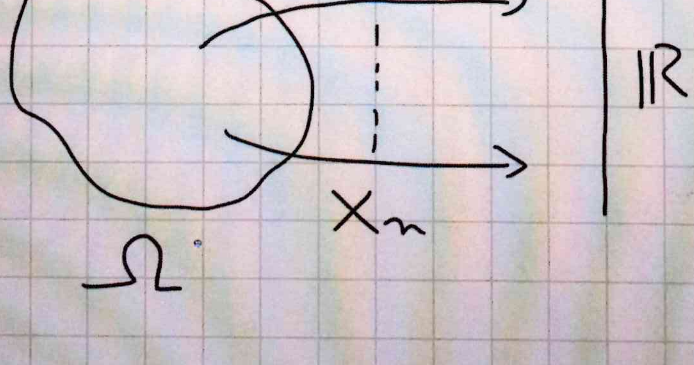
PROBABILITY THEORY

RANDOM VARIABLE



Ω = SAMPLE SPACE

RANDOM VECTOR

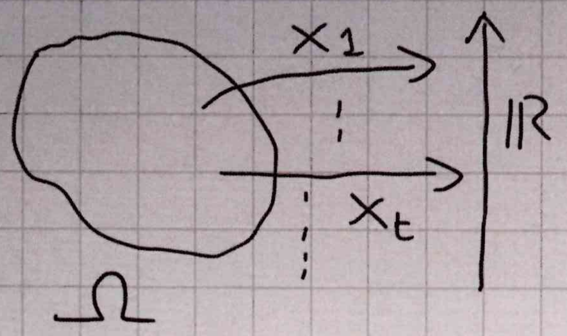


Ω

$$X = [X_1 \ldots X_n]^T$$

COLUMN VECTOR

STOCHASTIC PROCESS



TIME t NOW INDEXES THE R.V.

$$x_t = s(t), \quad t \in \mathbb{Z}$$

PROBABILITY

$X(\omega)$ IN PRACTICE IS NEVER GIVEN.

WE NEED A TOOL TO COMPUTE PROBABILITY OF EVENTS.

LET EACH R.V. BE CONTINUOUS (IT CAN ASSUME AN INFINITE NUMBER OF VALUES) AND X EQUIPPED WITH A PROBABILITY DENSITY

$$p_x(x) = p_x(x_1, \dots, x_n)$$

THEN WE CAN COMPUTE

- PROBABILITY OF EVENTS

$$P(x \in A) = \int_A p_x(x)\,dx$$

- EXPECTATION

$$E[x_i] = \int x_i \, p_x(x)\,dx$$

- EXPECTATION

$$E[x_i] = \int x_i \, p_x(x) \, dx$$

$$E[x] := \begin{bmatrix} E[x_1] \\ \vdots \\ E[x_n] \end{bmatrix}$$

- VARIANCE

$$VAR(x_i) = E\left[(x_i - E x_i)^2 \right]$$

$$= E x_i^2 - (E x_i)^2$$

FUNDAMENTAL EXAMPLE:

GAUSSIAN

SCALAR

$$x \sim N(\mu, \sigma^2)$$

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

VECTOR

$$x \sim N(\mu, \Sigma), \quad \Sigma = \Sigma^T \geq 0, \quad \Sigma \in \mathbb{R}^{n \times n}$$

VECTOR

$x \sim N(\mu, \Sigma)$, $\Sigma = \Sigma^T \geq 0$, $\Sigma \in \mathbb{R}^{n \times n}$

Σ = COVARIANCE, FOR SIMPLICITY FULL RANK BELOW

$$p(x) = \frac{1}{\sqrt{|2\pi \Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

$|\Sigma| = \det \Sigma$

$$Ex = \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}, \quad \text{VAR}\,x = \Sigma$$

JOINTLY GAUSSIAN

x, y ARE JOINTLY GAUSSIAN

IF $z = \begin{bmatrix} x \\ y \end{bmatrix}$ IS A GAUSSIAN VECTOR

$$z \sim N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \Sigma\right), \quad Ex = \mu_x, \quad Ey = \mu_y$$

$$\Sigma = \begin{bmatrix} \Sigma_x & \Sigma_{xy} \end{bmatrix}, \quad \Sigma_{xy} = \Sigma_{yx}^T$$

$$\Sigma = \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{bmatrix}, \quad \Sigma_{xy} = \Sigma_{yx}^T$$

$$\text{VAR } X = \Sigma_x, \quad \text{VAR } Y = \Sigma_y$$

INDEPENDENCE

$\{X_i\}_{i=1}^{n}$ INDEPENDENT

$$\Updownarrow$$

$$p_X(x) = \prod_{i=1}^{n} p_{X_i}(x_i)$$

GAUSSIAN EXAMPLE

$$\Sigma = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{bmatrix} \Rightarrow p(x) = \prod_{i=1}^{n} \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2}}{\sqrt{2\pi}\,\sigma_i}$$

INDEPENDENT $\Longleftrightarrow$ UNCORRELATED

GAUSSIAN
PECULIARITY

QUESTION:

QUESTION:

$X, Y \overset{?}{\Longrightarrow} X+Y$ GAUSSIAN

GAUSSIAN

NO IN GENERAL

YES IF X, Y JOINTLY GAUSSIAN,

E.G. IF X, Y INDEPENDENT

CALCULATIONS OF

EXPECTATIONS AND VARIANCES

- $E[aX_1 + bX_2]$, $a, b \in \mathbb{R}$

$$\| $$

$a E X_1 + b E X_2$

- $E[g(X)] = \int g(x)\, p(x)\, dx$

- $VAR[aX] = a^2 \, VAR\, X$

- $VAR[X_1 + X_2] = VAR\, X_1 + VAR\, X_2$

IF X_1, X_2 UNCORRELATED

NOTE ON UNCORRELATION

WE OFTEN WRITE $x_1 \perp x_2$ TO SAY " x_1 IS UNCORRELATED FROM x_2"

$$x_1 \perp x_2 \iff \mathrm{COV}(x_1, x_2) = 0$$
$$\mathrm{COV}(x_1, x_2) = E(x_1 - \mu_1)(x_2 - \mu_2)$$

OFTEN WE ASSUME ZERO MEAN

$$x_1 \perp x_2 \iff E x_1 x_2 = 0$$

WITH
$E x_1 = E x_2 = 0$

LINEAR AND AFFINE MAPS

IF x IS A RANDOM VECTOR OF DIMENSION n AND $A \in \mathbb{R}^{m \times n}$, ONE HAS

- Ax IS A LINEAR TRANSFORMATION OF X
- $Ax + b$ IS AN AFFINE " " "

$$\mathrm{VAR}(Ax + b) = \mathrm{VAR}\, Ax = A \Sigma_x A^T$$

- $Ax + b$ IS AN AFFINE " " "
- $\text{VAR}(Ax+b) = \text{VAR}\,Ax = A\Sigma_x A^T$

WHERE $\text{VAR}\,x = \Sigma_x$

COVARIANCES FROM LINEAR TRANSFORMATION

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}$$

$A \in \mathbb{R}^{a \times n}$, $B \in \mathbb{R}^{b \times p}$

$$\text{COV}(x,y) = \Sigma_{xy} = \underbrace{E\left[(x-\mu_x)(y-\mu_y)^T\right]}_{\in \mathbb{R}^{n \times p}}$$

Σ_{xy} IN GENERAL IS NOT SYMMETRIC, HENCE IT IS NOT SEMIDEFINITE POSITIVE

$$\text{COV}(Ax, By) = E\left[A(x-\mu_x)(y-\mu_y)^T B^T\right]$$

$$= A\Sigma_{xy}B^T$$

(IF $x = y$ AND $A = B$ THE PREVIOUS FORMULA IS OBTAINED)

NOTE ON $\Sigma \geq 0$

- Σ IS S.D.P., I.E. $\Sigma \geq 0$, IF

  $\Sigma = \Sigma^T$ AND $v^T \Sigma v \geq 0 \quad \forall v \in \mathbb{R}^n$

  $\Updownarrow$

  ALL THE EIGENVALUES OF $\Sigma = \Sigma^T$ ARE ≥ 0

- IF $\Sigma = \text{VAR}X$, THEN $\Sigma \geq 0$

- $\Sigma_1 \geq \Sigma_2$ MEANS $\Sigma_1 - \Sigma_2 \geq 0$

- $\Sigma_1 \geq \Sigma_2 \implies \Sigma_1(i,i) \geq \Sigma_2(i,i) \quad \forall i$

BAYESIAN ESTIMATION

BAYESIAN ESTIMATION

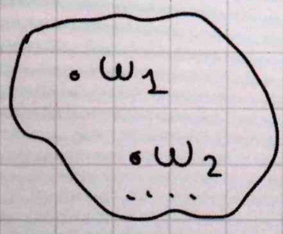
OFTEN A GOOD MODEL FOR m MEASUREMENTS $\{y_K\}_{K=1}^{m}$ IS

$$y_K = h_K(x) + \varepsilon_K$$

WHERE

- h_K IS A KNOWN FUNCTION $\forall K$
- ε_K ARE MEASUREMENT ERRORS

THIS IS DESCRIBED USING A PROBABILITY SPACE



ω_i = OUTCOME OF AN EXPERIMENT

SO, ALL THE ω_i ACCOUNT FOR ALL THE POSSIBLE EXPERIMENTAL CONDITIONS LEADING TO THE y_K

OFTEN, THE ε_K ARE INDEPENDENT ZERO-MEAN GAUSSIAN

KEY POINT: WHAT IS x AND HOW CAN WE OBTAIN IT JUST ASSUMING TO KNOW h, TO KNOW THE PDFS OF $\{\varepsilon_k\}$, AND TO OBSERVE $\{y_k\}$

FISHER APPROACH

x IS A DETERMINISTIC VECTOR

BAYES APPROACH

x IS ALSO A RANDOM VECTOR.

THIS VIEW IS USEFUL SINCE WE CAN INCLUDE "A PRIORI" INFORMATION, I.E. BEFORE SEEING $\{y_k\}$, IN THE PDF $p_x(x)$ OF x

$$p_x(x) = p_x(x_1, \dots, x_n) \quad \text{PRIOR}$$

BAYESIAN ESTIMATION PROBLEM

TO RECONSTRUCT THE RANDOM VECTOR x FROM E.G. THE MODEL

$$y_K = h_K(x) + \varepsilon_K$$

AND THE REALIZATIONS OF THE RANDOM VARIABLES $\{y_K\}_{K=1}^{m}$

FOR A BAYESIAN THE SOLUTION IS THE A POSTERIORI DENSITY FUNCTION (POSTERIOR) THAT IS DEFINED BY THE BAYES RULE

POSTERIOR

INGREDIENTS

- LIKELIHOOD, I.E. THE DENSITY OF $y = [y_1 \dots y_m]$ GIVEN x

  $f(y|x)$ OR $p(y|x)$

  EXAMPLE

  $m = 1$ AND $\varepsilon \sim N(0, \sigma^2)$ AND ε INDEPENDENT OF x

$$f(y|x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{y-h(x)}{\sigma}\right)^2}$$

$$f(y|x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{y-h(x)}{\sigma}\right)^2}$$

- PRIOR DENSITY $p_x(x)$
- MEASUREMENTS y^r (REALIZATIONS OF y)
- BAYES RULE

OFTEN OMITTED

$$p(x \mid y = y^r) = \frac{p(y|x)\, p_x(x)}{p_y(y)}$$

POSTERIOR

$$\propto p(y|x)\, p_x(x)$$

LIKELIHOOD PRIOR

NOTE THAT

$$p_y(y) = \int p_{xy}(x,y)\, dx$$

JOINT DENSITY

$$= \int p(y|x)\, p_x(x)\, dx$$

JOINT DENSITY

$p_y(y)$ BECOMES JUST A NUMBER SETTING $y = y^r$, I.E. IT IS

$p_y(y)$ BECOMES JUST A NUMBER SETTING $y = y^r$, I.E. IT IS A SCALE FACTOR AFTER OBSERVING THE MEASUREMENTS

EXAMPLE

$y = h(x, w)$, x INDEP. OF w, BOTH RANDOM VARIABLES

CALCULATE $p(y|x)$ SAYING IF IT IS NECESSARY TO KNOW BOTH $p_w(w)$ AND $p_x(x)$

SOL.

KNOWLEDGE OF x DOES NOT AFFECT THE DENSITY OF w. SO

$$y = h(x, w) = g_x(w)$$

WHERE $g_x(\cdot)$ IS KNOWN WITHOUT USING p_x.

WE CAN NOW EXPLOIT THE GENERAL FORMULA

$$y = g(w) \Rightarrow p_y(y) = \sum_{\substack{w \\ \text{s.t.}}} \frac{p_w(w)}{\left|\frac{\partial g(w)}{\partial w}\right|}$$

$$y = g(w) \Rightarrow p_y(y) = \sum_{\substack{w \\ \text{s.t.} \\ y = g(w)}} \frac{p_w(w)}{\left| \frac{\partial g(w)}{\partial w} \right|}$$

NOTE: RECALL THAT IN THE CASE OF RANDOM VECTORS THE FORMULA REMAINS CORRECT: IF $g : \mathbb{R}^n \to \mathbb{R}^n$

$$\frac{\partial g}{\partial w} \in \mathbb{R}^{n \times n}, \quad \left. \frac{\partial g}{\partial w} \right|_{ij} = \frac{\partial g_i}{\partial w_j},$$

$|\cdot|$ = ABSOLUTE VALUE OF THE DETERMINANT

IF $g : \mathbb{R}^n \to \mathbb{R}^p$ WITH $p < n$ ONE CAN ADD $n-p$ g_i LIKE $g_i(w) = w_i$ AND STILL USE THE FORMULA

EXAMPLE

LET y_i "$\sim N(\mu, \sigma^2)$" BUT ACTUALLY NOT GAUSSIAN SINCE σ^2 IS A RANDOM VARIABLE

$\{y_i\}^n$ CANNOT BE INDEPENDENT

EXAMPLE

LET y_i "$\sim N(\mu, \sigma^2)$" BUT ACTUALLY NOT GAUSSIAN SINCE σ^2 IS A RANDOM VARIABLE

$\{y_i\}_{i=1}^n$ CANNOT BE INDEPENDENT

BUT WE ASSUME $y_i | \sigma^2$ INDEP.

COMPUTING $p(y|\sigma^2)$, $y = [y_1 \dots y_n]^T$

$$p(y|\sigma^2) = \prod_{i=1}^n p(y_i|\sigma^2)$$

$$= \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} \prod_{i=1}^n e^{-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2}$$

$$= \quad '' \quad \cdot e^{-\frac{S}{2\sigma^2}},$$

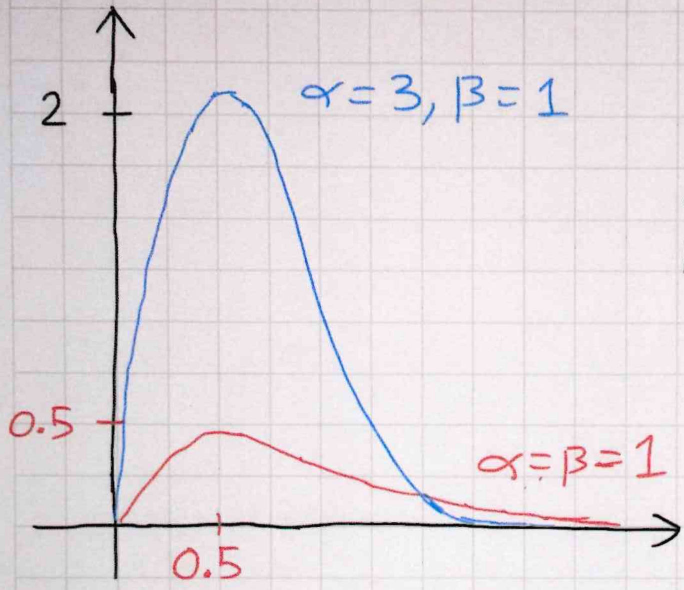
$$S := \sum_{i=1}^n (y_i - \mu)^2$$

LET $\sigma^2 \sim \text{INVGAMMA}(\alpha, \beta)$, THEN COMPUTE THE POSTERIOR OF σ^2

USING $\Theta = \sigma^2$, THE INVGAMMA HAS

USING $\Theta = \sigma^2$, THE INVGAMMA HAS PDF

$$p_\Theta(\theta) \propto \theta^{-\alpha-1} e^{-\beta/\theta}$$



OUR INFO ON σ^2 CAN BE CHANGED E.G. USING DIFFERENT α.
IF $\alpha \downarrow$ THE PRIOR IS FLAT, LESS INFO

(β COULD ALSO VARY TO DESCRIBE OTHER SHAPES)

NOW:

$$p(\sigma^2 | y) \propto p(y|\sigma^2)\, p_{\sigma^2}(\sigma^2)$$

$$\propto \frac{e^{-\frac{S}{2\sigma^2}}}{(\sigma^2)^{n/2}} \cdot (\sigma^2)^{-\alpha-1} e^{-\beta/\sigma^2}$$

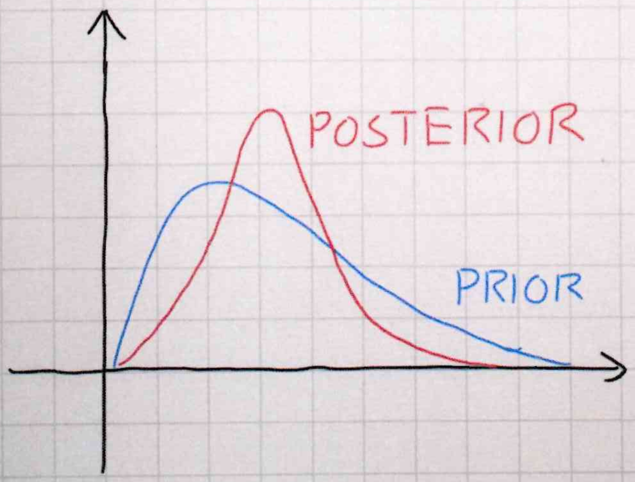
"NEW α" "NEW β"

$$= (\sigma^2)^{-\alpha-\frac{n}{2}-1} e^{-\left(\frac{S}{2}+\beta\right)/\sigma^2}$$

$$\propto \text{INVGAMMA}\left(\alpha + \frac{n}{2}, \beta + \frac{S}{2}\right)$$

NOTE THAT
$$\left.\begin{aligned} \alpha &\to \alpha + \frac{n}{2} \\ \beta &\to \beta + \frac{S}{2} \end{aligned}\right\} \text{BOTH INCREASE}$$

SO THE POSTERIOR TENDS TO BE MORE CONCENTRATED, DECAYING FASTER TO ZERO, WITH A DIFFERENT PEAK REGULATED ALSO BY β



IS THE POSTERIOR IN GENERAL ALWAYS LESS "UNCERTAIN" THAN THE PRIOR?

EXAMPLE

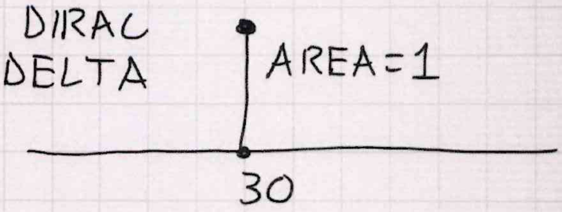
THE SCORE OF A STUDENT (OUTCOME OF AN EXAMINATION) IS A RANDOM VARIABLE V

PDF OF V IS DEFINED AS FOLLOWS:

EVENT A = "SCORE GIVEN BY PROFESSOR #1"

EVENT B = "SCORE GIVEN BY PROF. #2"

$p(V|A) =$



$p(V|B) =$ UNIFORM OVER $\{18, 19, \ldots, 30\}$

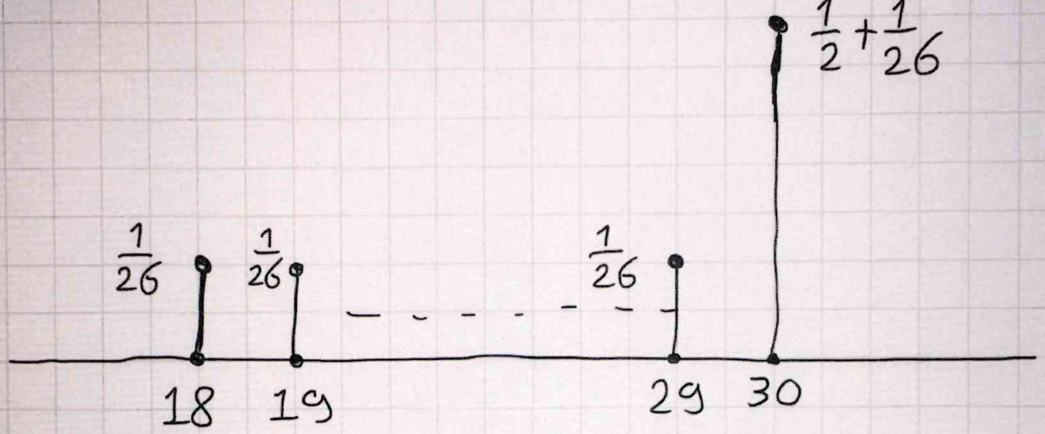
SO THAT $P(V = v|B) = \begin{cases} \frac{1}{13} & \text{IF } v \in \{18, \ldots, 30\} \\ 0 & \text{OTHERWISE} \end{cases}$

#1 GIVES THE SCORE WITH PROB. $1/2$

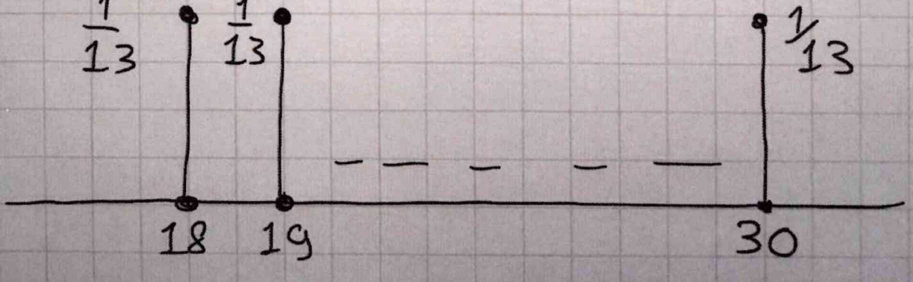
#2 " " " " " $1/2$

THIS DEFINES THE PRIOR ON V

$$p_V(v) = \frac{1}{2} p(v|A) + \frac{1}{2} p(v|B)$$



IF WE CONDITION ON B, THE PRIOR BECOMES THE POSTERIOR $p(V|B)$ WHICH IS THE UNIFORM (MUCH MORE "UNCERTAIN") DENSITY HERE DISPLAYED

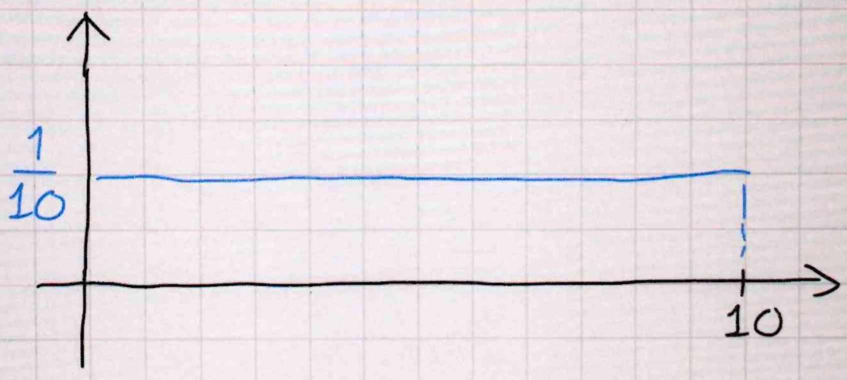


BAYESIAN ESTIMATION:

AN EXAMPLE

$x \sim U(0,10)$

$p_x(x)$



$y = x + \varepsilon$, $\quad \varepsilon \sim N(0,1)$ INDEP. OF x

ASSUME TO OBSERVE $y^2 = 10$

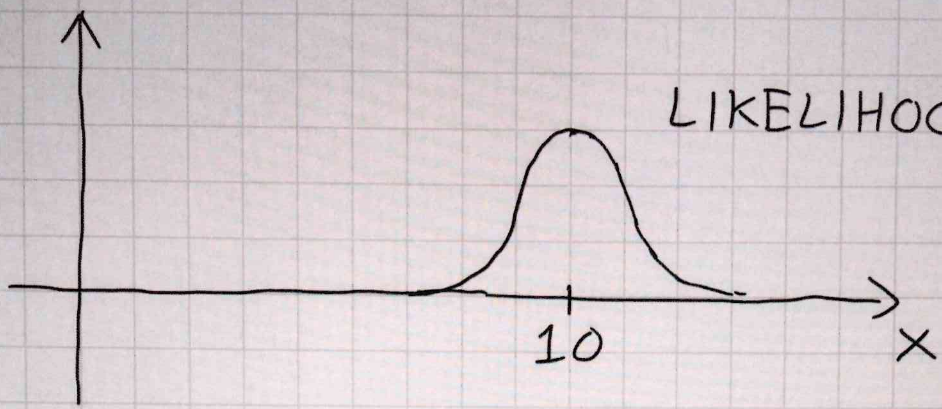
$$p(y|x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-10)^2}$$

$\quad \overset{\shortparallel}{y^2}$

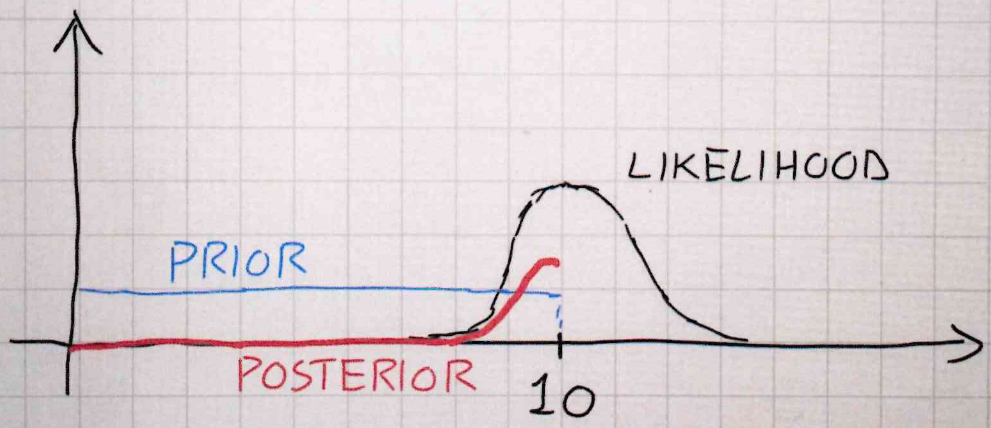
WE PLOT $p(y|x)$ WITH $y=10$

AS A FUNCTION OF x

AS A FUNCTION OF x



THE POSTERIOR IS OBTAINED (APART FROM A NORMALIZATION FACTOR) BY MULTIPLYING BY $p_x(x)$



SUCH POSTERIOR THUS IS:

$$p(x|y=10) = \begin{cases} K e^{-\frac{1}{2}(x-10)^2}, & 0 \leq x \leq 10 \end{cases}$$

SUCH POSTERIOR THUS IS:

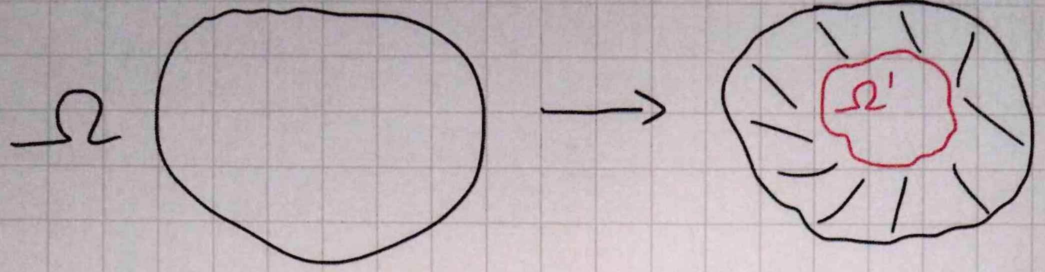
$$p(x|y=10) = \begin{cases} K e^{-\frac{1}{2}(x-10)^2} & , \ 0 \leq x \leq 10 \\ 0 & , \ \text{ELSEWHERE} \end{cases}$$

WITH

$$K = \left(\int_0^{10} e^{-\frac{1}{2}(x-10)^2} dx \right)^{-1}$$

NOTES

- THE POSTERIOR IS THE SOLUTION OF AN ESTIMATION PROBLEM IN A BAYESIAN SETTING
- THE EFFECT OF $|y=10$ HERE (AND IN GENERAL) IS TO REMOVE FROM THE SAMPLE SPACE Ω ALL THE ω NOT COMPATIBLE WITH THE OBSERVATIONS



PROBABILITY IS NOW CONCENTRATED ONLY ON Ω' AND NORMALIZED TO ONE

THE FORMULA

$$p(x|y) = \frac{p(y|x)\, p_x(x)}{p(y)}$$

EXACTLY HAS THIS EFFECT

- OFTEN FROM $p(x|y)$ WE NEED TO OBTAIN POINT ESTIMATES AND BAYES INTERVALS, E.G.

A = BAYES INTERVAL OF 99% LEVEL IF

$$\int_A p(x\,|\,y=y^*)\, dx = 0.99,$$

SO THAT, AFTER SEEING THE

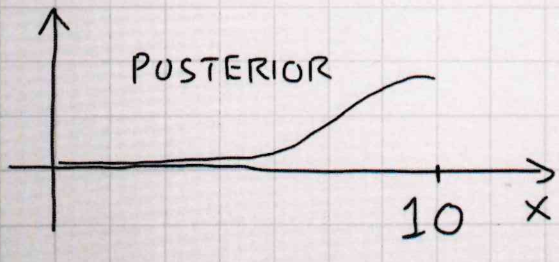
SO THAT, AFTER SEEING THE DATA, $x \in A$ WITH PROBABILITY 0.99,

$$P(x \in A \mid y = y^2) = 0.99$$

POINT ESTIMATES ARE DISCUSSED BELOW

MAP ESTIMATE

$$\hat{x}^{MAP} = \underset{x}{\mathrm{ARGMAX}}\ p(x \mid y = y^2)$$



IN THE PREVIOUS EXAMPLE

$$\hat{x}^{MAP} = 10$$

POSTERIOR MEAN / BAYES ESTIMATE / MINIMUM VARIANCE ESTIMATE

$$\hat{x} = E[x \mid y = y^2]$$

$$= \int x\, p(x \mid y = y^2)\, dx$$

$$\hat{x} = \int_0^{10} x K e^{-\frac{1}{2}(x-10)^2} dx$$ IN THE PREVIOUS EXAMPLE

NOTE THAT $\hat{x}$ CAN BE DIFFICULT TO OBTAIN SINCE IT CAN REQUIRE DIFFICULT INTEGRALS, OFTEN ALSO IN HIGH-DIMENSION

$\hat{x}$ HAS A FUNDAMENTAL PROPERTY.

TO GRASP IT WE HAVE TO MOVE

FROM ESTIMATE (JUST A NUMBER) $\longrightarrow$ ESTIMATOR (RANDOM)

BAYES ESTIMATOR

$\hat{x} = \int x \, p(x|y = y^2) \, dx$ HAS NOW

TO BE THOUGHT NOT AS A

NUMBER BUT AS A FUNCTION OF y^2

$\hat{x}(y^2)$. BUT NOW WE REPLACE

y^2 WITH THE RANDOM VECTOR y

$\hat{x}(y) = \int x \, p(x|y) \, dx$ IS A RANDOM VECTOR

EXAMPLE

$$\hat{x}(y) = \underbrace{\int_0^{10} x \, K \, e^{-\frac{1}{2}(x-y)^2} dx}$$

FUNCTION OF Y, WITH Y TO BE SEEN AS A RANDOM VARIABLE

$y =$ " $U(0,10) + N(0,1)$ "

SUM OF INDEPENDENT R.V. $\Rightarrow$ PDF IS THE CONVOLUTION OF THE TWO PDFs

NOTATION

$\hat{x}(y) = E[x|y]$ = CONDITIONAL MEAN

= ESTIMATOR OF x

= RANDOM VECTOR

MINIMUM VARIANCE ESTIMATOR

STARTING POINT (SCALAR)

x, y RANDOM VARIABLES

WITH KNOWN p_{xy}

AIM: TO RECONSTRUCT x FROM y

AIM: TO RECONSTRUCT x FROM y

SO, WE LOOK FOR A $g(y)$ THAT IS "CLOSE" TO x. MANY TYPES OF DISTANCES, E.G. $E|x-g(y)|$, BUT IT HAS ADVANTAGES TO USE

$$E(x-g(y))^2$$ MEAN SQUARED ERROR

THE OPTIMAL g WILL BE THE MINIMUM VARIANCE ERROR ESTIMATOR

IT SEEMS REALLY HARD TO MINIMIZE W.R.T. g, INSTEAD:

THEOREM: CONSIDER THE PROBLEM

$$\hat{g} = \underset{g}{\mathrm{ARGMIN}} \; E(x-g(y))^2$$

THEN

$$\hat{g}(y) = E[x|y]$$

THEN

$$\hat{g}(y) = E[x|y].$$

SO

- $\hat{g}(y)$ IS THE POSTERIOR MEAN

- POSTERIOR MEAN IS THE MINIMUM VAR. ESTIMATOR

- IT IS UNBIASED!

$$E\left(x - \hat{g}(y)\right) = E\left[E\left[(x - \hat{g}(y))|y\right]\right]$$

$$= E\left[E[x|y] - \hat{g}(y)\right]$$

$$= E\left[E[x|y] - E[x|y]\right]$$

$$= E\,0 = 0$$

PROOF:

$$E\left(x - g(y)\right)^2 = \int \left(x - g(y)\right)^2 p_{xy}(x,y)\,dx\,dy$$

PROOF:

$$E(x-g(y))^2 = \int (x-g(y))^2 \, p_{xy}(x,y)\, dx\, dy$$

$$= \int \left[\int (x-g(y))^2 \, p(x|y)\, dx \right] p_y(y)\, dy$$

$$= E\left[E\left[(x-g(y))^2 | y \right] \right]$$

LET US STUDY THE INNER E

$$E\left[(x-g(y))^2 | y \right]$$

$$= E\left[\left(x - E[x|y] + E[x|y] - g(y) \right)^2 | y \right]$$

$$= E\left[(x - E[x|y])^2 | y \right] \qquad (1)$$

$$+ \; E\left[(E[x|y] - g(y))^2 | y \right] \qquad (2)$$

$$+ \; 2E\Big[\underbrace{(x - E[x|y])}_{\text{HAS ZERO MEAN WHEN CONDITIONED ON Y}} \underbrace{(E[x|y] - g(y))}_{\text{DETERMINISTIC WHEN CONDITIONED ON Y}} | y \Big] \qquad (3)$$

3 = 0, 1 = DOES NOT DEPEND ON g

2 = 0 IF $\forall y$ WE DEFINE g S.T. $g(y) = E[x|y]$

$\blacksquare$

VECTOR CASE

$$(x-g(y))^2 \longrightarrow (x-g(y))^T Q (x-g(y))$$

$$Q = Q^T \geq 0$$

AND ONE HAS

$$\text{OPTIMAL } g = \begin{bmatrix} E[x_1|y] \\ \vdots \\ E[x_n|y] \end{bmatrix} = E[x|y]$$

($Q > 0 \Rightarrow$ SUCH SOLUTION IS UNIQUE)

GAUSSIAN ESTIMATION

LET X, Y BE JOINTLY GAUSSIAN

$$x \sim N(\mu_x, \Sigma_x)$$

$$y \sim N(\mu_y, \Sigma_y), \quad \Sigma_y > 0$$

$$z = \begin{bmatrix} x \\ y \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \Sigma_z\right)$$

$$\Sigma_z = \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{bmatrix}$$

PROPOSITION:

$$E[x|y] = \mu_x + \Sigma_{xy} \Sigma_y^{-1} (y - \mu_y)$$

$$VAR[x|y] = VAR(x - E[x|y])$$

COVARIANCE MATRIX

SCALAR CASE

$$\left(= E(x - E[x|y])^2 \right)$$

$$= \Sigma_x - \Sigma_{xy} \Sigma_y^{-1} \Sigma_{yx}$$

$$VAR[x|y] = VAR(x - E[x|y])$$

COVARIANCE MATRIX

$$\left(\begin{array}{l}\text{SCALAR CASE} \\ = E\left(x - E[x|y]\right)^2\end{array}\right)$$

$$= \Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}$$

(DOES NOT DEPEND ON y!)

PROOF:

$$\bar{z} = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} = \begin{bmatrix} x - \mu_x \\ y - \mu_y \end{bmatrix}$$

$$\tilde{x} = \bar{x} + A\bar{y}$$

$$\tilde{y} = \bar{y}$$

WHICH A MAKES $\tilde{x} \perp \tilde{y}$?

$$cov(\tilde{x}, \tilde{y}) = cov(\bar{x} + A\bar{y}, \bar{y})$$

$$= \Sigma_{xy} + A\Sigma_y$$

SO, IT IS $= 0$ IF

$$A = -\Sigma_{xy}\Sigma_y^{-1}$$

$$A = -\Sigma_{xy}\Sigma_y^{-1}$$

USING SUCH A:

$$E[\bar{x}|\bar{y}] = E[\hat{x} - A\bar{y} \,|\, \bar{y}] \qquad (\bar{y} = \hat{y})$$

$$= E[\hat{x}\,|\,\bar{y}] - E[A\bar{y}\,|\,\bar{y}]$$

$$= E[\hat{x}] - A\bar{y}$$

$$= -A\bar{y}$$

$$= \Sigma_{xy}\Sigma_y^{-1}\,\bar{y}$$

PLUGGING BACK THE MEANS

$$E[x|y] = \mu_x + \Sigma_{xy}\Sigma_y^{-1}(y - \mu_y)$$

DETERMINISTIC IF $|\bar{y}$

$$VAR[\bar{x}|\bar{y}] = VAR(\hat{x} - A\bar{y} \,|\, \bar{y})$$

$$= VAR(\hat{x}\,|\,\bar{y})$$

$$= VAR(\hat{x})$$

$$= VAR(\bar{x} + A\bar{y})$$

$$= COV(\bar{x} + A\bar{y}, \bar{x} + A\bar{y})$$

$$= \mathrm{VAR}(\bar{x} + A\bar{y})$$

$$= \mathrm{COV}(\bar{x} + A\bar{y}, \bar{x} + A\bar{y})$$

$$= \mathrm{VAR}(\bar{x}) + \mathrm{VAR}(A\bar{y}) + \mathrm{COV}(\bar{x}, A\bar{y}) + \mathrm{COV}(A\bar{y}, \bar{x})$$

$$\boxed{A^T = -\Sigma_y^{-1}\Sigma_{yx}}$$

$$= \Sigma_x + A\Sigma_y A^T + \Sigma_{xy} A^T + A\Sigma_{yx}$$

$$= \Sigma_x + \Sigma_{xy}\Sigma_y^{-1}\Sigma_y\Sigma_y^{-1}\Sigma_{yx} - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx} - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}$$

$$= \Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx} \qquad \circledast$$

$\blacksquare$

NOTES:

- $\tilde{x} = x - E[x|y]$ = RECONSTRUCTION ERROR
- $\tilde{x} \perp y$

- $\tilde{x} \perp y$

- $\mathrm{VAR}(\tilde{x}) = \mathrm{VAR}(x - E[x|y])$

$$= \mathrm{VAR}(x - E[x|y] \,|\, y)$$

SINCE $\tilde{x} \perp y$!

$$= \mathrm{VAR}(x|y)$$

- POINTS OUT THAT THE PRIOR VARIANCE (COVARIANCE) Σ_x IS REDUCED BY $\Sigma_{xy} \Sigma_y^{-1} \Sigma_{yx}$

- SUCH TERM IS VAR $E(x|y)$, I.E.

$$\mathrm{VAR}[E(x|y)] = \mathrm{VAR}\left(\mu_x + \Sigma_{xy} \Sigma_y^{-1} (y - \mu_y)\right)$$

$$= \mathrm{VAR}\ \Sigma_{xy} \Sigma_y^{-1} (y - \mu_y)$$

$$= \Sigma_{xy} \Sigma_y^{-1} \Sigma_y \Sigma_y^{-1} \Sigma_{yx}$$

$$= \Sigma_{xy} \Sigma_y^{-1} \Sigma_{yx}$$

MSE = MEAN SQUARED ERROR

- MSE = MEAN SQUARED ERROR

$$MSE_{\substack{\text{GAUSSIAN} \\ \text{ESTIMATOR}}} = TRACE\left[VAR\left(x - E[x|y]\right)\right]$$

$$= TRACE\left[VAR\left(x|y\right)\right]$$

$$= \sum_i E\left(x_i - E[x_i|y]\right)^2$$

POSTERIOR MEAN AND VARIANCE: ADDITIONAL REMARKS

x, y $\quad p_{xy}$ KNOWN, x RANDOM VAR. (FOR SIMPLICITY)

y = RANDOM VECTOR, EACH EXPERIMENT GIVES DIFFERENT REALIZATIONS

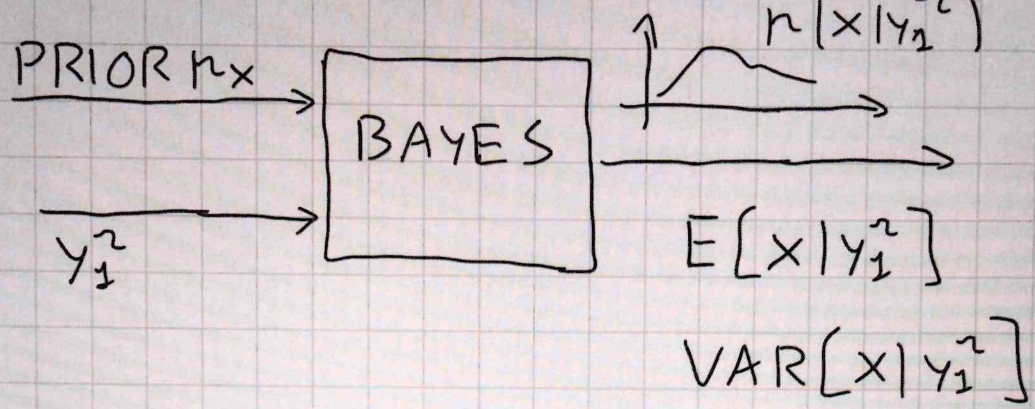
1^{ST} EXPERIMENT $\quad y_1^2$

$\vdots \qquad\qquad\qquad \vdots$

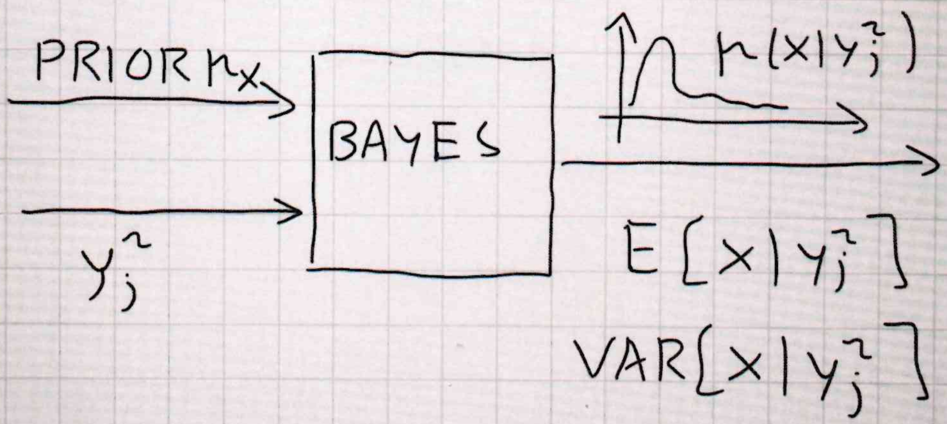
j-TH $\quad$ " $\quad y_j^2$

EACH EXPERIMENT LEADS TO A DIFFERENT POSTERIOR

1



j



WE CAN ALSO THINK OF THE POSTERIOR DENSITY AS RANDOM

$$p(x|\underbrace{y}_{\text{RANDOM}})$$

SO, ANY THING COMPUTE BY $p(x|y)$ IS RANDOM, E.G.

$E[x|y]$ BAYES ESTIMATOR

$VAR[x|y]$ POSTERIOR VARIANCE

REGARDING $VAR[x|y]$

x SCALAR FOR SIMPLICITY

$$VAR[x|y] = E[(x - E[x|y])^2 | y]$$

$$= \int (x - E[x|y])^2 \, p_{x|y}(x) \, dx$$

$= h(y)$ AND BECOMES RANDOM IF WE THINK y AS RANDOM VECTOR

SO, SUCH VARIANCE IN GENERAL DEPENDS ON y (RECALL E.G. STUDENT SCORE)

IN THE GAUSSIAN CASE

$h(y) =$ COSTANT !

$h(y) = $ COSTANT !

NOTE:

IF $VAR(x|y)$ IS RANDOM,
WHAT IS $E[VAR(x|y)]$?

$$E[VAR(x|y)] = E(x - E[x|y])^2$$

AND SO IT IS THE MSE
OF THE BAYES ESTIMATOR.
IT IS COMPUTED BEFORE
SEEING THE DATA (UNCONDITIONAL)
(IN THE GAUSSIAN CASE
IT DOES NOT MAKE ANY
DIFFERENCE TO COMPUTE
IT CONDITIONAL OR UNCOND.)

EXERCISE

PROVE THAT

EXERCISE

PROVE THAT

$$\tilde{x} = x - E[x|y] \perp \hat{x} = E[x|y]$$

$\perp$ = UNCORRELATION

SOL.

$$E\left[E[x|y]\right] = E\hat{x} = Ex \Rightarrow E\tilde{x} = 0$$

$$\Downarrow$$

$$\operatorname{cov}(\tilde{x}, \hat{x}) = E(x - \hat{x})(\hat{x} - Ex)$$

$$= E\left[E\left[(x - \hat{x})\underbrace{(\hat{x} - Ex)}_{\substack{\text{DETERMINISTIC} \\ \text{WHEN } |y}}|y\right]\right]$$

$$= E\left[(\hat{x} - Ex)\underbrace{E[x - \hat{x}\;|y]}_{= E[x|y] - E[x|y] = 0}\right]$$

$$= E[0] = 0$$

EXERCISE

EXERCISE

$x \sim U(0, y)$

$y \sim U(0, 2)$

(X IS NOT UNIFORM SINCE Y IS RANDOM)

COMPUTE

$\hat{x} = E[x|y]$, $MSE_{\hat{x}}$

SOL.

IF $z \sim U(0, b)$ ONE HAS

$E[z] = b/2$, $Ez^2 = b^2/3$

$VAR\ z = b^2/12$

$x|y \sim U(0, y)$

$\downarrow$ DETERMINISTIC

$E[x|y] = \frac{y}{2}$ NOW WE SEE Y AS RANDOM

$VAR(x|y) = \frac{y^2}{12}$ NOW WE SEE Y AS

$$\mathrm{VAR}(x|y) = \frac{y^2}{12}$$ NOW WE SEE Y AS RANDOM

$$\mathrm{MSE} = E\,\mathrm{VAR}(x|y)$$

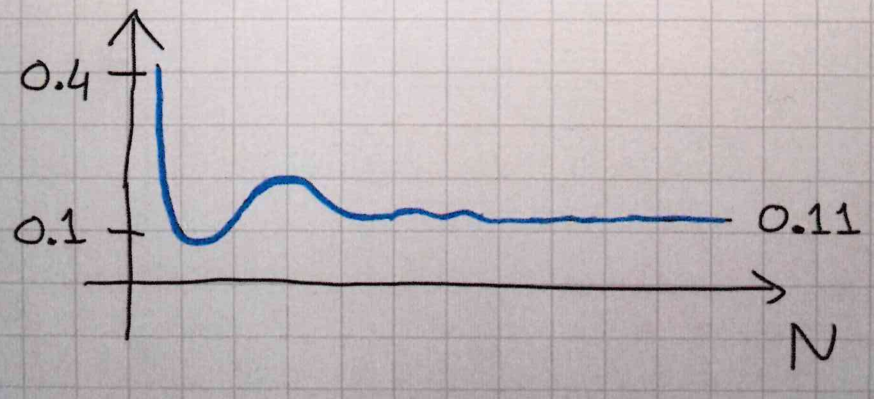
$$= \frac{E\,y^2}{12} = \left(\frac{2^2}{3}\right)/12 = \frac{1}{9}$$

MATLAB TEST USING LAW OF LARGE NUMBERS

$y^i \sim U(0,2)$, $y(i) = 2 \cdot \mathrm{RAND}$

$x^i \sim U(0,y)$, $x(i) = y(i) \cdot \mathrm{RAND}$

$$\mathrm{MSE} \underset{\text{LARGE } N}{\simeq} \sum_{i=1}^{N} \left(x(i) - \frac{y(i)}{2}\right)^2 / N$$



EXAMPLE

GAUSSIAN CASE:

SUMMARY

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{bmatrix} \right)$$

$$E[x|y] = \mu_x + \Sigma_{xy} \Sigma_y^{-1} (y - \mu_y) =: \hat{x}$$

$$\tilde{x} := x - E[x|y]$$

$$VAR[\tilde{x}] = \Sigma_x - \Sigma_{xy} \Sigma_y^{-1} \Sigma_{yx}$$

$$MSE_{\hat{x}} = TRACE(VAR\, \tilde{x})$$

CONSIDER NOW $x | y = y^2$

$p(x | y = y^2)$ COMPUTABLE BY BAYES RULE

$$\Downarrow$$

$x | y = y^2$ IS GAUSSIAN WITH

$$E[x | y = y^2] = \mu_x + \Sigma_{xy} \Sigma_y^{-1} (y^2 - \mu_y)$$

$$VAR[x | y = y^2] = \underbrace{\Sigma_x - \Sigma_{xy} \Sigma_y^{-1} \Sigma_{yx}}_{\text{DOES NOT DEPEND ON } y^2}$$

$$\begin{aligned} VAR(x - E[x|y]) &= VAR(x - E[x|y] \,|\, y = y^2) \\ &= VAR(x | y = y^2) \end{aligned}$$

EXAMPLE

$$y = x + e, \quad x \sim N(0, \lambda)$$
$$e \sim N(0, \sigma^2)$$
$$x \perp e$$

$$\hat{x} = ?$$

$$\hat{x} = \Sigma_{xy} \Sigma_y^{-1} y$$

$$\Sigma_{xy} = \mathrm{cov}(x, y)$$
$$= \mathrm{cov}(x, x+e)$$
$$= \mathrm{cov}(x, x) + \mathrm{cov}(x, e)$$
$$= \mathrm{VAR}(x) + 0$$
$$= \lambda$$

$$\Sigma_y = \mathrm{VAR}\, y = \mathrm{cov}(y, y)$$
$$= \mathrm{cov}(x+e, x+e)$$
$$= \mathrm{cov}(x, x) + \mathrm{cov}(e, e)$$
$$+ \mathrm{cov}(x, e) + \mathrm{cov}(e, x)$$
$$= \mathrm{VAR}\, x + \mathrm{VAR}\, e$$
$$= \lambda + \sigma^2$$

$$\left(\begin{array}{l}\text{SINCE } y = \text{SUM OF TWO INDEP. GAUSSIANS,} \\ \text{THEN } y \text{ IS GAUSSIAN} \\ \Downarrow \\ y \sim N(0, \lambda + \sigma^2)\end{array}\right)$$

$\hat{x} = \dfrac{\lambda}{\lambda+\sigma^2} y$, $\hat{x}$ IS GAUSSIAN

$$\left(\text{VAR } \hat{x} = \frac{\lambda^2}{(\lambda+\sigma^2)^2} \cdot \text{VAR} y = \frac{\lambda^2}{\lambda+\sigma^2} \right)$$

VAR $(\hat{x} - x) = ?$

$$\text{VAR}(\hat{x} - x) = \Sigma_x - \overbrace{\Sigma_{xy} \Sigma_y^{-1} \Sigma_{yx}}^{\text{VAR}\hat{x}}$$

$$\text{VAR}(\hat{x}) = \lambda - \frac{\lambda^2}{\lambda+\sigma^2}$$

$$= \frac{\lambda\sigma^2}{\lambda+\sigma^2}$$

$$\text{||}$$

"LIMITS"

$$\lambda \to +\infty \implies \hat{x} = y \,,\ VAR(\hat{x}) = \sigma^2$$

LIKE SAYING

"$x \sim U(-\infty, +\infty)$"

$$\sigma^2 \to +\infty \implies \hat{x} = 0 = Ex,$$

(NO MEASUREMENT) $\quad VAR\,\hat{x} = \lambda = VAR X$

(POSTERIOR $=$ PRIOR)

HILBERT SPACES

$\mathbb{R}^n$ CONTAINS VECTORS, I.E. ORDERED n-UPLES OF REAL NUMBERS $v = [v_1 \dots v_n]$

$\mathbb{R}^n$ IS A VECTOR SPACE SINCE, WHEN EQUIPPED WITH

$$+ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$$

$\mathbb{R}^n$ IS A VECTOR SPACE SINCE, WHEN EQUIPPED WITH

$$+ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$$

$v + h$ HAS COMPONENTS SUM OF THE COMPONENTS

$$\cdot : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$$

λv IS $[\lambda v_1 \ldots \lambda v_n]$,

SATISFIES THE 8 AXIOMS.

$\mathbb{R}^n$ IS ALSO HILBERT IF EQUIPPED WITH

$$< v, h > = \sum_{i=1}^{n} v_i h_i$$

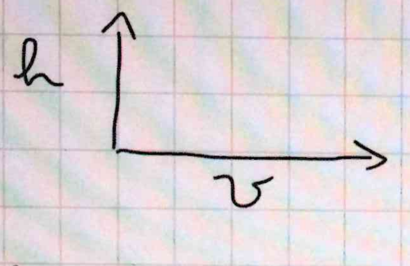
$$\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

$< \cdot , \cdot >$ ABOVE IS AN INNER-PRODUCT

SINCE IT SATISFIES 3 AXIOMS (SIMMETRY, BILINEARITY, POSITIVITY)

USEFULNESS OF INNER-PRODUCTS

a) NOTION OF ORTHOGONALITY



$\iff \langle v, h \rangle = 0$

(EASY TO SEE IN $\mathbb{R}^3$)

b) NOTION OF DISTANCE

$$\| v - h \|^2 = \langle v - h, v - h \rangle$$
$$= \sum_{i=1}^{n} (v_i - h_i)^2$$

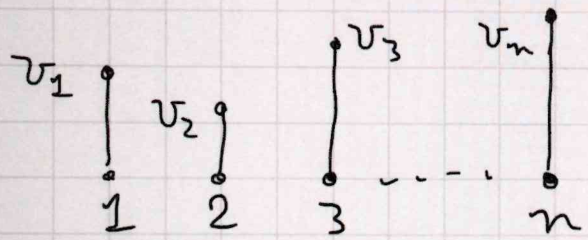
(ALLOWS TO DEFINE LIMITS,
HILBERT IF ANY CAUCY SEQ.
IS A CONVERGENT SEQUENCE)

OTHER HILBERT SPACES BEYOND $\mathbb{R}^n$:

OTHER HILBERT SPACES BEYOND $\mathbb{R}^n$:

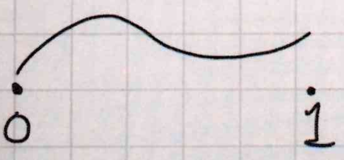
SPACES OF FUNCTIONS

A VECTOR CAN BE SEEN AS A FUNCTION $\{1, \dots, n\} \to \mathbb{R}$



$$\langle v, h \rangle = \sum_{i=1}^{n} v_i h_i$$

IF WE CONSIDER MORE GENERAL DOMAINS, WE OBTAIN GENERAL FUNCTIONS, E.G. $[0,1] \to \mathbb{R}$

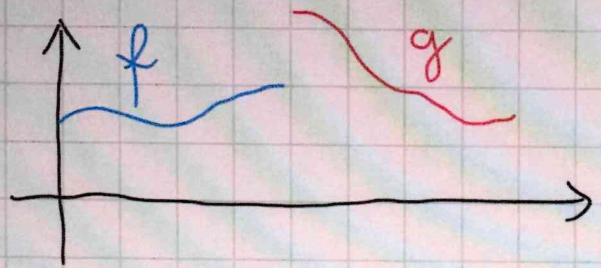


AND WE GENERALIZE THE INNER-PRODUCT AS FOLLOWS

$$\langle f, g \rangle_2 = \int f(x)\, g(x)\, dx$$

NOW I HAVE ALSO THE CONCEPT OF ORTHOGONAL FUNCTIONS AND DISTANCE BETWEEN FUNCTIONS

NOW I HAVE ALSO THE CONCEPT OF ORTHOGONAL FUNCTIONS AND DISTANCE BETWEEN FUNCTIONS

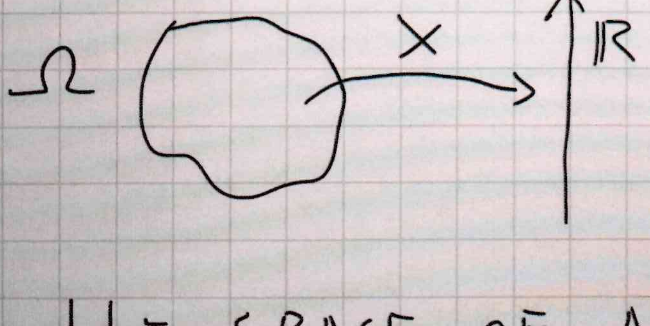


$$\langle f, g \rangle = 0$$

$$f \perp g$$

$$\| f - g \|_2^2 = \int (f(x) - g(x))^2 \, dx$$

ALSO THE RANDOM VARIABLES ARE FUNCTION EVEN IF ON AN ABSTRACT DOMAIN Ω AND A "STRANGE" INTEGRAL ON Ω



H = SPACE OF ALL RANDOM VARIABLES WITH

- ZERO MEAN
- FINITE VARIANCE

H IS HILBERT WITH

$$\langle x, y \rangle_H = E\,xy$$
$$= \int_\Omega X(\omega)\, Y(\omega)\, P(d\omega)$$
$$= \int_{\mathbb{R}^2} xy\, \mu_{xy}(x,y)\, dx\, dy$$

$$\|x\|_H^2 = E\,x^2 = \mathrm{VAR}\, x$$

$\Big(\|x\|_H = 0 \Rightarrow X(\omega) = 0$ FOR ALMOST ALL ω W.R.T. $P\Big)$

ADVANTAGES RELATED TO THIS VIEW

1) CONCEPTS OF ORTHOGONALITY AND DISTANCE BETWEEN R.V.

$$X \perp Y \iff \langle x, y \rangle_H = 0 \quad (= E\,xy)$$

2) IN MANY CASES $E[x|y]$ IS NOT EASILY COMPUTABLE, NOT LINEAR IN y AND AVAILABLE IN CLOSED FORM AS IN THE GAUSSIAN CASE

$$\hat{x} = \Sigma_{xy} \Sigma_y^{-1} y$$

THE GEOMETRICAL VIEW WILL ALLOW US TO FIND AN ALTERNATIVE SUBOPTIMAL ESTIMATOR LINEAR IN y

MINIMUM VARIANCE LINEAR ESTIMATORS

$Ex = Ey = 0$ FOR SIMPLICITY (BUT THEN WE GENERALIZE)

$E[x|y]$ CAN BE COMPLEX AND NOT LINEAR IN y

WE LOOK FOR THE BEST SUBOPTIMAL SOLUTION LINEAR IN y

PROBLEM

$$\hat{g} = \underset{\text{LINEAR } g}{\text{ARGMIN}} \; \|x - g(y)\|_H^2$$

EQUIVALENT TO SOLVING

$$\hat{A} = \underset{A}{\text{ARGMIN}} \; \|x - Ay\|_H^2$$

(Ay = SOMETHING IN THE SUBSPACE GENERATED BY THE $\{y_i\}$)

(Ay = SOMETHING IN THE SUBSPACE GENERATED BY THE $\{y_i\}$)

OBSERVATIONS:

1) IF WE LOOK FOR $Ay+b$ (AFFINE) WE WOULD FIND $b=0$ IF $Ex=Ey=0$

2) IF $Ex \neq 0$ AND $Ey \neq 0$, WE CAN JUST CONSIDER $\bar{x} = x - \mu_x$, $\bar{y} = y - \mu_y$ AND THEN IN THE FINAL RESULT WE REPLACE $\bar{x}, \bar{y}$ WITH $x - \mu_x$, $y - \mu_y$

3) IN THE GAUSSIAN CASE WE ALREADY KNOW

$$\hat{g} = \underset{g}{\text{ARG MIN}} \ldots = \underset{\text{LINEAR } g}{\text{ARG MIN}} \ldots$$

AND $\hat{g}(y) = \Sigma_{xy} \Sigma_y^{-1} y$

$$\Downarrow$$

$$\hat{A} = \Sigma_{xy} \Sigma_y^{-1}$$

SOLUTION USING HILBERT SPACES

DEFINITION:

IF $y = [y_1 \dots y_m]$,

$$H(y) = \text{SPAN}\{y_1, \dots, y_m\}$$
$$= \left\{ \sum_{i=1}^{m} a_i y_i \, , \, a_i \in \mathbb{R} \right\}$$

= SUBSPACE GENERATED BY $\{y_i\}$

NOTE 1:

$\{y_i\}_{i=1}^{m}$ INDEPENDENT IF $\left(\sum_{i=1}^{m} a_i y_i = 0 \Longleftrightarrow \begin{matrix} a_i = 0 \\ \forall i \end{matrix} \right)$

$$\Updownarrow$$

$$\Sigma_y = \text{VAR}\, y > 0$$

NOTE 2: $H(y)$ IS ALWAYS CLOSED SINCE ITS DIMENSION IS FINITE (ANY CAUCHY SEQUENCE IN

(ANY CAUCHY SEQUENCE IN

$H(y)$ IS ALSO CONVERGENT IN $H(y)$,

$\forall \varepsilon \ \exists\ n$ S.T. $\|z_i - z_j\|_H \leq \varepsilon \quad \forall i, j \geq m$

WITH $z_i \in H(y)\ \forall i$, THEN

$z_i \longrightarrow z \in H(y)$).

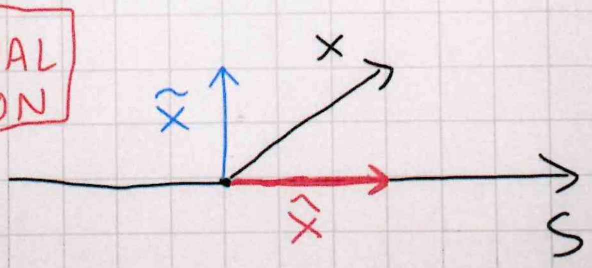
SO, NO HOLES IN $H(y)$

PROJECTION THEOREM

H = HILBERT SPACE, S = CLOSED SUBSPACE

$x \in H$

$\hat{x}$ = ORTHOGONAL PROJECTION OF x ONTO S



- $\forall x \in H$, ONE CAN WRITE

  $x = \hat{x} + \tilde{x}, \quad \hat{x} \in S, \ \tilde{x} \in S^{\perp}$ $\quad \boxed{\begin{array}{c}\forall v \in S \\ \tilde{x} \perp v\end{array}}$

- THE DECOMPOSITION IS UNIQUE

- $\hat{x} = \underset{v \in S}{\mathrm{ARGMIN}} \|x - v\|_H$

NOTE 1:

AFTER PROVING THAT THE DECOMP. $\exists$, UNICITY IS OBTAINED BY CONTRADICTION:

$$x = \underset{\in S}{\hat{x}'} + \underset{\in S^\perp}{\tilde{x}'} = \underset{\in S}{\hat{x}} + \underset{\in S^\perp}{\tilde{x}} \quad , \quad \hat{x} \neq \hat{x}'$$

$$\Downarrow$$

$$\underset{\in S}{\hat{x}' - \hat{x}} = \underset{\in S^\perp}{\tilde{x} - \tilde{x}'} =: v$$

SO, $v \in S$ AND $v \in S^\perp$

$$\Rightarrow \langle v, v \rangle = 0 = \|v\|^2$$

$\Rightarrow v = 0$ BY AXIOMS ON INNER PRODUCT

NOTE 2:

IN THE THEOREM $\dim(S)$ CAN BE ∞

NOW $H =$ HILBERT SPACE OF RANDOM VARIABLES

IF $S = H(y)$ THEN

$$\hat{x} = \Sigma_{xy} \Sigma_y^{-1} y$$

PROOF:

PROOF:

TO PROVE THAT $\hat{x}$ IS THE PROJECTION,

ONE CAN JUST PROVE THAT

$$x - \hat{x} \perp z \, , \quad \forall z \in H(y)$$

$$\Updownarrow$$

$$x - \hat{x} \perp y_i \quad \forall i = 1, \dots, m$$

WE HAVE ALREADY SEEN THIS
TREATING GAUSSIAN ESTIMATION,
BUT LET US REDO THE CALCULATION:

$$x - \hat{x} = x - \Sigma_{xy} \Sigma_y^{-1} y$$

$$\begin{aligned} \mathrm{COV}\left(x - \Sigma_{xy} \Sigma_y^{-1} y, y\right) &= E\, x y^T \\ &\quad - E\, \Sigma_{xy} \Sigma_y^{-1} y y^T \\ &= \Sigma_{xy} - \Sigma_{xy} \Sigma_y^{-1} E\, y y^T \\ &= \Sigma_{xy} - \Sigma_{xy} \Sigma_y^{-1} \Sigma_y = 0 \end{aligned}$$

(NULL MATRIX)

$\blacksquare$

0, FROM THE PROJECTION TH.

Note

SO, FROM THE PROJECTION TH.,

$$\hat{x} = \underset{v \in H(y)}{\mathrm{ARGMIN}} \|v - x\|_H^2 = E(v-x)^2$$

↓ SCALAR x

SO $\hat{x}$ IS THE LINEAR MINIMUM VARIANCE ESTIMATOR!

NOTE 3: x = RANDOM VECTOR,

$\hat{x} = \Sigma_{xy} \Sigma_y^{-1} y$ HAS THUS THE PROPERTY

$$\tilde{x}_i = (\hat{x}_i - x_i) \perp H(y) \quad \forall i,$$

I.E. EACH COMPONENT OF THE ERROR $\tilde{x}$ IS UNCORRELATED FROM ANY COMPONENT OF THE DATA VECTOR y AND FROM ANY LINEAR COMBINATION OF THE y_i

THE LINEAR MODEL

$$y = Sx + w, \quad x \perp w$$

THE LINEAR MODEL

$y = Sx + w, \quad x \perp w$

$\Sigma_x = VAR\,x = P$

$\Sigma_w = VAR\,w = R$

$Ex = Ew = 0$

LET US COMPUTE $\hat{E}[x|y]$ WHERE

$\hat{E}$ = PROJECTION (MINIMUM VARIANCE LINEAR ESTIMATOR)

$$\Sigma_{xy} = cov(x,y) = cov(x, Sx+w)$$
$$= cov(x, Sx) = cov(x,x)S^T = PS^T$$

$$\Sigma_y = VAR\,y = cov(y,y)$$
$$= cov(Sx+w, Sx+w)$$
$$= VAR(Sx) + VAR(w)$$
$$= SPS^T + R$$

HENCE

$$\hat{E}[x|y] = \Sigma_{xy}\Sigma_y^{-1}y$$
$$= PS^T(SPS^T+R)^{-1}y$$

$VAR(\tilde{x}) =$

$$\hat{E}[x|y] = \Sigma_{xy} \Sigma_y^{-1} y$$
$$= PS^T (SPS^T + R)^{-1} y$$

$$VAR(\tilde{x}) = \Sigma_x - \Sigma_{xy} \Sigma_y^{-1} \Sigma_{yx}$$
$$= P - PS^T (SPS^T + R)^{-1} SP$$

ALTERNATIVE FORMULA

A, C SQUARE AND INVERTIBLE

THEN INVERSION LEMMA SAYS

$$(A + BCD)^{-1} = A^{-1} - A^{-1} B (C^{-1} + DA^{-1} B)^{-1} DA^{-1}$$

THEN IN OUR CONTEXT:

$$\hat{E}[x|y] = (P^{-1} + S^T R^{-1} S)^{-1} S^T R^{-1} y$$

$$VAR\ \tilde{x} = (P^{-1} + S^T R^{-1} S)^{-1}$$

(LITTLE INFORMATIVE PRIOR MEANS $P^{-1} \to 0 \ldots$)

SUMMARY

$\hat{x} = \hat{E}[x|y] = \hat{E}[x \mid H(y)]$ SOLVES

MIN $E\|x - v\|^2$ SCALAR

SUMMARY

$\hat{x} = \hat{E}[x|y] = \hat{E}[x | H(y)]$ SOLVES

$$\min_{v \in H(y)} E \|x - v\|^2$$ SCALAR CASE

HENCE IT SOLVES

$$\min_{A} \|x - Ay\|_H^2 \Rightarrow \hat{A} = \Sigma_{xy} \Sigma_y^{-1}$$

(VALID ALSO FOR THE VECTOR CASE!)

WE HAVE ALSO UNDERSTOOD THAT

THE BEST LINEAR ESTIMATOR OF x BASED ON y ONLY DEPENDS ON THE 1^{ST} AND 2^{ND} MOMENTS OF x, y AND ONE HAS

$$\hat{E}[x|y] = \mu_x + \Sigma_{xy} \Sigma_y^{-1} (y - \mu_y)$$

$$\tilde{x} = x - \hat{x}$$

$$\mathrm{VAR}\, \tilde{x} = \Sigma_x - \Sigma_{xy} \Sigma_y^{-1} \Sigma_{yx}$$

IN THE GAUSSIAN CASE (x, y JOINTLY GAUSSIAN)

$$\hat{E}[x|y] = E[x|y]$$

IN THE GAUSSIAN CASE (x, y JOINTLY GAUSSIAN)

- $\hat{E}[x|y] = E[x|y]$
- $x|y$ IS GAUSSIAN WITH

MEAN $E[x|y]$, COVARIANCE $\mathrm{VAR}\,\tilde{x}$

$$\mathrm{VAR}\,\tilde{x} \equiv \mathrm{VAR}[x|y]$$

(IN FACT, $x - E[x|y]$ IS INDEPENDENT OF y, SO THAT $\mathrm{VAR}\,\tilde{x}$

$$= \mathrm{VAR}[x - E[x|y]]$$

$$= \mathrm{VAR}[x - E[x|y] \,|\, y]$$

$$= \mathrm{VAR}[x|y - E[x|y]]$$

$$= \mathrm{VAR}[x|y])$$

EXERCISE

PROVE THAT, IF $Ex = Ey = 0$, THEN

$$\hat{g} = \underset{\substack{g(y) \text{ S.T.} \\ g(y) = Ay + b}}{\mathrm{ARGMIN}} \; E \underbrace{\| g(y) - x \|^2}_{\substack{\text{EUCLIDEAN} \\ \text{CASE, VECTOR} \\ \text{CASE!}}}$$

EXERCISE

PROVE THAT, IF $Ex = Ey = 0$, THEN

$$\hat{g} = \underset{\substack{g(y) \text{ S.T.} \\ g(y) = Ay + b}}{\text{ARGMIN}} \; E \underbrace{\| g(y) - x \|^2}_{\substack{\text{EUCLIDEAN} \\ \text{CASE, VECTOR} \\ \text{CASE!}}}$$

$$= \underset{\substack{g(y) \text{ S.T.} \\ g(y) = Ay}}{\text{ARG MIN}} \; E \| g(y) - x \|^2$$

WE HAVE TO PROVE THAT $b = 0$.

$$\| x - Ay - b \|^2 = \| x - Ay \|^2 + \| b \|^2 - 2 \langle x - Ay, b \rangle$$

WE TAKE THE MEAN:

$$\left(E \| x - Ay \|^2 \right) + \| b \|^2 - 2 \langle \underbrace{E(x - Ay)}_{=0}, b \rangle$$

$\Rightarrow$ b MUST BE 0 TO MINIMIZE THE OBJECTIVE

EXERCISE

EXERCISE

n MEASUREMENTS OF A RESISTOR OF NOMINAL VALUE x_0 AND TOLERANCE α

PRIOR: WITH HIGH PROBABILITY

$$x_0 - \alpha x_0 \leq x \leq \alpha x_0 + x_0$$

$\Rightarrow$ USING A GAUSSIAN PRIOR REASONABLE TO SAY

$$x \sim N(x_0, \sigma_x^2)$$

WITH $2\sigma_x = \alpha x_0 \Rightarrow \sigma_x = \dfrac{\alpha x_0}{2}$

THE MODEL IS

$y_i = x + w_i, \quad i = 1, \ldots, n$ AND

W.H.P. WE KNOW

$$-\beta_i x_M \leq w_i \leq \beta_i x_M$$

SO IT IS REASONABLE TO SAY

$$w_i \sim N(0, \sigma_i^2)$$

WITH $2\sigma_i = \beta_i x_M \Rightarrow \sigma_i = \dfrac{\beta_i x_M}{2}$

WITH $2\sigma_i = \beta_i x_M \Rightarrow \sigma_i = \dfrac{\beta_i x_M}{2}$

$$\Downarrow$$

$$y = Sx + w, \quad x \perp w$$

$$S = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad P = \sigma_x^2$$

$$R = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{bmatrix}$$

SOME CALCULATIONS

$$S^T R^{-1} S = \sum_{i=1}^{n} \frac{1}{\sigma_i^2}$$

$$S^T R^{-1}(y - \mu_y) = S^T R^{-1}(y - \mu_x)$$

$$= S^T R^{-1}\left(y - \begin{bmatrix} x_0 \\ \vdots \\ x_0 \end{bmatrix}\right)$$

$$= \sum_{i=1}^{n} \frac{y_i - x_0}{\sigma_i^2}$$

$$= \sum_{i=1}^{n} \frac{y_i - x_0}{\sigma_i^2}$$

$$E[x|y] = x_0 + \Big(\underbrace{\frac{1}{\sigma_x^2}}_{P^{-1}} + \underbrace{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}_{S^T R^{-1} S} \Big)^{-1} \underbrace{\sum_{i=1}^{n} \frac{y_i - x_0}{\sigma_i^2}}_{S^T R^{-1}(y-\mu_y)}$$

EXERCISE

$$\{y_t\}_{t=1}^{3}, \quad y_t = \sin(\omega t), \quad \omega \sim U(-\pi, \pi)$$

I WANT TO ESTIMATE y_3 USING y_1, y_2

$$E y_t = \int_{-\pi}^{\pi} \sin(\omega t) \cdot \underbrace{\frac{1}{2\pi}}_{\text{PRIOR ON } \omega} d\omega = 0$$

$$E y_t y_s = \int_{-\pi}^{\pi} \sin(\omega t) \sin(\omega s) \frac{1}{2\pi} d\omega, \quad t \neq s$$

$$= 0 \quad (\text{ORTHOGONALITY OF FOURIER BASIS})$$

$$\hat{E}[y_3 | y_1, y_2] =$$

$$\big[\operatorname{COV}(y_3, y_1) \ \operatorname{COV}(y_3, y_2)\big] \big(\operatorname{VAR}(y_1, y_2)\big)^{-1} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$\hat{E}[y_3 \mid y_1, y_2] =$$

$$\underbrace{\left[\mathrm{COV}(y_3, y_1) \;\; \mathrm{COV}(y_3, y_2)\right]}_{=[0 \;\; 0]} \left(\mathrm{VAR}(y_1, y_2)\right)^{-1} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$= 0$$

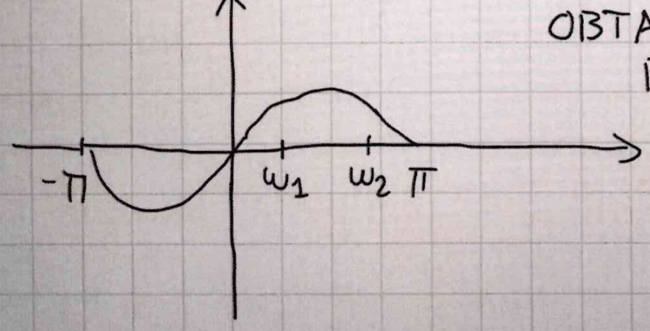
THE LINEAR ESTIMATOR DOES NOT CHANGE THE PRIOR:

$$\hat{E}[y_3 \mid y_1, y_2] = \hat{E}\, y_3 = 0$$

BUT WE CAN PERFECTLY RECONSTRUCT ω FROM y_1, y_2!

IN FACT:

$$y_1 = \sin \omega$$



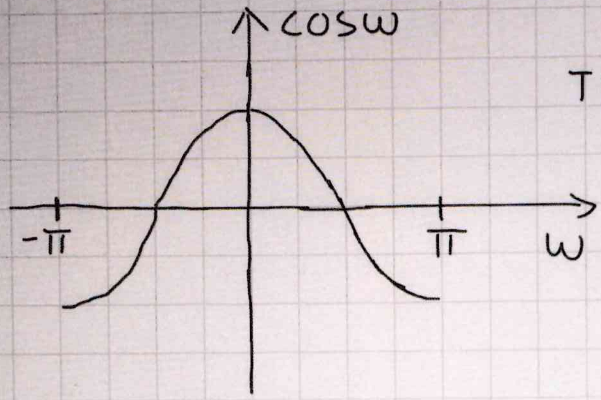
OBTAIN TWO POSSIBLE VALUES FOR ω, ω_1, ω_2

$$y_2 = \sin 2\omega = 2 \sin \omega \cos \omega$$

$$\Longrightarrow \cos \omega = y_2$$

$$y_2 = \sin 2\omega = 2 \sin\omega \cos\omega$$

$$\Rightarrow \cos\omega = \frac{y_2}{2y_1}$$



TWO POSSIBLE VALUES, ONE POSITIVE, ONE NEGATIVE, BUT I KNOW FROM y_1 THE SIGN!

UNIQUE VALUE OF ω OBTAINED,

PERFECT ESTIMATE

EXERCISE

RADAR / SONAR SIGNAL

$$y(t) = A \sin(\omega_0 t + \Theta) + w(t)$$

- A = KNOWN AMPLITUDE, ω_0 KNOWN
- Θ = ECO'S DELAY

  $\Theta \sim U(0, 2\pi)$
- $w(t)$ WHITE NOISE $\perp \Theta$,

  $w(t) \sim (0, \sigma^2)$

ESTIMATE Θ FROM $\{y_t\}_{t=1}^{N}$

ESTIMATE θ FROM $\{y_t\}_{t=1}^{N}$

SOL. LET US SEE AS A LINEAR ESTIMATION PROBLEM

$$x = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$$

$$\sin(\omega_0 t + \theta) = (\sin \omega_0 t) \cos\theta + (\cos \omega_0 t) \sin\theta$$

$$\Downarrow$$

$$y = Sx + w$$

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}, \quad S = A \begin{bmatrix} \sin \omega_0 & \cos \omega_0 \\ \vdots & \vdots \\ \sin \omega_0 N & \cos \omega_0 N \end{bmatrix}$$

$$S \in \mathbb{R}^{N \times 2},$$

$$w = \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix}$$

$$E\,x = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{(BY SIMMETRY OF THE INTEGRALS)}$$

$$E \cos^2\theta = \int_0^{2\pi} \frac{\cos^2\theta}{2\pi}\, d\theta$$

$$= \left. \frac{\theta + \sin\theta \cos\theta}{2 \cdot 2\pi} \right|_0^{2\pi} = \frac{1}{2}$$

$$E \cos\theta \sin\theta = 0$$

$$E \sin^2\theta = 1/2$$

$$\Downarrow$$

$$\mathrm{VAR}\, x = P = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}$$

$$\mathrm{VAR}\, w = R = \sigma^2 I_N$$

BEST LINEAR ESTIMATOR

$$S^T R^{-1} S \quad S^T S$$

BEST LINEAR ESTIMATOR

$$S^T R^{-1} S = \frac{S^T S}{\sigma^2}$$

$$= \frac{A^2}{\sigma^2} \begin{bmatrix} \sum_{t=1}^{N} \sin^2 \omega_0 t & \sum_{t=1}^{N} \sin \omega_0 t \cos \omega_0 t \\ \sum_{t=1}^{N} \sin \omega_0 t \cos \omega_0 t & \sum_{t=1}^{N} \cos^2 \omega_0 t \end{bmatrix}$$

$$S^T R^{-1} y \underset{(Ey = Ex = 0)}{=} \frac{A}{\sigma^2} \begin{bmatrix} \sum_{t=1}^{N} y(t) \sin \omega_0 t \\ \sum_{t=1}^{N} y(t) \cos \omega_0 t \end{bmatrix},$$

$$P^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\Downarrow$$

$$\hat{E}[x|y] = \left(P^{-1} + S^T R^{-1} S\right)^{-1} S^T R^{-1} y$$

HOMEWORK: WHAT HAPPENS FOR LARGE N?

FROM STATIC TO DYNAMIC ESTIMATE

$\{y_t\}_{t=t_0, t_0+1, \dots}$ COLLECTED OVER TIME, FLOW OF OBSERVATIONS...

$$y^t := \begin{bmatrix} y_1 \\ \vdots \\ y_t \end{bmatrix}$$

AND WE WANT TO ESTIMATE x

$$Ex = Ey^t = 0$$

$$\hat{x}_t = \Sigma_{xy^t} \Sigma_{y^t}^{-1} y^t$$

ESTIMATE OF x AT INSTANT t

WE APPLY THE STATIC FORMULA

BUT AS NEW DATA ARRIVE (t INCREASES) WE HAVE TO RECOMPUTE $\hat{x}_t$

- $\Sigma_{y^t}^{-1}$ REQUIRES $O(t^3)$ OPERATIONS AND $t \to +\infty$

AND $t \to +\infty$

- y^t HAS TO BE STORED IN MEMORY

IT WOULD BE INSTEAD FUNDAMENTAL TO COMPUTE $\hat{x}_{t+1}$ IN A RECURSIVE WAY

$$\hat{x}_{t+1} = f(\underbrace{\hat{x}_t}_{\text{PREVIOUS ESTIMATE}}, \underbrace{y_{t+1}}_{\text{ONLY THE LAST MEASUREMENT}})$$

BUT THERE IS AN ADDITIONAL KEY PROBLEM:

x ITSELF COULD BE DYNAMIC

$x_t = x$ VARIES OVER TIME

$\hat{x}_t$ = ESTIMATE OF x_t BASED ON y^t

ROBOT'S POSITION AT $t+1$

TO COMPUTE x_{t+1} IN A

RECURSIVE WAY

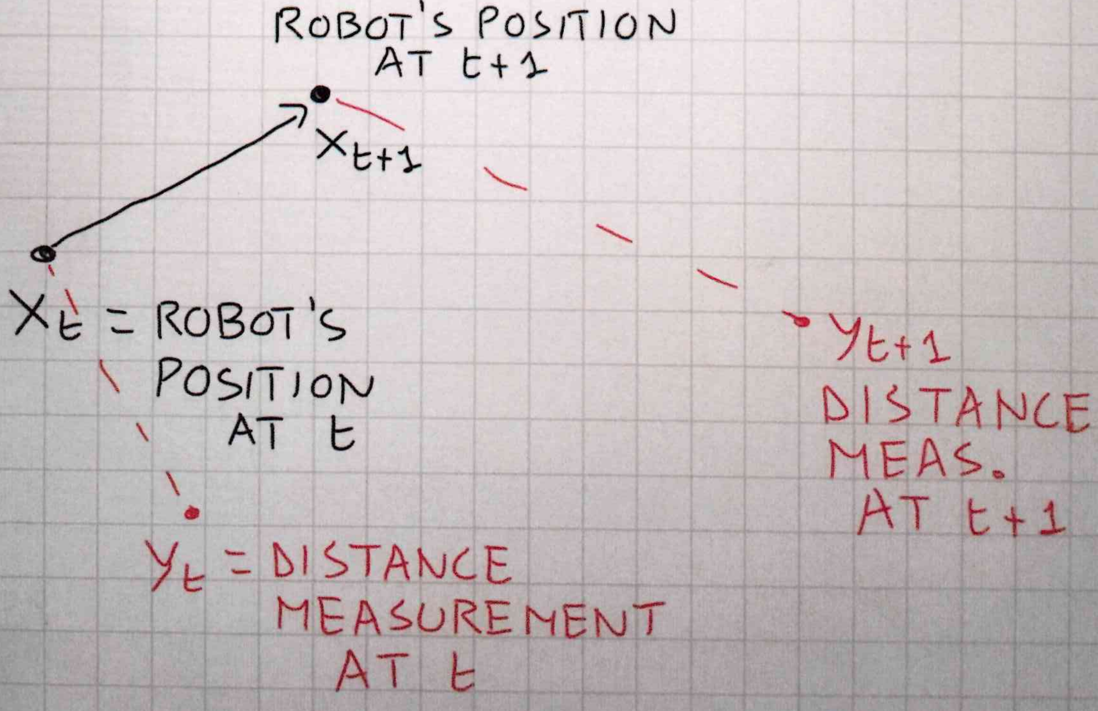
$$\hat{x}_{t+1} = f(\underbrace{\hat{x}_t}_{\text{PREVIOUS ESTIMATE}}, \underbrace{y_{t+1}}_{\text{ONLY THE LAST MEASUREMENT}})$$

BUT THERE IS AN ADDITIONAL KEY PROBLEM:

x ITSELF COULD BE DYNAMIC

x_t = x VARIES OVER TIME

$\hat{x}_t$ = ESTIMATE OF x_t BASED ON y^t



FILTERING AND PREDICTION

SCALAR CASE

$\{x_t\}$, $\{y_t\}$ ARE STOCHASTIC PROCESSES

DATA ARE $\{y(t), t \in I\}$

FILTERING

$I = [t_0, t]$ AND I WANT

$$E\left[x_t \mid \{y_t\}_{t \in I}\right] =: \hat{x}(t|t)$$

CAUSAL ESTIMATOR

PREDICTION

$I = [t_0, t]$ AND I WANT

$$E\left[x_{t+h} \mid \{y_t\}_{t \in I}\right] =: \hat{x}(t+h|t), \quad h>0$$

IT IS A PREDICTOR, STILL A CAUSAL ESTIMATOR

- SOMETIMES, AS TARGET ALSO

SOMETIMES, AS TARGET ALSO

$$E\left[y_{t+h} \mid \{y_t\}_{t\in I}\right] =: \hat{y}(t+h|t)$$

OUTPUT PREDICTOR

$\hat{E}$ IN PLACE OF E WILL BE THE TARGET FOR THE MOMENT

WIENER APPROACH

$\{x_t\}, \{y(t)\}$ JOINTLY STATIONARY (IN A WEAK SENSE)

AND $t_0 \to -\infty$, SO THAT

$$I = [-\infty, t]$$

WIENER SEES THE LINEAR ESTIMATOR / FILTER

$$\hat{x}(t|t) = \sum_{k=t_0}^{t} H(t,k)\, y(k)$$

AS A DYNAMIC SYSTEM

$$y_t \longrightarrow \boxed{H(t,\cdot)} \xrightarrow{\hat{x}(t|t)}$$

UNDER WIENER ASSUMPTIONS $H(t,K)$ BECOMES $H(t-K)$ (TIME-INVARIANT SYSTEM).

WE COMPUTE H JUST ONCE AND WE USE IT $\forall t$.

IT CAN BE OBTAINED BY SOLVING THE WIENER-HOPF EQUATION

COMPUTATION IS EASY ONLY IF THE SPECTRA OF THE PROCESSES ARE RATIONAL (SUCH PROCESSES ARE THE OUTPUTS OF FINITE-DIMENSIONAL TIME-INVARIANT LINEAR SYSTEMS WITH WHITE NOISE AS INPUT STARTING AT $t_0 = -\infty$)

SCALAR EXAMPLE

x, y SCALAR

$$\Sigma_x(\tau) = E\, x_t x_{t+\tau}$$

$$\Big\downarrow \text{Z-TRANSFORM} \left(\sum_k \Sigma_x(k) z^{-k} \right)$$

$$S_x(z) = \frac{\text{POLYNOMIAL}}{\text{POLYNOMIAL}}$$

SIMILAR DEFINITIONS THEN FOR

$S_{xy}(z)$, $S_y(z)$, $S_{yx}(z)$

IN THE MATRIX CASE

$$S_{xy}(z) = S_{yx}^T(z^*) = (S_{xy}(z))^*$$

$$z = \sigma + j\omega$$

$$z^* = \sigma - j\omega$$

FROM WIENER TO KALMAN

IMPORTANCE OF WIENER: TO SEE THE SIGNALS OF INTEREST AS STOCHASTIC

WIENER PREDICTOR

LET $s(t)$ BE A STATIONARY STOCHASTIC PROCESS OVER $t \in \mathbb{Z}$ WITH

$$R_s(\tau) = E\, s(t)\, s(t-\tau)$$

ITS POWER SPECTRUM IS

$$\Phi_s(\omega) = \sum_{\tau=-\infty}^{+\infty} R_s(\tau)\, e^{-i\tau\omega}, \quad \omega \in \mathbb{R}$$

IF $s(t)$ IS SCALAR (AS WE HEREBY ASSUME),

THEN:

1) $\Phi_s(\omega) \in \mathbb{R} \quad \forall \omega$

2) $\Phi_s(\omega) \geq 0 \quad \forall \omega$

3) $\Phi_s(\omega) = \Phi_s(-\omega)$

OFTEN, WE REASON IN TERMS OF

Z-TRANSFORM AND

$$\Phi_s(z) = \sum_{\tau=-\infty}^{+\infty} R_s(\tau)\, z^{-\tau}, \quad z \in \mathbb{C}$$

SO THAT THE POWER SPECTRUM IS GIVEN BY ITS EVALUATION ON THE UNIT CIRCLE

SO THAT THE POWER SPECTRUM IS GIVEN BY ITS EVALUATION ON THE UNIT CIRCLE

ALSO, WE OFTEN CONSIDER

$$y(t) = G(z)u(t) + H(z)e(t) \quad \circledast$$

TO INDICATE THAT y IS THE SUM OF THE OUTPUTS OF TWO LINEAR SYSTEMS WITH TRANSFER FUNCTIONS $G(z)$, $H(z)$ FED RESPECTIVELY WITH INPUTS $u(t)$, $e(t)$

THEOREM: LET $y(t)$ BE GIVEN BY $\circledast$ WITH u A STATIONARY STOCHASTIC PROCESS WITH SPECTRUM $\Phi_u(\omega)$ WHILE e IS WHITE NOISE OF VARIANCE λ. THEN, IF G AND H ARE STABLE, y IS STATIONARY WITH

$$\Phi_y(\omega) = |G(e^{j\omega})|^2 \Phi_u(\omega) + \lambda |H(e^{j\omega})|^2$$

SPECTRAL FACTORIZATION

LET $\Phi(\omega)$ BE RATIONAL IN $e^{j\omega}$ WITH $\Phi(\omega) > 0 \ \forall \omega$.

THEN, THERE EXISTS A MONIC RATIONAL TRANSFER FUNCTION OF z, $R(z)$, WITH POLES AND ZEROS ALL INSIDE THE UNIT CIRCLE S.T.

$$\Phi(\omega) = \lambda \left| R(e^{j\omega}) \right|^2 \quad \blacksquare$$

THIS MEANS THAT A STATIONARY DISTURBANCE v CAN BE WRITTEN AS

$$v(t) = R(z) e(t),$$

WITH $R(z)$ BEING MINIMUM PHASE AND e WHITE NOISE OF VARIANCE λ,

WITH $R(z)$ BEING MINIMUM PHASE AND e WHITE NOISE OF VARIANCE λ, FROM INFORMATION ABOUT ITS SPECTRUM ONLY.

SUMMARY

$$y(t) = G(z)\,u(t) + H(z)\,e(t)$$

IS THE BASIC DESCRIPTION OF A LINEAR SYSTEM S.T. DISTURBANCES WITH e WHITE NOISE.

G AND H ARE RATIONAL TRANSFER FUNCTIONS:

$$G(z) = \sum_{k=1}^{\infty} g(k)\, z^{-k}$$

$$H(z) = 1 + \sum_{k=1}^{\infty} h(k)\, z^{-k}, \quad \text{AND}$$

G STABLE, I.E. $\sum_{k=1}^{\infty} |g(k)| < \infty$,

H STABLE AND MINIMUM PHASE

IMPORTANCE OF MINIMUM-PHASE NOISE REPRESENTATION:

NOISE MODEL INVERTIBILITY

LET

$$v(t) = H(z)e(t) \quad ⊛$$

$$= \sum_{k=0}^{\infty} h(k)\, e(t-k), \quad h(0)=1$$

A CRUCIAL PROPERTY FOR PREDICTION IS THAT ⊛ SHOULD BE INVERTIBLE, I.E. WE SHOULD BE ABLE TO COMPUTE $e(t)$ FROM $\{v(s)\}_{s \le t}$

(MORE SPECIFICALLY, CAUSALLY INVERTIBLE).

IF $H(z)$ IS STABLE AND INVERSELY STABLE (MINIMUM PHASE), WE JUST HAVE

JUST HAVE

$$e(t) = \tilde{H}(z)\,v(t)$$
$$= \sum_{k=0}^{\infty} \tilde{h}(k)\,v(t-k)$$

WITH

$$\tilde{H}(z) = 1/_{H(z)} =: H^{-1}(z)$$

SO, $H(z)$ MUST HAVE NO ZEROS ON OR OUTSIDE THE UNIT CIRCLE.

THIS WELL RELATES TO THE SPECTRAL FACTORIZATION RESULT WHICH SAYS THAT, FOR RATIONAL STRICTLY POSITIVE SPECTRA, WE CAN ALWAYS FIND A REPRESENTATION $H(z)$ SATISFYING THIS.

ONE-STEP AND K-STEP AHEAD OUTPUT PREDICTION

IF

$$y(t) = G(z)\,u(t) + H(z)\,e(t)$$

G STABLE, H MINIMUM PHASE

Note

ONE-STEP AND K-STEP AHEAD OUTPUT PREDICTION

IF

$$y(t) = G(z)\,u(t) + H(z)\,e(t)$$

G STABLE, H MINIMUM PHASE

IT TURNS OUT THAT

$$\hat{y}(t|t-1) = H^{-1}(z)\,G(z)\,u(t) + \left[1 - H^{-1}(z)\right] y(t)$$

IN ADDITION, LETTING

$$\bar{H}_K(z) = \sum_{i=0}^{K-1} h(i)\,z^{-i}$$

$$W_K(z) = \bar{H}_K(z)\,H^{-1}(z)$$

ONE HAS

$$\hat{y}(t|t-K) = W_K(z)\,G(z)\,u(t) + \left[1 - W_K(z)\right] y(t)$$

A New Approach to Linear Filtering and Prediction Problems[1]

R. E. KALMAN
Research Institute for Advanced Study,[2]
Baltimore, Md.

The classical filtering and prediction problem is re-examined using the Bode-Shannon representation of random processes and the "state transition" method of analysis of dynamic systems. New results are:

(1) The formulation and methods of solution of the problem apply without modification to stationary and nonstationary statistics and to growing-memory and infinite-memory filters.

(2) A nonlinear difference (or differential) equation is derived for the covariance matrix of the optimal estimation error. From the solution of this equation the coefficients of the difference (or differential) equation of the optimal linear filter are obtained without further calculations.

(3) The filtering problem is shown to be the dual of the noise-free regulator problem.

The new method developed here is applied to two well-known problems, confirming and extending earlier results.

The discussion is largely self-contained and proceeds from first principles; basic concepts of the theory of random processes are reviewed in the Appendix.

Introduction

AN IMPORTANT class of theoretical and practical problems in communication and control is of a statistical nature. Such problems are: (i) Prediction of random signals; (ii) separation of random signals from random noise; (iii) detection of signals of known form (pulses, sinusoids) in the presence of random noise.

In his pioneering work, Wiener [1][3] showed that problems (i) and (ii) lead to the so-called Wiener-Hopf integral equation; he also gave a method (spectral factorization) for the solution of this integral equation in the practically important special case of stationary statistics and rational spectra.

Many extensions and generalizations followed Wiener's basic work. Zadeh and Ragazzini solved the finite-memory case [2]. Concurrently and independently of Bode and Shannon [3], they also gave a simplified method [2] of solution. Booton discussed the nonstationary Wiener-Hopf equation [4]. These results are now in standard texts [5-6]. A somewhat different approach along these main lines has been given recently by Darlington [7]. For extensions to sampled signals, see, e.g., Franklin [8], Lees [9]. Another approach based on the eigenfunctions of the Wiener-Hopf equation (which applies also to nonstationary problems whereas the preceding methods in general don't), has been pioneered by Davis [10] and applied by many others, e.g., Shinbrot [11], Blum [12], Pugachev [13], Solodovnikov [14].

In all these works, the objective is to obtain the specification of a linear dynamic system (Wiener filter) which accomplishes the prediction, separation, or detection of a random signal.[4]

---

[1] This research was supported in part by the U. S. Air Force Office of Scientific Research under Contract AF 49 (638)-382.

[2] 7212 Bellona Ave.

[3] Numbers in brackets designate References at end of paper.

[4] Of course, in general these tasks may be done better by nonlinear filters. At present, however, little or nothing is known about how to obtain (both theoretically and practically) these nonlinear filters.

Contributed by the Instruments and Regulators Division and presented at the Instruments and Regulators Conference, March 29– April 2, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

Manuscript received at ASME Headquarters, February 24, 1959. Paper No. 59—IRD-11.

Present methods for solving the Wiener problem are subject to a number of limitations which seriously curtail their practical usefulness:

(1) The optimal filter is specified by its impulse response. It is not a simple task to synthesize the filter from such data.

(2) Numerical determination of the optimal impulse response is often quite involved and poorly suited to machine computation. The situation gets rapidly worse with increasing complexity of the problem.

(3) Important generalizations (e.g., growing-memory filters, nonstationary prediction) require new derivations, frequently of considerable difficulty to the nonspecialist.

(4) The mathematics of the derivations are not transparent. Fundamental assumptions and their consequences tend to be obscured.

This paper introduces a new look at this whole assemblage of problems, sidestepping the difficulties just mentioned. The following are the highlights of the paper:

(5) *Optimal Estimates and Orthogonal Projections.* The Wiener problem is approached from the point of view of conditional distributions and expectations. In this way, basic facts of the Wiener theory are quickly obtained; the scope of the results and the fundamental assumptions appear clearly. It is seen that all statistical calculations and results are based on first and second order averages; no other statistical data are needed. Thus difficulty (4) is eliminated. This method is well known in probability theory (see pp. 75–78 and 148–155 of Doob [15] and pp. 455–464 of Loève [16]) but has not yet been used extensively in engineering.

(6) *Models for Random Processes.* Following, in particular, Bode and Shannon [3], arbitrary random signals are represented (up to second order average statistical properties) as the output of a linear dynamic system excited by independent or uncorrelated random signals ("white noise"). This is a standard trick in the engineering applications of the Wiener theory [2–7]. The approach taken here differs from the conventional one only in the way in which linear dynamic systems are described. We shall emphasize the concepts of *state* and *state transition;* in other words, linear systems will be specified by systems of first-order difference (or differential) equations. This point of view is

natural and also necessary in order to take advantage of the simplifications mentioned under (5).

(7) *Solution of the Wiener Problem*. With the state-transition method, a single derivation covers a large variety of problems: growing and infinite memory filters, stationary and nonstationary statistics, etc.; difficulty (3) disappears. Having guessed the "state" of the estimation (i.e., filtering or prediction) problem correctly, one is led to a nonlinear difference (or differential) equation for the covariance matrix of the optimal estimation error. This is vaguely analogous to the Wiener-Hopf equation. Solution of the equation for the covariance matrix starts at the time t_0 when the first observation is taken; at each later time t the solution of the equation represents the covariance of the optimal prediction error given observations in the interval (t_0, t). From the covariance matrix at time t we obtain at once, without further calculations, the coefficients (in general, time-varying) characterizing the optimal linear filter.

(8) *The Dual Problem*. The new formulation of the Wiener problem brings it into contact with the growing new theory of control systems based on the "state" point of view [17–24]. It turns out, *surprisingly*, that the Wiener problem is the *dual* of the noise-free optimal regulator problem, which has been solved previously by the author, using the state-transition method to great advantage [18, 23, 24]. The mathematical background of the two problems is identical—this has been suspected all along, but until now the analogies have never been made explicit.

(9) *Applications*. The power of the new method is most apparent in theoretical investigations and in numerical answers to complex practical problems. In the latter case, it is best to resort to machine computation. Examples of this type will be discussed later. To provide some feel for applications, two standard examples from nonstationary prediction are included; in these cases the solution of the nonlinear difference equation mentioned under (7) above can be obtained even in closed form.

For easy reference, the main results are displayed in the form of theorems. Only Theorems 3 and 4 are original. The next section and the Appendix serve mainly to review well-known material in a form suitable for the present purposes.

Notation Conventions

Throughout the paper, we shall deal mainly with *discrete* (or *sampled*) dynamic systems; in other words, signals will be observed at equally spaced points in time (*sampling instants*). By suitable choice of the time scale, the constant intervals between successive sampling instants (*sampling periods*) may be chosen as unity. Thus variables referring to time, such as t, t_0, τ, T will always be integers. The restriction to discrete dynamic systems is not at all essential (at least from the engineering point of view); by using the discreteness, however, we can keep the mathematics rigorous and yet elementary. Vectors will be denoted by small bold-face letters: **a**, **b**, ..., **u**, **x**, **y**, ... A *vector* or more precisely an *n-vector* is a set of n numbers x_1, ... x_n; the x_i are the *co-ordinates* or *components* of the vector **x**.

Matrices will be denoted by capital bold-face letters: **A**, **B**, **Q**, $\mathbf{\Phi}$, $\mathbf{\Psi}$, ...; they are $m \times n$ arrays of elements a_{ij}, b_{ij}, q_{ij}... The *transpose* (interchanging rows and columns) of a matrix will be denoted by the prime. In manipulating formulas, it will be convenient to regard a vector as a matrix with a single column.

Using the conventional definition of matrix multiplication, we write the *scalar product* of two n-vectors **x**, **y** as

$$\mathbf{x}'\mathbf{y} = \sum_{i=1}^{n} x_i y_i = \mathbf{y}'\mathbf{x}$$

The scalar product is clearly a scalar, i.e., not a vector, quantity. Similarly, the quadratic form associated with the $n \times n$ matrix **Q** is,

$$\mathbf{x}'\mathbf{Q}\mathbf{x} = \sum_{i,j=1}^{n} x_i q_{ij} x_j$$

We define the expression $\mathbf{xy}'$ where $\mathbf{x}'$ is an m-vector and **y** is an n-vector to be the $m \times n$ matrix with elements $x_i y_j$.

We write $E(\mathbf{x}) = E\mathbf{x}$ for the expected value of the random vector **x** (see Appendix). It is usually convenient to omit the brackets after E. This does not result in confusion in simple cases since constants and the operator E commute. Thus $E\mathbf{xy}'$ = matrix with elements $E(x_i y_j)$; $E\mathbf{x}E\mathbf{y}'$ = matrix with elements $E(x_i)E(y_j)$.

For ease of reference, a list of the principal symbols used is given below.

Optimal Estimates

t	time in general, present time.
t_0	time at which observations start.
$x_1(t), x_2(t)$	basic random variables.
$y(t)$	observed random variable.
$x_1^*(t_1 \mid t)$	optimal estimate of $x_1(t_1)$ given $y(t_0), \ldots, y(t)$.
L	loss function (non random function of its argument).
ε	estimation error (random variable).

Orthogonal Projections

$\mathcal{Y}(t)$	linear manifold generated by the random variables $y(t_0), \ldots, y(t)$.
$\bar{x}(t_1 \mid t)$	orthogonal projection of $x(t_1)$ on $\mathcal{Y}(t)$.
$\tilde{x}(t_1 \mid t)$	component of $x(t_1)$ orthogonal to $\mathcal{Y}(t)$.

Models for Random Processes

$\mathbf{\Phi}(t+1; t)$	transition matrix
$\mathbf{Q}(t)$	covariance of random excitation

Solution of the Wiener Problem

$\mathbf{x}(t)$	basic random variable.
$\mathbf{y}(t)$	observed random variable.
$\mathcal{Y}(t)$	linear manifold generated by $\mathbf{y}(t_0), \ldots, \mathbf{y}(t)$.
$\mathcal{Z}(t)$	linear manifold generated by $\tilde{\mathbf{y}}(t \mid t-1)$.
$\mathbf{x}^*(t_1 \mid t)$	optimal estimate of $\mathbf{x}(t_1)$ given $\mathcal{Y}(t)$.
$\tilde{\mathbf{x}}(t_1 \mid t)$	error in optimal estimate of $\mathbf{x}(t_1)$ given $\mathcal{Y}(t)$.

Optimal Estimates

To have a concrete description or the type of problems to be studied, consider the following situation. We are given signal $x_1(t)$ and noise $x_2(t)$. Only the sum $y(t) = x_1(t) + x_2(t)$ can be observed. Suppose we have observed and know exactly the values of $y(t_0)$, ..., $y(t)$. What can we infer from this knowledge in regard to the (unobservable) value of the signal at $t = t_1$, where t_1 may be less than, equal to, or greater than t? If $t_1 < t$, this is a *data-smoothing (interpolation)* problem. If $t_1 = t$, this is called *filtering*. If $t_1 > t$, we have a *prediction* problem. Since our treatment will be general enough to include these and similar problems, we shall use hereafter the collective term *estimation*.

As was pointed out by Wiener [1], the natural setting of the estimation problem belongs to the realm of probability theory and statistics. Thus signal, noise, and their sum will be random variables, and consequently they may be regarded as random processes. From the probabilistic description of the random processes we can determine the probability with which a particular sample of the signal and noise will occur. For any given set of measured values $\eta(t_0)$, ..., $\eta(t)$ of the random variable $y(t)$ one can then also determine, in principle, the probability of simultaneous occurrence of various values $\xi_1(t)$ of the random variable $x_1(t_1)$. This is the conditional probability distribution function

New Results in Linear Filtering and Prediction Theory[1]

R. E. KALMAN
Research Institute for Advanced Study,[2] Baltimore, Maryland

R. S. BUCY
The Johns Hopkins Applied Physics Laboratory, Silver Spring, Maryland

A nonlinear differential equation of the Riccati type is derived for the covariance matrix of the optimal filtering error. The solution of this "variance equation" completely specifies the optimal filter for either finite or infinite smoothing intervals and stationary or nonstationary statistics.

The variance equation is closely related to the Hamiltonian (canonical) differential equations of the calculus of variations. Analytic solutions are available in some cases. The significance of the variance equation is illustrated by examples which duplicate, simplify, or extend earlier results in this field.

The Duality Principle relating stochastic estimation and deterministic control problems plays an important role in the proof of theoretical results. In several examples, the estimation problem and its dual are discussed side-by-side.

Properties of the variance equation are of great interest in the theory of adaptive systems. Some aspects of this are considered briefly.

1 Introduction

At present, a nonspecialist might well regard the Wiener-Kolmogorov theory of filtering and prediction [1, 2][3] as "classical'—in short, a field where the techniques are well established and only minor improvements and generalizations can be expected.

That this is not really so can be seen convincingly from recent results of Shinbrot [3], Steeg [4], Pugachev [5, 6], and Parzen [7]. Using a variety of time-domain methods, these investigators have solved some long-standing problems in *nonstationary* filtering and prediction theory. We present here a unified account of our own independent researches during the past two years (which overlap with much of the work [3–7] just mentioned), as well as numerous new results. We, too, use time-domain methods, and obtain major improvements and generalizations of the conventional Wiener theory. In particular, our methods apply without modification to multivariate problems.

The following is the historical background of this paper.

In an extension of the standard Wiener filtering problem, Follin [8] obtained relationships between time-varying gains and error variances for a given circuit configuration. Later, Hanson [9] proved that Follin's circuit configuration was actually optimal for the assumed statistics; moreover, he showed that the differential equations for the error variance (first obtained by Follin) follow rigorously from the Wiener-Hopf equation. These results were then generalized by Bucy [10], who found explicit relationships between the optimal weighting functions and the error variances; he also gave a rigorous derivation of the variance equations and those of the optimal filter for a wide class of nonstationary signal and noise statistics.

Independently of the work just mentioned, Kalman [11] gave a new approach to the standard filtering and prediction problem. The novelty consisted in combining two well-known ideas:

(i) the "state-transition" method of describing dynamical systems [12–14], and

(ii) linear filtering regarded as orthogonal projection in Hilbert space [15, pp. 150–155].

As an important by-product, this approach yielded the *Duality Principle* [11, 16] which provides a link between (stochastic) filtering theory and (deterministic) control theory. Because of the duality, results on the optimal design of linear control systems [13, 16, 17] are directly applicable to the Wiener problem. Duality plays an important role in this paper also.

When the authors became aware of each other's work, it was soon realized that the principal conclusion of both investigations was identical, in spite of the difference in methods:

Rather than to attack the Wiener-Hopf integral equation directly, it is better to convert it into a nonlinear differential equation, whose solution yields the covariance matrix of the minimum filtering error, which in turn contains all necessary information for the design of the optimal filter.

2 Summary of Results: Description

The problem considered in this paper is stated precisely in Section 4. There are two main assumptions:

(A_1) A sufficiently accurate model of the message process is given by a linear (possibly time-varying) dynamical system excited by white noise.

(A_2) Every observed signal contains an additive white noise component.

Assumption (A_2) is unnecessary when the random processes in question are sampled (discrete-time parameter); see [11]. Even in the continuous-time case, (A_2) is no real restriction since it can be removed in various ways as will be shown in a future paper. Assumption (A_1), however, is quite basic; it is analogous to but somewhat less restrictive than the assumption of rational spectra in the conventional theory.

Within these assumptions, we seek the best linear estimate of the message based on past data lying in either a finite or infinite time-interval.

The fundamental relations of our new approach consist of five equations:

[1] This research was partially supported by the United States Air Force under Contracts AF 49(638)-382 and AF 33(616)-6952 and by the Bureau of Naval Weapons under Contract NOrd-73861.

[2] 7212 Bellona Avenue.

[3] Numbers in brackets designate References at the end of paper.

Contributed by the Instruments and Regulators Division of The American Society of Mechanical Engineers and presented at the Joint Automatic Controls Conference, Cambridge, Mass., September 7–9, 1960. Manuscript received at ASME Headquarters, May 31, 1960. Paper No. 60—JAC-12.

$$\mathcal{E}[\mu(t) - \mathcal{E}\mu(t)]^2 \geqq \|\mathbf{x}^*\|^2_{\mathbf{M}^{-1}(t_0,\, t)} \tag{32}$$

Every costate $\mathbf{x}^*$ *has a minimum-variance unbiased estimator for which the equality sign holds in (32) if and only if* $\mathbf{M}$ *is positive definite.* This motivates the use of condition (A_4') in Theorem 3 and the term "completely observable."

(*i*) It can be shown [17] that in the constant case complete observability is equivalent to the easily verified condition:

$$\operatorname{rank}[\mathbf{H}',\ \mathbf{F}'\mathbf{H}',\ \ldots,\ (\mathbf{F}')^{n-1}\mathbf{H}'] = n \tag{33}$$

where the square brackets denote a matrix with n rows and np columns.

(9) *Stability of the optimal filter.* It should be realized now that the *optimality* of the filter (I) does not at the same time guarantee its *stability*. The reader can easily check this by constructing an example (for instance, one in which (10–11) consists of two noninteracting systems). To establish weak sufficient conditions for stability entails some rather delicate mathematical technicalities which we shall bypass and state only the best final result currently available.

First, some additional definitions.

We say that the model (10–11) is *uniformly completely observable* if there exist fixed constants, α_1, α_2, and σ such that

$$\alpha_1\|\mathbf{x}^*\|^2 \leqq \|\mathbf{x}^*\|^2_{\mathbf{M}(t-\sigma,\, t)} \leqq \alpha_2\|\mathbf{x}^*\|^2 \quad \text{for all} \quad \mathbf{x}^* \text{ and } t.$$

Similarly, we say that a model is *completely controllable* [*uniformly completely controllable*] if the dual model is completely observable [uniformly completely observable]. For a discussion of these motions, the reader may refer to [17]. It should be noted that the property of "uniformity" is always true for constant systems.

We can now state the central theorem of the paper:

THEOREM 4. *Assume that the model of the message process is*

(A_4'') *uniformly completely observable;*

(A_5) *uniformly completely controllable;*

(A_6) $\alpha_3 \leqq \|\mathbf{Q}(t)\| \leqq \alpha_4$, $\alpha_5 \leqq \|\mathbf{R}(t)\| \leqq \alpha_6$ for all t;

(A_7) $\|\mathbf{F}(t)\| \leqq \alpha_7$.

Then the following is true:

(i) *The optimal filter is uniformly asymptotically stable;*

(ii) *Every solution* $\mathbf{\Pi}(t;\ \mathbf{P}_0,\ t_0)$ *of the variance equation* (IV) *starting at a symmetric nonnegative matrix* $\mathbf{P}_0$ *converges to* $\bar{\mathbf{P}}(t)$ *(defined in Theorem 3) as* $t \to \infty$.

Remarks. (*j*) A filter which is not *uniformly* asymptotically stable may have an unbounded response to a bounded input [21]; the practical usefulness of such a filter is rather limited.

(*k*) Property (ii) in Theorem 4 is of central importance since it shows that the variance equation is a "stable" computational method that may be expected to be rather insensitive to roundoff errors.

(*l*) The speed of convergence of $\mathbf{P}_0(t)$ to $\bar{\mathbf{P}}(t)$ can be estimated quite effectively using the second method of Lyapunov; see [17].

(10) *Solution of the classical Wiener problem.* Theorems 3 and 4 have the following immediate corollary:

THEOREM 5. *Assume the hypotheses of Theorems 3 and 4 are satisfied and that* $\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$, $\mathbf{Q}$, $\mathbf{R}$, *are constants.*

Then, if $t_0 = -\infty$, *the solution of the estimation problem is obtained by setting the right-hand side of* (IV) *equal to zero and solving the resulting set of quadratic algebraic equations. That solution which is nonnegative definite is equal to* $\bar{\mathbf{P}}$.

To prove this, we observe that, by the assumption of constancy, $\bar{\mathbf{P}}(t)$ is a constant. By Theorem 4, all solutions of (IV) starting at nonnegative matrices converge to $\bar{\mathbf{P}}$. Hence, if a matrix $\mathbf{P}$ is found for which the right-hand side of (IV) vanishes and if this matrix is nonnegative definite, it must be identical with $\bar{\mathbf{P}}$. Note, however, that the procedure may fail if the conditions of Theorems 3 and 4 are not satisfied. See Example 4.

(11) *Solution of the Dual Problem.* For details, consult [17]. The only facts needed here are the following: The optimal control law is given by

$$\mathbf{u}^*(t^*) = -\mathbf{K}^*(t^*)\mathbf{x}(t^*) \tag{34}$$

where $\mathbf{K}^*(t^*)$ satisfies the duality relation

$$\mathbf{K}^*(t^*) = \mathbf{K}'(t) \tag{35}$$

and is to be determined by duality from formula (III). The value of the performance index (20) may be written in the form

$$\min_{\mathbf{u}^*} V(\mathbf{x}^*;\ t^*,\ t_0^*,\ \mathbf{u}^*) = \|\mathbf{x}^*\|^2_{\mathbf{\Pi}^*(t^*;\ \mathbf{x}^*,\, t_0^*)}$$

where $\mathbf{\Pi}^*(t^*;\ \mathbf{x}^*,\ t_0^*)$ is the solution of the dual of the variance equation (IV).

It should be carefully noted that the hypotheses of Theorem 4 are invariant under duality. Hence essentially the same theory covers both the estimation and the regular problem, as stated in Section 5.

The vector-matrix block diagram for the optimal regulator is shown in Fig. 11.

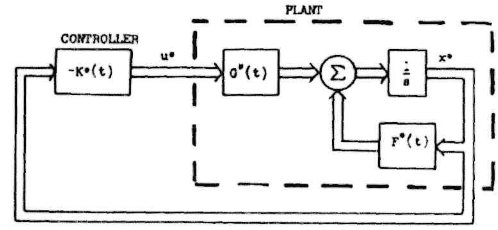


Fig. 11 General block diagram of optimal regulator

(12) *Computation of the covariance matrix for the message process.* To apply Theorem 1, it is necessary to determine cov $[\mathbf{x}(t_0),\ \mathbf{x}(t_0)]$. This may be specified as part of the problem statement as in Example 5. On the other hand, one might assume that the message model has reached steady state (see (A_3)), in which case from (13) and (12) we have that

$$\mathbf{S}(t) = \operatorname{cov}[\mathbf{x}(t),\ \mathbf{x}(t)] = \int_{-\infty}^{t} \mathbf{\Phi}(t, \tau)\mathbf{G}(\tau)\mathbf{Q}(\tau)\mathbf{G}'(\tau)\mathbf{\Phi}'(t, \tau)d\tau$$

provided the model (10) is asymptotically stable. Differentiating this expression with respect to t we obtain the following differential equation for $\mathbf{S}(t)$

$$d\mathbf{S}/dt = \mathbf{F}(t)\mathbf{S} + \mathbf{S}\mathbf{F}'(t) + \mathbf{G}(t)\mathbf{Q}(t)\mathbf{G}'(t) \tag{36}$$

This formula is analogous to the well-known lemma of Lyapunov [21] in evaluating the integrated square of a solution of a linear differential equation. In case of a constant system, (36) reduces to a system of linear algebraic equations.

8 Derivation of the Fundamental Equations

We first deduce the matrix form of the familiar Wiener-Hopf integral equation. Differentiating it with respect to time and then using (10–11), we obtain in a very simple way the fundamental equations of our theory.

Much cumbersome manipulation of integrals can be avoided by recognizing, as has been pointed out by Pugachev [27], that the Wiener-Hopf equation is a special case of a simple geometric principle: *orthogonal projection.*

Consider an abstract space $\mathfrak{X}$ such that an inner product (X, Y) is defined between any two elements X, Y of $\mathfrak{X}$. The norm is defined by $\|X\| = (X, X)^{1/2}$. Let $\mathfrak{U}$ be a subspace of $\mathfrak{X}$. We

FROM WIENER TO KALMAN

IMPORTANCE OF WIENER: TO SEE THE SIGNALS OF INTEREST AS STOCHASTIC PROCESSES ⇒ STATISTICAL SIGNAL PROCESSING

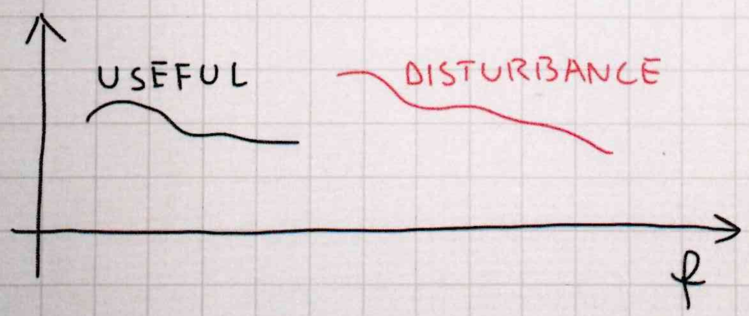
MORE POWERFUL THAN DETERMINISTIC,

MANY TIMES WE CAN MEASURE

(SIGNAL OF INTEREST) + (DISTURBANCE NOISE)

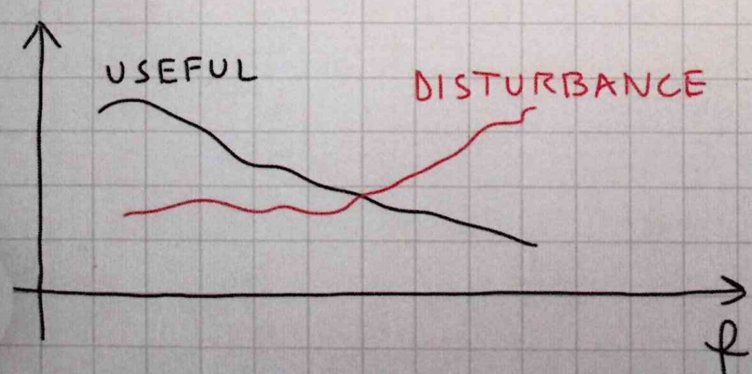
AND THESE TWO SIGNALS DO NOT LIVE ON DIFFERENT FREQUENCIES

IF ONE HAD



JUST A LOW-PASS FILTER IS NEEDED

BUT ONE COULD WELL HAVE



IF STATISTICAL PROPERTIES OF THE SIGNALS ARE KNOWN WE CAN STILL OBTAIN A GOOD ESTIMATE OF THE SIGNAL OF INTEREST

LIMITATIONS OF THE WIENER APPROACH

WE HAVE SAID THAT IT RELIES ON STATIONARY ASSUMPTIONS.
EVEN IF IT CAN ALSO (IN PRINCIPLE) HANDLE NON RATIONAL SPECTRA, STATIONARITY IS AN ASSUMPTION OFTEN VIOLATED IN REAL APPLICATIONS

IN ROBOTICS, BIOMEDICINE, BIOINFORMATICS STRONGLY NONSTATIONARY SIGNALS ARISE.
IN INDUSTRIAL PROCESSES LOAD DISTURBANCES ARE IMPULSIVE INPUTS TO THE SYSTEM.
MEASUREMENT NOISES CAN BE OF THE FORM

$$e_i \sim N(0, \sigma_i^2), \quad \text{E.G. } \sigma_i^2 \sim (CV\% \cdot y_i)^2$$

WIENER CAN ALSO HANDLE

∞-DIM LINEAR SYSTEMS

BUT THE FORMULA TO OBTAIN THE FILTER IN PRACTICE CAN BE USED ONLY WHEN SPECTRA ARE RATIONAL

FILTER IN PRACTICE CAN BE USED ONLY WHEN SPECTRA ARE RATIONAL, AND TO HAVE A FINITE-MEMORY FILTER WE NEED TO START FROM RATIONAL SPECTRA

KALMAN

SINCE RATIONAL SPECTRA ARE GENERATED BY FINITE-DIMENSIONAL LINEAR SYSTEMS, THE FUNDAMENTAL IDEA OF KALMAN WAS TO START USING SUCH SYSTEMS AND IN STATE SPACE FORM

VERY STRONG CONSEQUENCES

- THE PROBLEM OF THE EXPLICIT CALCULATION OF THE TRANSFER FUNCTION OF THE FILTER DISAPPEARS.
  THE ESTIMATOR HAS A KNOWN ALGORITHMIC STRUCTURE: FINITE-DIMENSIONAL LINEAR SYSTEM IN STATE SPACE DRIVEN BY THE OBSERVATIONS (REACHING MOON WITH NO POWERFUL COMPUTERS)

- NO NEED TO ASSUME TIME-INVARIANT AND STABLE SYSTEMS AS

- NO NEED TO ASSUME TIME-INVARIANT AND STABLE SYSTEMS AS SIGNAL GENERATORS

- NO NEED TO START FROM $t_0 = -\infty$ UNCERTAINTY ON SYSTEM INITIAL CONDITION CAN BE HANDLED

- CALCULATIONS MADE VIA PROJECTIONS ONTO FINITE-DIMENSIONAL SUBSPACES OF RANDOM VARIABLES

  ("SIMPLER" BUT MORE POWERFUL THEORY)

- DUALITY WITH OPTIMAL CONTROL

FINITE-DIMENSIONAL STOCHASTIC STATE-SPACE MODEL IN DISCRETE-TIME

$$\Sigma \begin{cases} x(t+1) = A(t)\,x(t) + B(t)\,n(t), & t \geq t_0 \\ y(t) = C(t)\,x(t) + D(t)\,n(t) \\ x(t_0) = x_0 \end{cases}$$

- $n(t)$ IS p-DIMENSIONAL WHITE NOISE OF UNIT VARIANCE

$$E\left[n(t)\,n^T(s)\right] = I_p\,\delta(t-s)$$

FINITE-DIMENSIONAL STOCHASTIC STATE-SPACE MODEL IN DISCRETE-TIME

$$\Sigma \begin{cases} x(t+1) = A(t)\,x(t) + B(t)\,n(t), \quad t \geq t_0 \\ y(t) = C(t)\,x(t) + D(t)\,n(t) \\ x(t_0) = x_0 \end{cases}$$

- $n(t)$ IS p-DIMENSIONAL WHITE NOISE OF UNIT VARIANCE

$$E\left[n(t)\,n^T(s)\right] = I_p\,\delta(t-s)$$

$$E\left[n(t)\right] = 0 \quad \forall t$$

- $\{x(t)\}$ ARE THE STATES,

  n-DIM. RANDOM VECTORS

- x_0 IS THE INITIAL STATE, A RANDOM VECTOR

$$\left.\begin{aligned} E\,x_0 &= \mu_0, \\ VAR\,x_0 &= \Sigma_0 \end{aligned}\right. \Rightarrow x_0 \sim (\mu_0, \Sigma_0)$$

IT IS UNCORRELATED FROM $n(t)$, I.E.

$$COV(x_0, n(t)) = E\left[(x_0-\mu_0)\,n^T(t)\right] = 0_{n\times p}$$

$$\forall t$$

- $\{y(t)\}$ ARE THE m-DIM OUTPUTS, THEY ARE RANDOM VECTORS

DEFINITION:

A PROCESS $\{y(t)\}$ IS SAID TO HAVE FINITE DIMENSION IF IT ADMITS A REPRESENTATION WITH THE Σ ABOVE. Σ CAN BE NON UNIQUE, ANY Σ THAT REPRESENTS $\{y(t)\}$ IS SAID A REALIZATION OF $\{y(t)\}$

ASSUMPTION ONLY DONE TO SIMPLIFY NOTATION

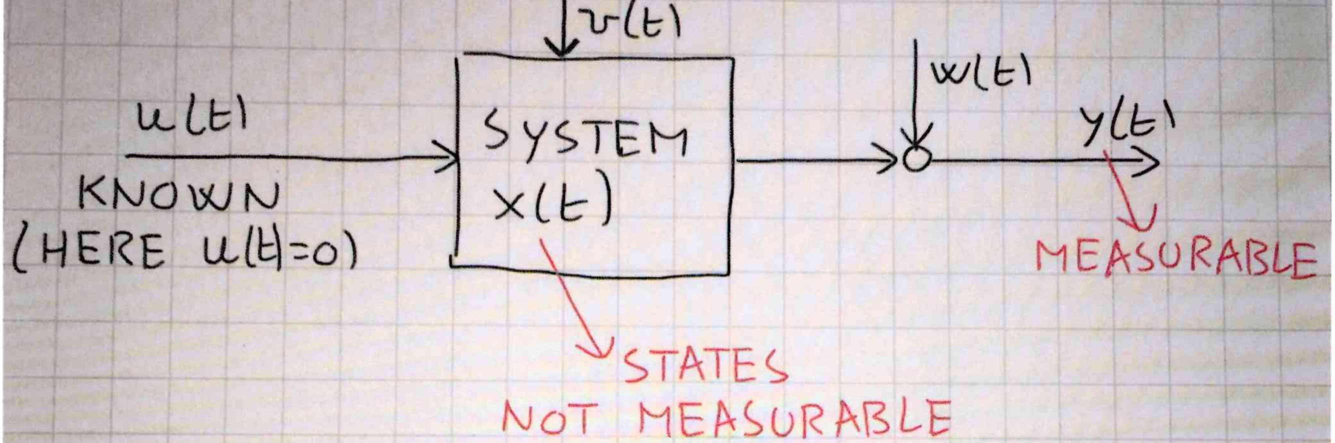
A, B, C, D DO NOT DEPEND ON t

$$A \in \mathbb{R}^{n \times n}, \; B \in \mathbb{R}^{n \times p}, \; C \in \mathbb{R}^{m \times n}, \; D \in \mathbb{R}^{m \times p}$$

FUNDAMENTAL ENGINEERING PARADIGM FOR MANY APPLICATIONS

FUNDAMENTAL ENGINEERING PARADIGM FOR MANY APPLICATIONS

$$v(t) := B\, n(t), \quad w(t) := D\, n(t)$$



WE WANT TO PREDICT FUTURE OUTPUTS OR STATES ACCOUNTING FOR UNEXPECTE EVENTS DESCRIBED BY DISTURBANCES $v(t)$
(VERY HARD TO CAST THEM AS DETERMINISTIC SYSTEM INPUTS)

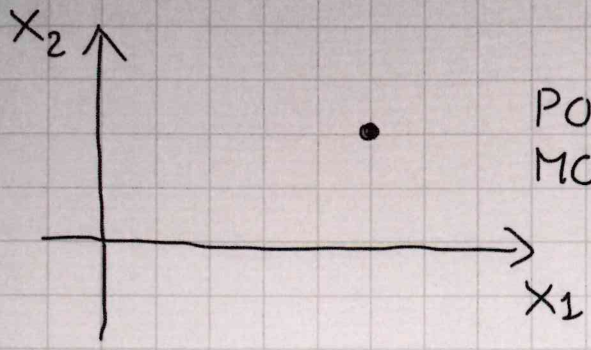
NOTE: MANY TIMES $v(t) \perp w(s)$

$$\forall t, s$$

EXAMPLE:

EXAMPLE:

ROBOT LOCALIZATION



POSITION OF A MOBILE AGENT ON THE PLANE

$$x_1(t+1) = x_1(t) + \boxed{u_1(t)} + \boxed{v_1(t)}$$
$$x_2(t+1) = x_2(t) + \boxed{u_2(t)} + \boxed{v_2(t)}$$

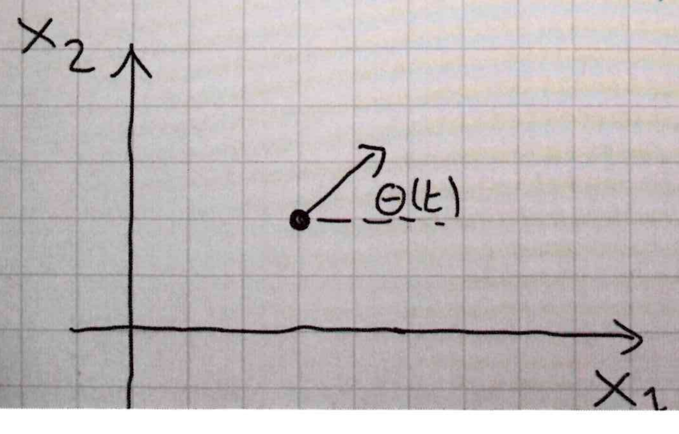
KNOWN INPUTS TO MAKE THE ROBOT MOVE

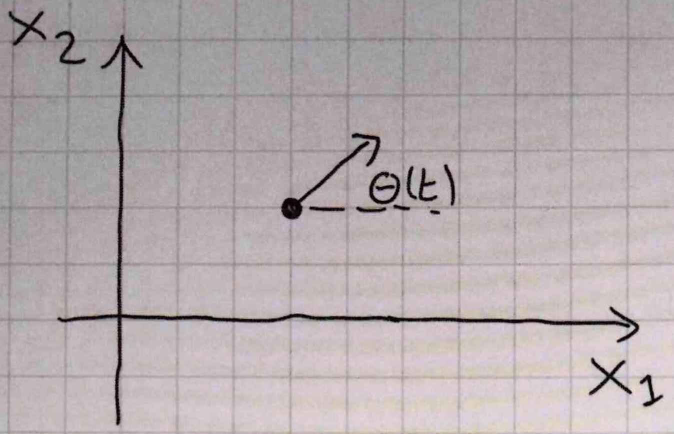
OBSTACLES, WALLS, ...

$$y_1(t) = x_1(t) + w_1(t)$$
$$y_2(t) = x_2(t) + w_2(t)$$

LINEAR

POSE ESTIMATION (BEARING)





$T(t)$ = TRANSLATION SPEED

$r(t)$ = ROTATION SPEED

$$x_1(t+1) = x_1(t) + \frac{T(t)}{r(t)} \cdot \sin(\theta(t) + r(t)) - \frac{T(t)}{r(t)} \cdot \sin(\theta(t)) + v_1(t)$$

$$x_2(t+1) = x_2(t) + \frac{T(t)}{r(t)} \cdot \cos\theta(t) - \frac{T(t)}{r(t)} \cdot \cos(\theta(t) + r(t)) + v_2(t)$$

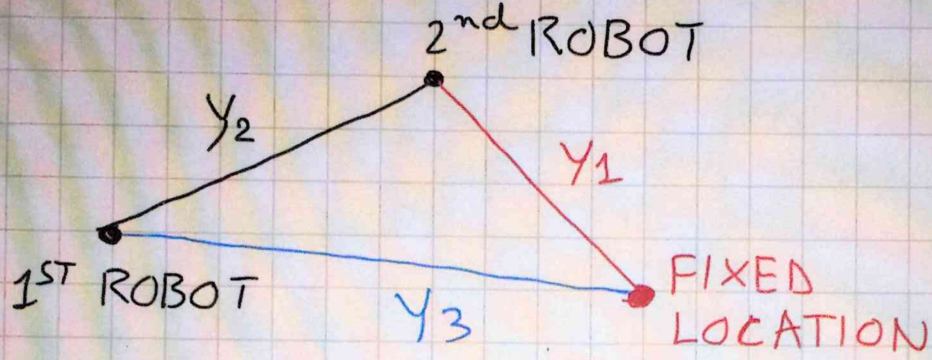
$$\theta(t+1) = \theta(t) + r(t)$$

NONLINEAR

THEN GROUP OF ROBOTS CAN

NONLINEAR

THEN GROUP OF ROBOTS CAN BE CONSIDERED, WITH DISTANCE MEASUREMENTS



RATIONAL TRANSFER FUNCTIONS: REALIZATION THEORY

LET US RECALL WHICH I/O RELATIONS $\Sigma = (A, B, C, D)$ CAN DESCRIBE (THINK $n(t)$ AS THE DETERMINISTIC INPUT $u(t)$ SEEN IN PREVIOUS COURSES)

$$Y(z) = W_{\Sigma}(z)\, U(z)$$

RATIONAL TRANSFER FUNCTIONS: REALIZATION THEORY

LET US RECALL WHICH I/O RELATIONS $\Sigma = (A, B, C, D)$ CAN DESCRIBE (THINK $n(t)$ AS THE DETERMINISTIC INPUT $u(t)$ SEEN IN PREVIOUS COURSES)

$$Y(z) = W_\Sigma(z)\, U(z)$$

$$W_\Sigma(z) = C(zI - A)^{-1} B + D$$

↓

MATRIX WITH ENTRIES GIVEN BY RATIONAL TRANSFER FUNCTIONS

LET A BE STABLE ($|\lambda_i| < 1 \ \forall i$)

IF $u(t)$ IS STOCHASTIC, IN PARTICULAR

$u(t) =$ WHITE NOISE $n(t)$ OF UNIT VARIANCE

AND $t_0 = -\infty$ (THE INPUT STARTS IN THE REMOTE PAST)

- THEN $y(t)$ IS A STATIONARY

THEN $y(t)$ IS A STATIONARY STOCHASTIC PROCESS WITH SPECTRUM

$$S_y(z) = W_{\Sigma}(z)\, W_{\Sigma}^T(z^{-1})$$

NOW, CONSIDER THE INVERSE PROBLEM:

GIVEN A RTF $W(z)$ $\Longrightarrow$ FIND A SYSTEM Σ WHICH IS A REALIZATION OF $W(z)$

THEOREM: GIVEN A PROPER RTF $W(z)$ $\left(= \dfrac{\text{POLYNOMIAL}}{\text{POLYNOMIAL}} = \dfrac{n(s)}{d(s)},\ \deg n(s) \leq \deg d(s) \right)$

THERE ALWAYS EXISTS $\Sigma = (A, B, C, D)$ WHICH IS A REALIZATION OF $W(z)$.

SO, RATIONAL SPECTRA ARE ASSOCIATED WITH FINITE-DIM. STATE SPACE MODELS

SKETCH

$$W(z) = \bar{W}(z) + \lim_{z \to \infty} W(z)$$

$$= \underbrace{\bar{W}(z)}_{\text{STRICTLY PROPER}} + D$$

AND E.G. IN THE SISO CASE, IF

$$\bar{W}(z) = \frac{b_0 + b_1 z + \ldots + b_{n-1} z^{n-1}}{a_0 + a_1 z + \ldots + z^n}$$

ONE REALIZATION E.G. IS

$$A = \begin{bmatrix} 0 & 1 & & \\ & & \ddots & \\ & & & 1 \\ -a_0 & -a_1 & \cdots & -a_{n-1} \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$

$$C = \begin{bmatrix} b_0 & b_1 & \cdots & b_{n-1} \end{bmatrix}$$

$$C = [\, b_0 \; b_1 \; \ldots \; b_{n-1} \,]$$

THIS IS THE CANONICAL CONTROL FORM

$\blacksquare$

THE STATE PROCESS

$$x(t+1) = Ax(t) + Bn(t)$$

PROPOSITION:

$\{x(t)\}$ IS A MARKOV PROCESS IN WEAK SENSE, I.E.

$$\hat{E}\left[x(t) \,\middle|\, x_0, \ldots, x_s\right] = \hat{E}\left[x(t) \mid H_s(x)\right]$$

$$= \hat{E}\left[x(t) \mid x(s)\right]$$

$$\forall t \geq s$$

PROOF:

IT SUFFICES WRITING

PROOF:

IT SUFFICES WRITING

$$x(s+1) = A x(s) + B n(s)$$

$$x(s+2) = A(A x(s) + B n(s)) + B n(s+1)$$

$$= A^2 x(s) + [B \;\; AB] \begin{bmatrix} n(s+1) \\ n(s) \end{bmatrix}$$

$$\vdots$$

$$x(t) = \underbrace{A^{t-s} x(s)}_{\in H_s(x)} + \underbrace{\sum_{k=s}^{t-1} A^{t-1-k} B n(k)}_{\substack{\perp x_0, \dots, x_s, \\ \text{I.E. } \perp H_s(x)}}$$

$$= f_{H_s} + f_{H_s^\perp}$$

SO WE HAVE OBTAINED THE

UNIQUE DECOMPOSITION RELATED

UNIQUE DECOMPOSITION RELATED TO THE PROJECTION THEOREM AND THE PROJECTION OF $x(t)$ ONTO $H_s(x)$ IS $A^{t-s}x(s)$ PROVING THE THEOREM

$\blacksquare$

NOTE: IN THE GAUSSIAN CASE $\{x(t)\}$ IS MARKOV SINCE IT ALSO HOLDS THAT

$$p\left(x(t) \mid H_s(x)\right) = p\left(x(t) \mid x(s)\right)$$

HOMEWORKS

PROVE THAT

a) $$\hat{E}\left[x(t) \mid H(x^s, y^{s-1})\right], \quad t \geq s$$

$$\parallel$$

$$\hat{E}\left[x(t) \mid x(s)\right]$$

(x^s = ALL THE STATES UNTIL INSTANT s)

NOTE: IN THE GAUSSIAN CASE $\{x(t)\}$ IS MARKOV SINCE IT ALSO HOLDS THAT

$$p\left(x(t) \mid H_s(x)\right) = p\left(x(t) \mid x(s)\right)$$

HOMEWORKS

PROVE THAT

a) $$\hat{E}\left[x(t) \mid H(x^s, y^{s-1})\right], \quad t \geq s$$

$$\parallel$$

$$\hat{E}\left[x(t) \mid x(s)\right]$$

(x^s = ALL THE STATES UNTIL INSTANT s)

b) $$\hat{E}\left[y(t) \mid H(x^s, y^{s-1})\right], \quad t \geq s$$

$$\parallel$$

$$\hat{E}\left[y(t) \mid x(s)\right]$$

SOLUTION OF POINT a)

WE WANT TO PROVE THAT

$$\hat{E}\left[x(t) \mid H(x^s, y^{s-1})\right]$$
$$\parallel$$
$$\hat{E}\left[x(t) \mid x(s)\right] \quad , \quad t \geq s$$

PRELIMINARY NOTE:

- WE ALREADY PROVED THAT

$$\hat{E}\Big[x(t) \mid \overbrace{H(x_0, \ldots, x(s))}^{= H(x^s) \text{ OR JUST } x^s}\Big] = \hat{E}\left[x(t) \mid x(s)\right]$$

- $x(t_0) = x_0$

$x(t_0+1) = A x_0 + B n(t_0) \Rightarrow$ $x(t_0+1)$ IS LINEARLY GENERATED BY $x_0, n(t_0)$

$x(t_0+2) = A x(t_0+1) + B n(t_0+1) \Rightarrow x(t_0+2)$ IS LIN. GEN. BY $x(t_0), n(t_0), n(t_0+1)$

$\vdots$

$\Rightarrow x_0, x(t_0+1), \ldots, x(s)$ ARE LINEARLY GENERATED BY $x_0, n(t_0), \ldots, n(s-1)$

GENERATED BY $x_0, n(t_0), \dots, n(s-1)$

$$\Downarrow$$

$$H(x^s) \subseteq H(x_0, n(t_0), \dots, n(s-1))$$

- $y(t_0) = Cx_0 + Dn(t_0)$

$\vdots$

$y(s-1) = Cx(s-1) + Dn(s-1)$

$$\Downarrow$$

$$H(y^{s-1}) \subseteq H(x_0, n(t_0), \dots, n(s-1))$$

$$\Downarrow$$

$$H(x^s, y^{s-1}) \subseteq H(x_0, n(t_0), \dots, n(s-1)) \circledast$$

NOW WE CAN PROVE a)

WE KNOW THAT

$$x(t) = \underbrace{A^{t-s} x(s)}_{\in H(x^s, y^{s-1})} + \underbrace{\sum_{k=s}^{t-1} A^{t-1-k} B n(k)}_{\subseteq H(n(s), \dots, n(t-1))}$$

$$\Downarrow$$

$$\perp H(x_0, n(t_0), \dots, n(s-1))$$

$\Downarrow$ USING $\circledast$

$$\Downarrow$$

$$H(y^{s-1}) \subseteq H(x_0, n(t_0), \dots, n(s-1))$$

$$\Downarrow$$

$$H(x^s, y^{s-1}) \subseteq H(x_0, n(t_0), \dots, n(s-1)) \quad \circledast$$

NOW WE CAN PROVE a)

WE KNOW THAT

$$x(t) = \underbrace{A^{t-s} x(s)}_{\in H(x^s, y^{s-1})} + \underbrace{\sum_{k=s}^{t-1} A^{t-1-k} B n(k)}_{\subseteq H(n(s), \dots, n(t-1))}$$

$$\Downarrow$$

$$\perp H(x_0, n(t_0), \dots, n(s-1))$$

$\Downarrow$ USING $\circledast$

$$\perp H(x^s, y^{s-1})$$

SO, BY THE PROJ. TH. ONE HAS

$$\hat{E}\left[x(t) \mid H(x^s, y^{s-1})\right] = A^{t-s} x(s)$$

$$\left(= \hat{E}\left[x(t) \mid H(x^s)\right] = \hat{E}\left[x(t) \mid x(s)\right] \right)$$

FIRST- AND SECOND-ORDER MOMENTS OF $\{x(t)\}$, $\{y(t)\}$

$$x(t+1) = A x(t) + B n(t)$$
$$y(t) = C x(t) + D n(t)$$

$$\left.\begin{aligned} E x_0 &= \mu_0 \\ VAR x_0 &= \Sigma_0 \end{aligned}\right\} \text{HOW DO THEY CHANGE IN TIME?}$$

MEANS

$$\mu_x(t+1) = A \mu_x(t), \quad \mu_x(t_0) = \mu_0$$
$$\mu_y(t) = C \mu_x(t)$$

COVARIANCE

$$COV(x(s+1), x(s)) =: \Sigma_x(s+1, s)$$
$$= COV(A x(s) + B n(s), x(s))$$
$$= A\, COV(x(s), x(s))$$
$$= A\, VAR(x(s)) = A \Sigma(s)$$

ONE ALSO OBTAINS

ONE ALSO OBTAINS

$$\Sigma_x(s, s+1) = \Sigma(s) A^T$$

AND IN GENERAL

$$\Sigma_x(t,s) = A^{t-s}\,\Sigma(s), \quad t \geq s$$

$$\Sigma_x(t,s) = \Sigma(t)\,(A^T)^{s-t}, \quad t \leq s$$

VARIANCE

$$\mathrm{VAR}\; x(t+1) = \mathrm{VAR}\left[A x(t) + B n(t) \right]$$

$$= A\,\mathrm{VAR}\,x(t)\,A^T + B\,\mathrm{VAR}\,n(t)\,B^T$$

$$= A\,\underbrace{\mathrm{VAR}\,x(t)}_{=\,\Sigma(t)}\,A^T + BB^T$$

$$\Downarrow$$

$$\Sigma(t+1) = A\,\Sigma(t)\,A^T + BB^T$$

$$\Sigma(t_0) = \Sigma_0$$

STATE-SPACE MODELS AND STATIONARY PROCESSES

$x(t+1) = Ax(t) + Bn(t)$, $v(t) := Bn(t)$

$x(t_0) = x_0 \sim (\mu_0, \Sigma_0)$, $\quad \mathrm{VAR}\, v(t) = BB^T = Q$

PROPOSITION:

IF A IS STABLE ($|\lambda_i| < 1 \ \forall i$, λ_i = EIGENVALUE)

$x(t)$, FOR $t \to +\infty$, TENDS TO BECOME A STATIONARY PROCESS IN WEAK SENSE (I.E. STATIONARY IN MEAN AND COVARIANCE).

ONE HAS

$$\lim_{t \to +\infty} \mu(t) = 0 \ , \quad \mathrm{VAR}\, \underset{t \to +\infty}{x(t)} = \bar{\Sigma}$$

$$\lim_{t \to +\infty} \mu(t) = 0 , \quad \lim_{t \to +\infty} \mathrm{VAR}\, x(t) = \bar{\Sigma}$$

$$\Sigma_x(t-s) \simeq A^{t-s} \bar{\Sigma} , \quad t \geq s$$

WHERE

$\bar{\Sigma} = \lim_{t \to +\infty} \Sigma_x(t)$ IS THE SOLUTION OF

$$\bar{\Sigma} = A \bar{\Sigma} A^T + BB^T \quad \text{LYAPUNOV}$$

IF ONE ALSO HAS

$\mu_0 = 0 , \quad \Sigma_0 = \bar{\Sigma}$

THEN

$$\mu(t) = 0 , \quad \mathrm{VAR}\, x(t) = \bar{\Sigma} \quad \forall t$$

PROOF:

$$\mu(t_0) = \mu_0 , \quad \mu(t) = A^{t-t_0} \mu_0 \xrightarrow{t \to +\infty} 0$$

(SINCE A IS STABLE)

$\Sigma(t_0) = \Sigma_0$

PROOF:

$$\mu(t_0) = \mu_0, \quad \mu(t) = A^{t-t_0}\mu_0 \xrightarrow{t\to+\infty} 0$$

(SINCE A IS STABLE)

$$\Sigma(t_0) = \Sigma_0$$

$$\Sigma(t+1) = A\Sigma(t)A^T + BB^T \quad (*)$$

$$\Sigma(t) = A^{t-t_0}\Sigma_0 (A^{t-t_0})^T$$ (→0 SINCE A IS STABLE)

$$+ \sum_{k=0}^{t-t_0-1} A^k BB^T (A^T)^k$$

SO

$$\lim_{t\to+\infty} \Sigma(t) = \sum_{k=0}^{+\infty} A^k BB^T (A^T)^k$$ (CONVERGES SINCE A IS STABLE)

$$\| =: \bar{\Sigma} < \infty$$

NOW, IF WE TAKE THE LIMIT ON THE LEFT AND RIGHT OF (*) ONE OBTAINS THAT $\bar{\Sigma}$ SATISFIES

ONE OBTAINS THAT $\bar{\Sigma}$ SATISFIES

$$\bar{\Sigma} = A \bar{\Sigma} A^T + BB^T$$

FINALLY,

$$\mu_0 = 0 \Rightarrow \mu(t) = A^{t-t_0} \mu_0 = 0 \quad \forall t$$

$$\begin{aligned}
\Sigma_0 = \bar{\Sigma} \Rightarrow \Sigma(t_0+1) &= A \bar{\Sigma} A^T + BB^T \\
&= \bar{\Sigma} \\
\Sigma(t_0+2) &= A \, \Sigma(t_0+1) A^T \\
&\quad + BB^T \\
&= A \bar{\Sigma} A^T + BB^T \\
&= \bar{\Sigma}
\end{aligned}$$

$$\Downarrow$$

$$\Sigma(t) = \bar{\Sigma} \quad \forall t$$

NOTE: IF A IS STABLE, THERE IS A UNIQUE SOLUTION OF THE LYAPUNOV EQUATION.

SO THE $\bar{\Sigma}$ OBTAINED IN THE

$$\Sigma_0 = \bar{\Sigma} \Rightarrow \Sigma(t_0+1) = A\bar{\Sigma}A^T + BB^T$$
$$= \bar{\Sigma}$$

$$\Sigma(t_0+2) = A\,\Sigma(t_0+1)\,A^T + BB^T$$
$$= A\bar{\Sigma}A^T + BB^T$$
$$= \bar{\Sigma}$$

$$\Downarrow$$

$$\Sigma(t) = \bar{\Sigma} \quad \forall t$$

NOTE: IF A IS STABLE, THERE IS A UNIQUE SOLUTION OF THE LYAPUNOV EQUATION.

SO, THE $\bar{\Sigma}$ OBTAINED IN THE PROOF WITH A LIMIT IS THE UNIQUE $\bar{\Sigma}$ SOLVING

$$\bar{\Sigma} = A\bar{\Sigma}A^T + BB^T$$

UNIQUENESS OF $\bar{\Sigma}$

LET (BY CONTRADICTION) $\bar{\bar{\Sigma}} \neq \bar{\Sigma}$ ANOTHER SOLUTION OF THE LYAPUNOV EQUATION.

WE PROVED THAT, FOR ANY Σ_0,

$$\lim_{t \to +\infty} \Sigma(t) = \bar{\Sigma}$$

LET $\Sigma_0 = \bar{\bar{\Sigma}}$. BUT THEN

$\Sigma(t) = \bar{\bar{\Sigma}} \quad \forall t$, SO

$\bar{\bar{\Sigma}}$ MUST BE EQUAL TO $\bar{\Sigma}$

AND THE CONTRADICTION IS OBTAINED

$\blacksquare$

KALMAN FILTER EQUATIONS

$x_0 \sim (\mu_0, P_0)$

$x(t+1) = A x(t) + v(t)$

$y(t) = C x(t) + w(t)$

HANDLING NOISE CORRELATION

$x_0 \perp v(t), w(t)$ AND $v(t) \perp w(s)$
$t \neq s$

BUT

$$\mathrm{VAR} \begin{bmatrix} v(t) \\ w(t) \end{bmatrix} = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix}$$, $R > 0$ BY ASSUMPTION

IT IS NOW CONVENIENT TO WORK WITH UNCORRELATED NOISES

$v(t), w(t) \longrightarrow \tilde{v}(t), w(t)$
UNCORRELATED

WITH

$$\tilde{v}(t) = v(t) - E\left[v(t) \mid \underbrace{H(w)}_{\text{SUBSPACE GENERATED BY } \{w(t)\}}\right]$$

$$\tilde{v}(t) = v(t) - E[v(t) \mid \underbrace{H(w)}_{\substack{\text{SUBSPACE} \\ \text{GENERATED} \\ \text{BY } \{w(t)\}}}]$$

$$= v(t) - E[v(t) \mid w(t)]$$

$$= v(t) - COV(v(t), w(t)) (VAR\, w(t))^{-1} w(t)$$

$$= v(t) - SR^{-1} w(t)$$

$$= v(t) - SR^{-1} y(t) + SR^{-1} C x(t)$$

$$\Downarrow$$

$$x(t+1) = A x(t) + SR^{-1} y(t) - SR^{-1} C x(t) + \tilde{v}(t)$$

$$= \underbrace{(A - SR^{-1} C)}_{F} x(t) + \underbrace{SR^{-1} y(t)}_{\substack{\text{OUTPUT} \\ \text{INJECTION}}} + \tilde{v}(t)$$

WITH

$$VAR \begin{bmatrix} \tilde{v}(t) \\ w(t) \end{bmatrix} = \begin{bmatrix} \tilde{Q} & 0 \\ 0 & R \end{bmatrix}$$

$$\tilde{Q} = Q - SR^{-1} S^T$$

$$\tilde{Q} = Q - SR^{-1}S^T$$

VAR v VAR w

COV(v, w)

KALMAN FILTER

NOW, WE CAN START FROM

$$x(t+1) = Fx(t) + SR^{-1}y(t) + \hat{v}(t)$$

$$y(t) = Cx(t) + w(t)$$

$$x(t_0) = x_0 \sim (\mu_0, P_0)$$

$$\text{VAR } \hat{v} = \tilde{Q}$$

$$\text{VAR } w = R$$

PREDICTION AND FILTERING

$$\hat{x}(t+1|t) = \hat{E}\left[x(t+1) \mid H_t(y)\right]$$

$$\hat{x}(t+1|t+1) = \hat{E}\left[x(t+1) \mid H_{t+1}(y)\right]$$

PREDICTION AND FILTERING ERRORS

$$\tilde{x}(t+1|t) = x(t+1) - \hat{x}(t+1|t)$$

PREDICTION AND FILTERING ERRORS

$$\tilde{x}(t+1|t) = x(t+1) - \hat{x}(t+1|t)$$

$$\tilde{x}(t|t) = x(t) - \hat{x}(t|t)$$

COVARIANCES OF THE ERRORS

$$P(t+1|t) = E\left[\tilde{x}(t+1|t)\left(\tilde{x}(t+1|t)\right)^T\right]$$

$$P(t|t) = E\left[\tilde{x}(t|t)\left(\tilde{x}(t|t)\right)^T\right]$$

INNOVATION

$$e(t) = y(t) - C\,\hat{x}(t|t-1)$$

ONE HAS (ASSUMING $\mu_0 = 0$, BUT THEN WE REPLUG IT IN THE INITIAL CONDITION AND OBTAIN THE SAME EQUATIONS)

NOTES ON $\perp$ IN STATE SPACE

1) $S = S_a \oplus S_b , \quad S_a \perp S_b$

$$\Downarrow$$

$$E[x|S] = E[x|S_a] + E[x|S_b]$$

2) $\tilde{v}(t) \perp x(t)$ ($\tilde{v}(t)$ FORMS $x(t+1)$)

3) $\tilde{v}(t) \perp x_0, \tilde{v}^{t-1}, w^t$ (GENERATE ALSO $y(t_0), \dots, y(t)$)

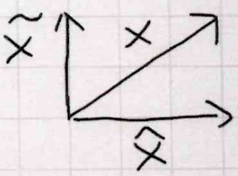
$$\Downarrow$$

$\tilde{v}(t) \perp H_t(y)$ (CONTAINS $\hat{x}(t|t)$)

$$\Downarrow$$

$$\tilde{v}(t) \perp \hat{x}(t|t)$$

4) $\tilde{v}(t) \perp \tilde{x}(t|t)$



SINCE

$\tilde{x}(t|t) = x(t) - \hat{x}(t|t)$ AND

$\tilde{v}(t) \perp x(t)$ AND $\tilde{v}(t) \perp \hat{x}(t|t)$

BY 2 BY 3

5) $e(t+1) = y(t+1) - C\hat{x}(t+1|t) = y(t+1) - \hat{y}(t+1|t)$

$\Rightarrow e(t+1)$ = ONE-STEP AHEAD OUTPUT PROJECTION ERROR

$\Rightarrow e(t+1) \perp H_t(y) , \quad H_{t+1}(y) = H_t(y) \oplus H(e(t+1))$

TIME UPDATE

$$\hat{x}(t+1|t) = F\hat{x}(t|t) + SR^{-1}y(t)$$
$$P(t+1|t) = FP(t|t)F^T + \tilde{Q}$$

PROOF

$$x(t+1) = Fx(t) + SR^{-1}y(t) + \tilde{v}(t)$$

AND WE WANT TO PROJECT ONTO $H_t(y)$. NOTE THAT

$$\tilde{v}(t) \perp \text{SPAN}\{x_0, \tilde{v}(s-1), w(s); s \leq t\} \supset H_t(y)$$

IN FACT

$x(t+1) = \dots + \tilde{v}(t)$ DOES NOT "BUILD" $y(t)$

$y(t) = Cx(t) + \dots$

THEN $\tilde{v}(t) \perp H_t(y)$ AND

$$E[x(t+1)|H_t(y)] = F\hat{x}(t+1|t) + SR^{-1}y(t)$$

SO THAT

SO THAT

$$\tilde{x}(t+1|t) = x(t+1) - \hat{x}(t+1|t)$$
$$= F x(t) + \tilde{v}(t) + SR^{-1} y(t)$$
$$- F\hat{x}(t|t) - SR^{-1} y(t)$$
$$= F(\underbrace{x(t)}_{\perp \tilde{v}(t)} - \underbrace{\hat{x}(t|t)}_{\in H_t(y)}) + \underbrace{\tilde{v}(t)}_{\perp H_t(y)}$$
$$= F\tilde{x}(t|t) + \tilde{v}(t)$$ (SUM OF TWO ⊥ R.V.)

$$\text{VAR}\ \tilde{x}(t+1|t) = F\ \text{VAR}\ \tilde{x}(t|t) F^T + \text{VAR}\ \tilde{v}(t)$$

$$\Rightarrow \underbrace{P(t+1|t)}_{\text{JUST NOTATION}} = F\, P(t|t) F^T + \tilde{Q}$$

MEASUREMENT UPDATE

$$\hat{x}(t+1|t+1) = \hat{x}(t+1|t) + L(t+1) \cdot e(t+1)$$

$$\underbrace{e(t+1)}_{\text{INNOVATION}} = y(t+1) - C\hat{x}(t+1|t)$$

$$L(t+1) = P(t+1|t)\, C^T \Lambda^{-1}(t+1)$$

$$\Lambda(t) = \text{VAR}\ e(t) = C P(t|t-1) C^T + R$$

INNOVATION

$$L(t+1) = P(t+1|t)\, C^T \Lambda^{-1}(t+1)$$

$$\Lambda(t) = VAR\, e(t) = C P(t|t-1) C^T + R$$

$$P(t+1|t+1) = P(t+1|t)$$
$$- P(t+1|t)\, C^T \Lambda^{-1}(t+1)\, C P(t+1|t)$$

PROOF

KEY POINT: USE OF INNOVATION TO PROJECT ONTO

$$H_{t+1}(y) = H_t(y) \oplus H(e(t+1))$$

USING THE FACT THAT PROJECTION ONTO TWO ORTHOGONAL SUBSPACES (LIKE $H_t(y)$ AND $H(e(t+1))$) IS THE SUM OF THE TWO PROJECTIONS.

NOTE ALSO THAT:

$e(t+1)$ BY DEFINITION IS $\perp$ $H_t(y)$ SINCE $C\hat{x}(t+1|t)$ IS THE OPTIMAL LINEAR ESTIMATE OF $y(t+1)$ BASED ON $y(t_0), \dots, y(t)$.

LINEAR ESTIMATE OF $y(t+1)$ BASED ON $y(t_0), \dots, y(t)$.

$$e(t+1) = y(t+1) - C\hat{x}(t+1|t)$$
$$= C[x(t+1) - \hat{x}(t+1|t)] + w(t+1)$$
$$= C\tilde{x}(t+1|t) + w(t+1)$$

$$\hat{E}[x(t+1)|H_t(y)] = \hat{E}[x(t+1)|H_t(y) \oplus H(e(t+1))]$$
$$= \hat{E}[x(t+1)|H_t(y)] + \hat{E}[x(t+1)|H(e(t+1))]$$

$$\Downarrow$$

$$\hat{x}(t+1|t+1) = \hat{x}(t+1|t)$$
$$+ \underbrace{cov(x(t+1), e(t+1))}_{(a)} \underbrace{[VAR(e(t+1))]^{-1}}_{(b)} e(t+1)$$

(a) $cov(x(t+1), e(t+1))$

$$= cov(x(t+1), y(t+1) - C\hat{x}(t+1|t))$$
$$= cov(x(t+1), C(x(t+1) - \hat{x}(t+1|t)) + w(t+1))$$
$$cov(x(t+1), C\tilde{x}(t+1|t))$$

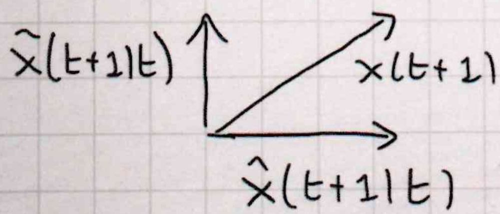
(a) $\mathrm{cov}(x(t+1), e(t+1))$

$$= \mathrm{cov}(x(t+1), y(t+1) - C\hat{x}(t+1|t))$$

$$= \mathrm{cov}\big(x(t+1), C(x(t+1) - \hat{x}(t+1|t)) + w(t+1)\big)$$

$$= \mathrm{cov}\big(x(t+1), C\tilde{x}(t+1|t)\big) + \underbrace{\mathrm{cov}(x(t+1), w(t+1))}_{=0 \text{ SINCE } x(t+1) \perp w(t+1)}$$

$$= \mathrm{cov}\big(\underbrace{\tilde{x}(t+1|t)}_{\perp H_t(y)} + \underbrace{\hat{x}(t+1|t)}_{\perp \tilde{x}(t+1|t)}, \tilde{x}(t+1|t)\big) C^T$$



$$= \mathrm{cov}\big(\tilde{x}(t+1|t), \tilde{x}(t+1|t)\big) C^T$$

$$= P(t+1|t)\, C^T$$

(b) $\mathrm{VAR}(e(t+1))$

b $\mathrm{VAR}(e(t+1))$

$$= \mathrm{VAR}\left(C\tilde{x}(t+1|t) + w(t+1) \right)$$

$\perp \hat{x}(t+1|t),\ \perp x(t+1)$

$\Longrightarrow \perp \tilde{x}(t+1|t)$

$$= \mathrm{VAR}\left(C\tilde{x}(t+1|t)\right) + \mathrm{VAR}\, w(t+1)$$

$$= C\ \mathrm{VAR}\,\tilde{x}(t+1|t)\, C^T + R$$

$$= C\ P(t+1|t)\, C^T + R$$

FINALLY, TO UPDATE THE COVARIANCES:

$$x(t+1) - \hat{x}(t+1|t+1)$$

$$= x(t+1) - \hat{x}(t+1|t) - L(t+1)\,e(t+1)$$

$$\Downarrow$$

$$\tilde{x}(t+1|t+1) = \tilde{x}(t+1|t) - L(t+1)\,e(t+1)$$

$$\Downarrow$$

$$\underbrace{\tilde{x}(t+1|t+1)} + \underbrace{L(t+1)\,e(t+1)} = \tilde{x}(t+1|t)$$

$$\Downarrow$$

$$\underbrace{\tilde{x}(t+1|t+1)}_{\perp H_{t+1}(y)} + \underbrace{L(t+1)e(t+1)}_{\in H_{t+1}(y)} = \tilde{x}(t+1|t)$$

THEN WE TAKE THE VARIANCES TO OBTAIN

$$P(t+1|t+1) + L(t+1)\underbrace{\Lambda(t+1)}_{\text{VAR}\,e(t+1)}L(t+1)^T$$

$$= P(t+1|t)$$

$\blacksquare$

INITIAL CONDITION
(NOTATION)

$$\underbrace{\hat{x}(t_0|t_0-1) = \mu_0}_{\substack{\text{PROJECTION OF}\\ x_0 \text{ ONTO A SPACE}\\ \text{WITH NULL INFORMATION}}}, \quad P(t_0|t_0-1) = P_0$$

(NOTATION)

$$\underbrace{\hat{x}(t_0 | t_0 - 1)}_{\substack{\text{PROJECTION OF} \\ x_0 \text{ ONTO A SPACE} \\ \text{WITH NULL INFORMATION}}} = \mu_0 , \quad P(t_0 | t_0 - 1) = P_0$$

OUTPUT PREDICTION

$$y(t+1) = C\,x(t+1) + w(t+1)$$

$$\Big\Downarrow \ \text{PROJECTION ONTO } H_t(y)$$

$$\hat{y}(t+1|t) = C\,\underbrace{\hat{x}(t+1|t)}_{\substack{\text{IT IS} \\ \text{RECURSIVELY} \\ \text{GIVEN BY THE} \\ \text{KF THAT THUS ALSO} \\ \text{GIVES THE OUTPUT PRED.}}} + \underbrace{0}_{\substack{\text{SINCE} \\ w(t+1) \perp H_t(y)}}$$

MULTI-STEP AHEAD PREDICTOR

$$\hat{x}(t+s+1|t) = F^{s}\,\hat{x}(t+1|t)$$

PREDICTOR AND FILTER DYNAMICS

USEFUL TO OBTAIN MORE COMPACT EQUATIONS

PREDICTOR

$$\hat{x}(t+1|t) = F\hat{x}(t|t-1) + SR^{-1}y(t)$$
$$+ K(t)\left(y(t) - C\hat{x}(t|t-1)\right)$$

$$K(t) = FL(t) = FP(t|t-1)C^T\Lambda^{-1}(t)$$

$$\Lambda(t) = VAR\, e(t)$$

RICCATI EQUATION

$$P(t+1|t) = F\left[P(t|t-1) - P(t|t-1)C^T\Lambda^{-1}(t) \cdot CP(t|t-1)\right]F^T + Q$$

$$P(t_0|t_0-1) = P_0$$

PREDICTOR AS SYSTEM IN FEEDBACK

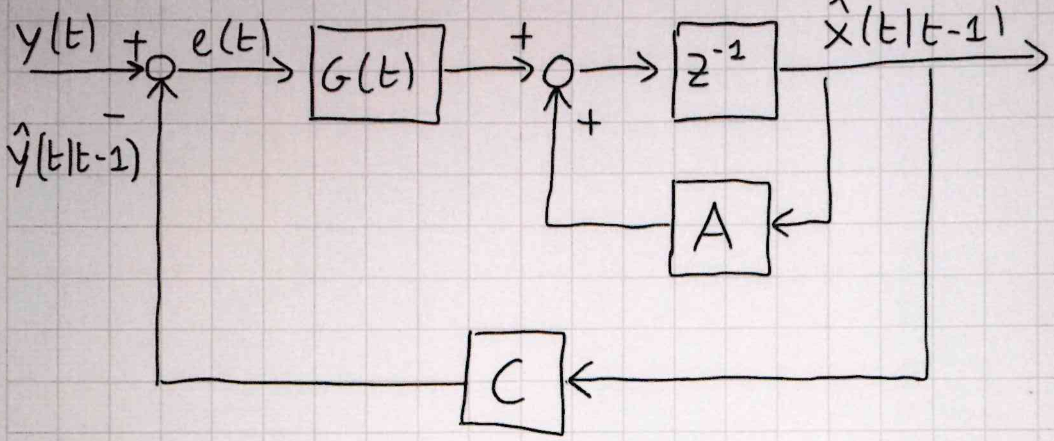
$$\hat{x}(t+1|t) = A\hat{x}(t|t-1) + G(t)\cdot e(t)$$

$$G(t) = K(t) + SR^{-1}$$

$$e(t) = y(t) - C\hat{x}(t|t-1)$$

$$G(t) = K(t) + S R^{-1}$$

$$e(t) = y(t) - C\,\hat{x}(t|t-1)$$



THE STATE IS THE PREDICTION
THE SYSTEM IS FED BY THE INNOVATION

THE STATE TRANSITION MATRIX

WE CAN ALSO WRITE

$$\hat{x}(t+1|t) = \underbrace{\left(F - K(t)C\right)}_{\Gamma(t)} \hat{x}(t|t-1) + \left(SR^{-1} + K(t)\right) y(t)$$

$$P(t+1|t) = \Gamma(t) P(t|t-1) \Gamma^T(t) + K(t) R K^T(t) + \tilde{Q}$$

$$P(t_0|t_0-1) = P_0$$

$$\Downarrow$$

$\Gamma(t)$ ALSO REGULATES THE ERROR DYNAMICS

$$\tilde{x}(t+1|t) = \Gamma(t)\,\tilde{x}(t|t-1) - K(t)\,w(t) + \tilde{v}(t)$$

IF $\Gamma(t)$ IS STABLE

- SMALL SENSITIVITY TO PARAMETER VARIATION
- SMALL SENSITIVITY TO MODELING ERRORS
- DISTURBANCE REJECTION

$$\Downarrow$$

FUNDAMENTAL ISSUES

$$\lim_{t \to +\infty} \Gamma(t)$$

FUNDAMENTAL ISSUES

$$\lim_{t \to +\infty} \Gamma(t)$$

(IF IT EXISTS)

WHICH IN TURN REQUIRES STUDY OF $K(t)$ AND $P(t|t-1)$ FOR $t \to +\infty$

FILTER EQUATIONS (COMPACT)

LET $S=0$ (NO NOISE CORRELATION)

$$\hat{x}(t+1|t+1) = A\hat{x}(t|t) + P(t+1|t)C^T\Lambda^{-1}(t+1)\left(y(t+1) - CA\hat{x}(t|t)\right)$$

$$\Lambda(t+1) = C\underbrace{P(t+1|t)}_{\text{FROM THE RICCATI}}C^T + R$$

$$P(t+1|t+1) = \left[I - P(t+1|t)C^T\Lambda^{-1}(t+1)C\right]P(t+1|t)$$

NOTE:

NOTE:

LET $L(t+1) = P(t+1|t)C^T \Delta^{-1}(t+1)$

THEN THE TRANSITION MATRIX FOR $\hat{x}(t|t)$ IS

$$\left[I - L(t+1)C\right]A$$

WHILE THAT FOR $\hat{x}(t|t-1)$ IS

$$A\left[I - L(t)C\right] \quad (\text{WITH } S=0)$$

SO THAT THEY SHARE THE SAME EIGENVALUES

EXAMPLES

1

$$x(t+1) = Ax(t) + v(t)$$

$$y(t) = w(t)$$

$$v(t) \perp w(t)$$

$$\hat{x}(t+1|t) = ?$$

SOL.

$$\hat{x}(t+1|t) = F\hat{x}(t|t-1) + SR^{-1}y(t)$$

SOL.

$$\hat{x}(t+1|t) = F\,\hat{x}(t|t-1) + SR^{-1}y(t)$$
$$+ FP(t|t-1)C^T\Lambda^{-1}(t)\left(y(t) - C\,\hat{x}(t|t-1)\right)$$

$F = A, \quad S = 0, \quad C = 0$

$$\Downarrow$$

$$\hat{x}(t+1|t) = A\,\hat{x}(t|t-1)$$

AND THIS IS OBVIOUS SINCE $y(t)$ DOES NOT CARRY ANY INFORMATION ON $x(t)$

2

$x(t+1) = v(t)$

$y(t) = Cx(t) + w(t)$

$v(t) \perp w(t)$

NOW $A = 0$, $S = 0$, $F = A = 0$

$$\Downarrow$$

$\hat{x}(t+1|t) = 0$

$$\Downarrow$$

$$\hat{x}(t+1|t) = 0$$

AND THIS IS DUE TO THE FACT THAT $y(t)$ CARRIES INFORMATION ON $x(t)$ BUT NOT ON $x(t+1)$ SINCE $\{x(t)\}$ IS WHITE NOISE (NO DYNAMICS)

3

$$x(t+1) = x(t)$$

$$y(t) = x(t) + w(t), \quad w(t) \sim N(0,1)$$

$$t_0 = 1$$

$$x(1) \sim N(x_1, 1), \quad x_1 = \hat{x}(1|0)$$

$$P_0 = P(1|0) = 1$$

RICCATI:

$$P(t+1|t) = F\left[P(t|t-1) - P(t|t-1)C^T\Delta^{-1}(t) \cdot CP(t|t-1)\right]F^T + \tilde{Q},$$

$$F = 1, \ \tilde{Q} = 0, \ C = 1$$

$$F = 1, \quad \tilde{Q} = 0, \quad C = 1$$

$$\Downarrow$$

$$P(t+1|t) = P(t|t-1) - \frac{P^2(t|t-1)}{1 + P(t|t-1)}$$

$$= \frac{P(t|t-1)}{1 + P(t|t-1)}$$

$$\Downarrow$$

$$P(1|0) = 1,$$

$$P(2|1) = 1/2,$$

$$P(3|2) = \frac{1/2}{1 + 1/2} = \frac{1}{3}$$

$$\vdots$$

$$P(t+1|t) = \frac{1}{t+1}$$

$$P(t+1|t) = \frac{1}{t+1}$$

$$\Downarrow$$

$$\hat{x}(t+1|t) = \hat{x}(t|t-1) + \frac{P(t|t-1)\left[y(t) - \hat{x}(t|t-1)\right]}{P(t|t-1)+1}$$

$$= \hat{x}(t|t-1) + \frac{y(t) - \hat{x}(t|t-1)}{t+1}$$

$$= \frac{x_1 + \sum_{i=1}^{t} y_i}{t+1}$$

(OBVIOUSLY THE MEAN IN VIEW OF THE SYSTEM EQUATIONS!)

$\Downarrow$

IN FACT:

$$\hat{x}(2|1) = x_1 + \frac{y(1) - x_1}{2} = \frac{x_1 + y(1)}{2}$$

$$\hat{x}(3|2) = \hat{x}(2|1) + \frac{y(2) - \hat{x}(2|1)}{3}$$

$$= \frac{x_1 + y(1)}{2} + \frac{y(2) - \frac{x_1 + y(1)}{2}}{3}$$

$$\hat{x}(3|2) = \hat{x}(2|1) + \frac{y(2) - \hat{x}(2|1)}{3}$$

$$= \frac{x_1 + y(1)}{2} + \frac{y(2) - \frac{x_1 + y(1)}{2}}{3}$$

$$= \frac{x_1 + y(1) + y(2)}{3}$$

SEMIDEFINITE POSITIVE MATRIX

$A \in \mathbb{R}^{n \times n}$

DEFINITION: $A = A^T \geq 0$

IF

$x^T A x \geq 0 \quad \forall\, x \in \mathbb{R}^n,$

$A = A^T > 0$

IF

$x^T A x > 0 \quad \forall\, x \in \mathbb{R}^n$

CONSEQUENCES

$$x^T A x > 0 \quad \forall x \in \mathbb{R}^n$$

CONSEQUENCES

- IF $A \geq B$, I.E. $A - B \geq 0$, THEN

$$[A]_{ii} \geq [B]_{ii} \quad \forall i$$

(BUT NOT $[A]_{ij} \geq [B]_{ij} \quad \forall i, j$)

- $B \in \mathbb{R}^{m \times n} \Rightarrow BB^T \in \mathbb{R}^{m \times m}$ AND $BB^T \geq 0$

- $A \geq 0$, $B \in \mathbb{R}^{m \times n} \Rightarrow BAB^T \in \mathbb{R}^{m \times m}$ AND $BAB^T \geq 0$

- $A > 0 \Rightarrow \exists\, \rho_1, \rho_2 > 0$ S.T.

$$\rho_1 I \leq A \leq \rho_2 I$$

TO PROVE THIS RECALL THAT

$$A = A^T > 0 \iff \underbrace{\lambda_i(A)}_{\text{EIGENVALUES}} > 0 \quad \forall i$$

$$\left(A = A^T \geq 0 \iff \lambda_i(A) \geq 0 \quad \forall i \right)$$

- $B \in \mathbb{R}^{m \times n} \Rightarrow BB^T \in \mathbb{R}^{m \times m}$

  AND $BB^T \geq 0$

- $A \geq 0,\ B \in \mathbb{R}^{m \times n} \Rightarrow BAB^T \in \mathbb{R}^{m \times m}$

  AND $BAB^T \geq 0$

- $A > 0 \Rightarrow \exists\ \rho_1, \rho_2 > 0$ S.T.

  $$\rho_1 I \leq A \leq \rho_2 I$$

TO PROVE THIS RECALL THAT

$$A = A^T > 0 \iff \underbrace{\lambda_i(A)}_{\text{EIGENVALUES}} > 0 \quad \forall i$$

$$\left(A = A^T \geq 0 \iff \lambda_i(A) \geq 0 \quad \forall i \right)$$

AND USE $\rho_1 = \min(\lambda_i)$

$\rho_2 = \max(\lambda_i)$

OPTIMAL ONE-STEP AHEAD PREDICTOR: SIMPLIFIED NOTATION

$$\Sigma : \begin{cases} x(t+1) = A x(t) + v(t) \\ y(t) = C x(t) + w(t) \end{cases}$$

$$v(t) \perp w(s) \quad \forall s, t$$

$$VAR \begin{bmatrix} v(t) \\ w(t) \end{bmatrix} = \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix}, \quad R > 0$$

$t_0 = 1$

$x(1) \sim (x_1, P(1))$

$x(1) \perp \{v(t), w(t)\}$

$E\, x(1) = x_1, \quad VAR\, x(1) = P(1)$

$\hat{x}(1|0) = x_1$

$P(1|0) =: P(1) = P_0$

$P(t|t-1) =: P(t)$

$P(t|t-1) =: P(t)$

$\parallel\!|$ $\qquad$ $\parallel\!|$

COVARIANCE MATRIX OF THE ONE-STEP AHEAD PREDICTOR ERROR

KALMAN PREDICTOR

$$\hat{\Sigma}: \begin{cases} \hat{x}(t+1|t) = A\hat{x}(t|t-1) + K(t)e(t) = (A-K(t)C)\hat{x}(t|t-1) + K(t)y(t) \\ \hat{y}(t+1|t) = C\hat{x}(t+1|t) \end{cases}$$

$K(t)$ = KALMAN GAIN

$$= AP(t)C^T(CP(t)C^T+R)^{-1}$$

WITH

DRE: $P(t+1) = AP(t)A^T + Q$

$$- AP(t)C^T(CP(t)C^T+R)^{-1}CP(t)A^T$$

$e(t)$ = INNOVATION

$$= y(t) - C\hat{x}(t|t-1)$$

< Note Fine

ASYMPTOTIC ISSUES

Σ IS TIME INVARIANT SUBJECT TO STATIONARY NOISES BUT THE PREDICTOR $\hat{\Sigma}$ IS TIME VARYING SINCE $K(t)$ VARIES IN TIME

MOTIVATIONS FOR STUDYING THE ASYMPTOTIC PREDICTOR

1. IF $\lim_{t \to +\infty} K(t) = \bar{K}$

I COULD ALREADY USE $\bar{K}$ OBTAINING A SUBOPTIMAL PREDICTOR CONVERGING TO THE OPTIMAL ONE

THIS MOTIVATES THE STUDY OF

$P(t)$ FOR $t \to +\infty$ (ASYMPTOTICS OF DRE)

IF $P(t) \to \bar{P}$, $\bar{P} = \bar{P}^T \geq 0$ AND

$\bar{P}$ HAS TO SOLVE THE

ALGEBRAIC RICCATI EQUATION (ARE)

$$P = APA^T + Q - APC^T(CPC^T+R)^{-1}CPA^T$$

QUESTIONS:

- WHEN $\lim_{t \to +\infty} P(t)$ EXISTS?

- DOES THE LIMIT $\bar{P}$ (IF IT EXISTS) DEPEND ON P_0?

- CAN ARE ADMIT MORE THAN ONE SOLUTION? WHICH IS THE RIGHT $\bar{P}$ TO WHICH $P(t)$ CONVERGES?

② BEYOND COMPUTATIONAL CONSIDERATIONS, STUDY OF

$$\lim_{t \to +\infty} P(t)$$

MAKES UNDERSTAND IF THE $\hat{\Sigma}$ CAN PREDICT THE STATE WITH AN ERROR VARIANCE $P(t)$ WHICH REMAINS SMALL (BOUNDED BY U)

③ ASYMPTOTIC BEHAVIOUR STUDY

③ ASYMPTOTIC BEHAVIOUR STUDY PERMITS TO ASSESS IF THE $\hat{\Sigma}$ TENDS TO A STABLE SYSTEM

$$\hat{\Sigma}_\infty : \hat{x}(t+1|t) = (A - \bar{K}C)\,\hat{x}(t|t-1) + \bar{K}y(t)$$

$$A - \bar{K}C = A - A\bar{P}C^T (C\bar{P}C^T + R)^{-1} C$$

THE PREDICTOR $\hat{\Sigma}_\infty$ IS STABLE IF $\max_i |\lambda_i(A - \bar{K}C)| < 1$

WE WILL SEE THAT

$$\lim_{t \to +\infty} P(t) = \bar{P} \not\Longrightarrow \hat{\Sigma}_\infty \text{ STABLE}$$

DEFINITION: A SOLUTION $\bar{P}$ OF THE ARE IS STABILIZING IF

$$\max_i |\lambda_i(A - \bar{K}C)| < 1$$

DRE CONVERGENCE
AND
PREDICTOR STABILITY

WE WILL CONSIDER 3 DIFFERENT SCENARIOS GIVING 3 DIFFERENT THEOREMS. WE WILL PROVE ONLY THE LAST ONE SINCE IT WILL CONTAIN THE FIRST TWO AS SPECIAL CASES. THIS WILL GIVE CRUCIAL INSIGHTS ON THE DRE (NON TRIVIAL MATRIX EQUATION SINCE IT IS HIGHLY NON LINEAR)

SCENARIO #1:
STABLE Σ

$$x(t+1) = Ax(t) + v(t), \quad |\lambda_i| < 1 \quad \forall i$$

SCENARIO #1:
STABLE Σ

$x(t+1) = A x(t) + v(t), \quad |\lambda_i| < 1 \quad \forall i$

WE KNOW THAT $x(t)$ TENDS TO BECOME A STATIONARY PROCESS WITH COVARIANCE $\bar{\Sigma}$ THAT IS THE UNIQUE SOLUTION OF

$$\bar{\Sigma} = A \bar{\Sigma} A^T + Q$$

IS IT POSSIBLE THAT $P(t)$ DIVERGES?

NO: IN THE WORST CASE NO DATA $y(t)$ ARE AVAILABLE, THUS $K(t) = 0$ AND

$$\hat{x}(t+1|t) = A \hat{x}(t|t-1)$$

$$\hat{x}(t+1|t) \to 0 = \lim_{t \to +\infty} E\,x(t)$$

SO THAT

$P(t) \to \bar{\Sigma}$ SINCE

AND

$$\hat{x}(t+1|t) = A\,\hat{x}(t|t-1)$$

$$\hat{x}(t+1|t) \to 0 = \lim_{t\to+\infty} E\,x(t)$$

SO THAT

$P(t) \to \bar{\Sigma}$ SINCE

$$\mathrm{VAR}\left(x(t+1) - \hat{x}(t+1|t)\right) \underset{\substack{\text{LARGE}\\ t}}{\simeq} \mathrm{VAR}\,x(t+1)$$

$$\simeq \bar{\Sigma}$$

CAN $\hat{\Sigma}$ BE UNSTABLE?

Σ STABLE $\Rightarrow$ x, y STATIONARY

AND $y(t)$ IS THE INPUT

TO THE PREDICTOR $\hat{\Sigma}$

$$\hat{\Sigma}_\infty : \hat{x}(t+1|t) = \underbrace{(A - \bar{K}C)}_{\text{IF UNSTABLE, } \hat{x} \text{ NON STATIONARY,}}\hat{x}(t|t-1) + \bar{K}\,y(t)$$

IT SEEMS ABSURD TO PREDICT A STATIONARY PROCESS x WITH A NON STATIONARY $\hat{x}$

FIRST CONVERGENCE THEOREM

STABLE $\Sigma \Longrightarrow$ 1) $\forall\ P_0 = P_0^T \geq 0$

$$\lim_{t \to +\infty} P(t) = \bar{P}$$

2) $\hat{\Sigma}_\infty$ IS STABLE

3) $\bar{P}$ IS THE ONLY $\bar{P} = \bar{P}^T \geq 0$ WHICH SOLVES THE ARE

NOTE: REMOVING THE ASSUMPTION $R > 0$, $A - \bar{K}C$ COULD HAVE EIGENVALUES λ_i ON THE UNIT CIRCLE ($S_y(e^{j\omega}) \geq 0$ ONLY)

RELATIONSHIP WITH WIENER

IF Σ IS STABLE, IF $\tau = t - t_0$:

$$\lim_{\tau \to \infty} \| \underbrace{\hat{x}(\tau|\tau-1)}_{\text{KALMAN}} - \underbrace{\hat{x}_\infty(\tau|\tau-1)}_{\text{WIENER-KOLMOGOROV}} \| = 0$$

$t \to +\infty$

2) $\hat{\Sigma}_\infty$ IS STABLE

3) $\bar{P}$ IS THE ONLY $\bar{P} = \bar{P}^T \geq 0$ WHICH SOLVES THE ARE

NOTE: REMOVING THE ASSUMPTION $R>0$, $A-\bar{K}C$ COULD HAVE EIGENVALUES λ_i ON THE UNIT CIRCLE ($S_y(e^{j\omega}) \geq 0$ ONLY)

RELATIONSHIP WITH WIENER

IF Σ IS STABLE, IF $\tau = t - t_0$:

$$\lim_{\tau \to \infty} \| \underbrace{\hat{x}(\tau|\tau-1)}_{\text{KALMAN PREDICTOR}} - \underbrace{\hat{x}_\infty(\tau|\tau-1)}_{\text{WIENER-KOLMOGOROV PREDICTOR WHICH USES MEASUREMENTS FROM } t_0 = -\infty} \| = 0$$

$\|x\|$ = TRACE OF VAR x

SCENARIO #2:

OBSERVABILITY AND CONTROLLABILITY

STABILITY IS SUFFICIENT BUT NOT NECESSARY FOR PREDICTOR CONVERGENCE!

- POSSIBLE CONVERGENCE EVEN WITH UNSTABLE Σ
- $P(t)$ COULD SATISFY $P(t) \leq U \ \forall t$ EVEN IF $\hat{\Sigma}_{\infty}$ IS UNSTABLE

EXAMPLE 1

$x(t+1) = x(t)$

$y(t) = w(t)$

$x(1) = x_1, \quad P(1) = 0$

$$\Downarrow$$

$\hat{x}(t+1|t) = \hat{x}(t|t-1) = x_1$

NECESSARY FOR PREDICTOR CONVERGENCE!

- POSSIBLE CONVERGENCE EVEN WITH UNSTABLE Σ

- $P(t)$ COULD SATISFY $P(t) \leq U \ \forall t$ EVEN IF $\hat{\Sigma}_{\infty}$ IS UNSTABLE

EXAMPLE 1

$$x(t+1) = x(t)$$

$$y(t) = w(t)$$

$$x(1) = x_1, \quad P(1) = 0$$

$$\Downarrow$$

$$\hat{x}(t+1|t) = \hat{x}(t|t-1) = x_1$$

AND $P(t) = 0 \ \forall t$

SO, $P(t)$ CONVERGES, Σ AND $\hat{\Sigma}$ ARE UNSTABLE

EXAMPLE 2

$$\Sigma \begin{cases} x(t+1) = \alpha x(t) + v(t), & v(t) \sim (0, \beta^2) \\ y(t) = \gamma x(t) + w(t), & w(t) \sim (0, 1) \end{cases}$$

$$w \perp v$$

$|\alpha| > 1$

CASE A: $\beta \neq 0$, $\gamma \neq 0$

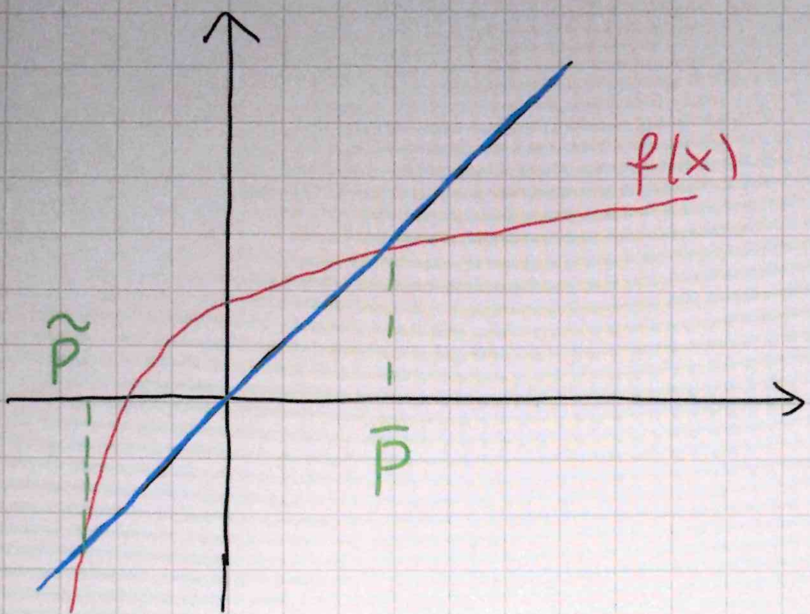
$x(t)$ DIVERGES IN A PROBABILISTIC SENSE ($\mathrm{VAR}\, x(t) \to +\infty$), BUT LET US STUDY THE DRE

$$P(t+1) = \beta^2 + \frac{\alpha^2 P(t)}{1 + \gamma^2 P(t)}$$

IT IS USEFUL TO PLOT

$$f(x) = \beta^2 + \frac{\alpha^2 x}{1 + \gamma^2 x}$$

WHOSE $\cap$ WITH $y = x$ GIVES ARE SOLUTION

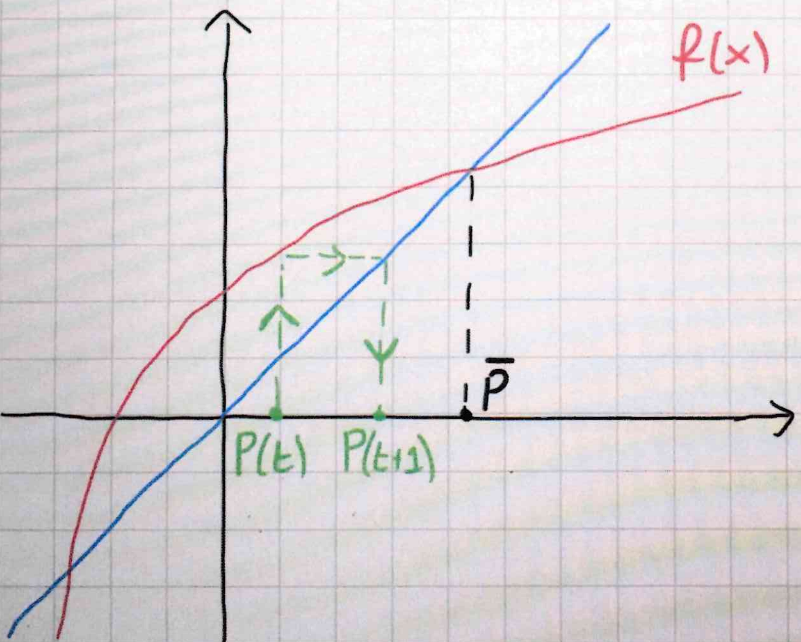


$\bar{P}$ IS THE ONLY SOLUTION (OF INTEREST SINCE ≥ 0) OF THE ARE

LET US NOW STUDY

THE TWO CASES FOR $P(1)$

$P_1 < \bar{P}$



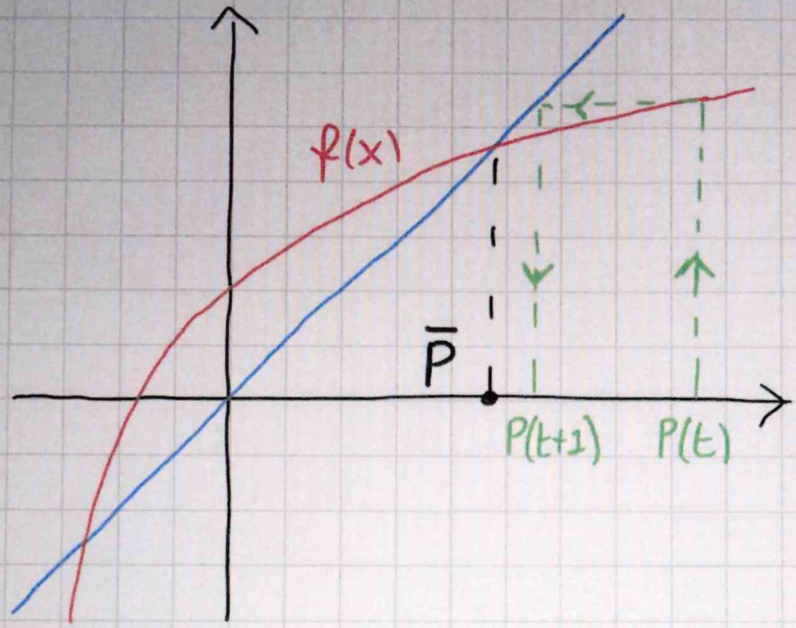
$f(x)$ STRICTLY MONOTONE INCREASING AND $f(x) \leq \bar{P}$ $\forall x \leq \bar{P}$

$\Downarrow$

CONVERGENCE TO $\bar{P}$

$P_1 > \bar{P}$

$P_1 > \bar{P}$



f(x) STRICTLY MONOTONE INCREASING AND $f(x) \geq \bar{P}$ $\forall x \geq \bar{P}$

$\Downarrow$

CONVERGENCE TO $\bar{P}$

IN CONCLUSION: $\forall P_1 \geq 0$ THERE IS CONVERGENCE TO $\bar{P}$

CHECK ALSO THAT $\bar{P}$ IS STABILIZING SINCE $|\alpha - \bar{k}\gamma| < 1$

NOTE THAT THE PREDICTOR IS FED WITH AN "UNSTABLE" $y(t)$ AND ALSO $\hat{x}(t+1|t)$ DIVERGES BUT THE PREDICTION ERROR $P(t+1)$ REMAINS

WELL BOUNDED. IT IS THE NON STATIONARITY OF $y(t)$ THAT ALLOWS THE STABLE PREDICTOR TO FOLLOW THE NONSTATIONARY PROCESS $x(t)$

CASE B: $\beta \neq 0, \gamma = 0$

DRE BECOMES

$$P(t+1) = \beta^2 + \alpha^2 P(t)$$

$$\Downarrow$$

$P(t)$ DIVERGES

$K(t)$ IS PROPORTIONAL TO γ

$$\Downarrow$$

$$K(t) = 0 \quad \forall t$$

SO, $\hat{\Sigma}$ IS IN OPEN LOOP:

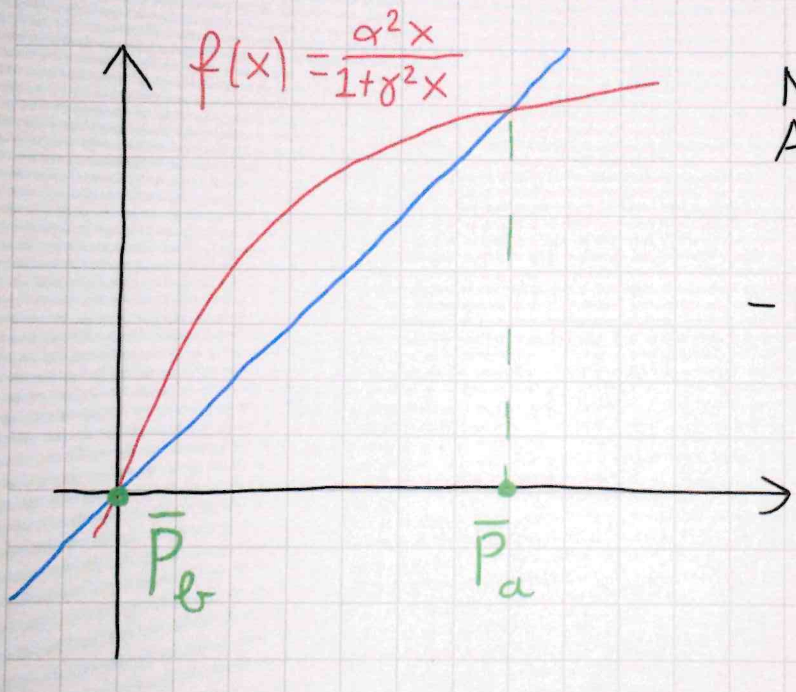
$$\hat{x}(t+1|t) = \alpha \hat{x}(t|t-1)$$

AND IS UNSTABLE

CASE C: $\beta=0, \gamma\neq0$

SO, NOW

$$x(t+1)=\alpha x(t)$$
$$y(t)=\gamma x(t)+w(t) \Rightarrow P(t+1)=\frac{\alpha^2 P(t)}{1+\gamma^2 P(t)}$$



NOW $f(0)=0$ AND ARE HAS TWO SOLUTIONS:

- $P_1=0 \Rightarrow P(t)=0 \ \forall t$

  AND $\hat{x}(t+1|t)=\alpha\hat{x}(t|t-1)$

  $\bar{P}_b$ IS NOT STABILIZING BUT THE ERROR IS NULL

- $P_1\neq0 \Rightarrow P(t)\to\bar{P}_a$

  AS IN THE CASE A

  AND $\bar{P}_a$ IS STABILIZING

SO, IN CASE C TWO SOLUTIONS FOR ARE, ONE STABILIZING

DISCUSSION ON EXAMPLE #2

FIRST, IT POINTS OUT THE IMPORTANCE OF $y(t)$. Σ IS UNSTABLE, BUT $P(t)$ DIVERGES ONLY IF $\gamma = 0$ (NO INFO FROM THE OUTPUT)

IN GENERAL, CONVERGENCE OF $P(t)$ REQUIRES $y(t)$ TO CARRY INFORMATION (EVEN IF TRIVIAL CASES SUCH AS EXAMPLE #1 EXIST WHERE $\gamma = 0$ AND $P(t) = 0$!)

CASE A WITH $\gamma \neq 0, \beta \neq 0$ ENSURES:

- UNIQUE ARE SOLUTION
- $P(t) \to \bar{P} \quad \forall \; P_1 \geq 0$
- STABILIZING $\bar{P}$

WHICH STRUCTURAL PROPERTIES DOES Σ HAVE IN CASE A?

$\gamma \neq 0 \Rightarrow$ OBSERVABLE Σ

OBSERVABILITY

$$x(t+1) = A\,x(t) \qquad\qquad y(1) = C\,x(1)$$

$$y(t) = C\,x(t) \qquad \Longrightarrow \qquad y(2) = CA\,x(1)$$

$$\vdots$$

$$y(n) = CA^{n-1}x(1)$$

$$n = \text{DIM } x(t)$$

Σ (OR THE COUPLE A,C) IS OBSERVABLE IF THERE ARE NO DISTINCT $x(1)$ THAT GENERATE THE SAME (FREE) OUTPUTS

OBSERVABLE Σ

$$\Updownarrow$$

$$\text{RANK}\left[C^T \,\middle|\, A^TC^T \,\middle|\, \ldots \,\middle|\, (A^T)^{n-1}C\right] = n$$

$$\Updownarrow$$

$$\text{RANK}\left[C^T \,\vdots\, A^T C^T \,\vdots\, \ldots \,\vdots\, (A^T)^{n-1} C\right] = n$$

$$\Updownarrow$$

$$\text{RANK}\left[sI - A^T \,\vdots\, C^T\right] = n$$

$\forall s \in \mathbb{C}$ (IT SUFFICES CHECKING $s = \lambda_k$, $k = 1, \ldots, n$)

THEOREM:

(A, C) OBSERVABLE $\Longrightarrow$ ARE ADMITS AT LEAST ONE SOLUTION $\bar{P} = \bar{P}^T \geq 0$ $\blacksquare$

$\beta \neq 0 \Rightarrow$ CONTROLLABILITY

EVEN IF (A, C) IS OBSERVABLE, ARE CAN HAVE MORE THAN ONE SOLUTION (RECALL THE CASE $\gamma \neq 0, \beta = 0$)

IF INSTEAD $\gamma \neq 0, \beta \neq 0$ WE OBTAINED

< Note

EVEN IF (A,C) IS OBSERVABLE, ARE CAN HAVE MORE THAN ONE SOLUTION (RECALL THE CASE $\gamma \neq 0, \beta = 0$)

IF INSTEAD $\gamma \neq 0, \beta \neq 0$, WE OBTAINED A UNIQUE $\bar{P}$.

$\beta \neq 0$ ENSURES THAT $x(t)$ IS INFLUENCED BY THE TRANSITION NOISE.

ARE HAS ONLY ONE SOLUTION IF $v(t) = Bn(t)$ REACHES ALL THE COMPONENTS OF $x(t)$. THIS IS EMBEDDED IN THE CONTROLLABILITY CONCEPT.

CONTROLLABILITY

$$x(t+1) = Ax(t) + Bn(t)$$

NOW WE HAVE TO CONSIDER THE

CONTROLLABILITY

$x(t+1) = Ax(t) + Bu(t)$

NOW WE HAVE TO CONSIDER THE COUPLE (A, B) OR, EQUIVALENTLY, (A, Q) OR $(A, Q^{1/2})$ WITH $Q = BB^T$.

$u(t)$ REACHES ALL THE STATE SPACE IF Σ IS CONTROLLABLE. ONE HAS

$$\text{CONTROLLABLE } \Sigma$$

$$\Updownarrow$$

$$\text{RANK} \left[B \,\vdots\, AB \,\vdots\, \cdots \,\vdots\, A^{n-1}B \right] = n$$

$$\Updownarrow$$

$$\text{RANK} \left[sI - A \,\vdots\, B \right] = n \quad \forall s \in \mathbb{C}$$

$$(s = \lambda_k,\ k = 1, \dots, n)$$

NOTE:

COUPLE (A, B) OR, EQUIVALENTLY, (A, Q) OR $(A, Q^{1/2})$ WITH $Q = BB^T$.

$x(t)$ REACHES ALL THE STATE SPACE IF Σ IS CONTROLLABLE. ONE HAS

$$\text{CONTROLLABLE } \Sigma$$

$$\Updownarrow$$

$$\text{RANK}\left[B \,\vdots\, AB \,\vdots\, \cdots \,\vdots\, A^{n-1}B\right] = n$$

$$\Updownarrow$$

$$\text{RANK}\left[sI - A \,\vdots\, B\right] = n \quad \forall s \in \mathbb{C}$$

$$(s = \lambda_k,\ k = 1, \ldots, n)$$

NOTE:

$$\underset{(A, C)}{\text{OBSERVABLE}} \iff \underset{(A^T, C^T)}{\text{CONTROLLABLE}}$$

SECOND DRE CONVERGENCE THEOREM

(A, C) OBSERVABLE, (A, B) CONTROLLABLE $\Longrightarrow$

1) ARE HAS A UNIQUE SOLUTION $\bar{P} \geq 0$
2) $P(t) \to \bar{P} \quad \forall P_0 \geq 0$
3) $\bar{P} > 0$ AND STABILIZING

NOTE a: LIKE 1^{ST} TH. EXCEPT $\bar{P} > 0$

NOTE b: 2) => 1). IN FACT, IF

$$P(t) \to \bar{P} \quad \forall P_0 \geq 0,$$

ASSUME $\exists$ A, B SOLUTIONS OF ARE, WITH $A \geq 0$, $B \geq 0$, $A \neq B$.

BUT IF $P_0 = A$, $P(t) = A \quad \forall t$

IF $P_0 = B$, $P(t) = B \quad \forall t$

AND THIS CONTRADICTS 2), I.E.

AND THIS CONTRADICTS 2), I.E.

IT WOULD NOT BE TRUE THAT

$P(t) \rightarrow$ SAME MATRIX $\quad \forall \; P_0 \geq 0$

THE 1^{ST} AND 2^{nd} THEOREM DO NOT GIVE NECESSARY AND SUFFICIENT CONDITIONS

STABLE Σ $\Longrightarrow$

(A, C) OBS.
(A, B) CONTR. $\Longrightarrow$

WE WILL SEE THAT BOTH OF THESE CONDITIONS (TAKEN ALONE) IMPLY A KEY STRUCTURE OF Σ

REPRESENTATION OF

REPRESENTATION OF NON CONTROLLABLE Σ

$$\dim\left[B \ AB \ldots A^{n-1}B\right] = \rho < n$$

$\exists \ T$ INVERTIBLE DEFINING

$x = T\bar{x}$, $\bar{x}$ = STATE IN THE NEW COORDINATES

S.T. Σ BECOMES (JUST A, B TO SEE):

$$\bar{A} = T^{-1}AT = \begin{array}{cc} & \begin{array}{cc} \rho & \phantom{\bar{A}_{12}} \end{array} \\ \rho & \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} \\ 0 & \bar{A}_{22} \end{bmatrix} \end{array},$$

$$\bar{B} = T^{-1}B = \begin{array}{cc} \begin{bmatrix} \bar{B}_1 \\ 0 \end{bmatrix} & \begin{array}{c} \rho \\ \end{array} \end{array}$$

AND THE SYSTEM IS WRITTEN

AND THE SYSTEM IS WRITTEN IN CONTROLLABLE FORM.

ONE HAS:

a) $(\bar{A}_{11}, \bar{B}_1)$ IS CONTROLLABLE,

$(\bar{A}_{22}, 0)$ IS THE NON CONTROLLABLE SUBSYSTEM

b) $\lambda_i(\bar{A}_{22})$ ARE EXACTLY THOSE $s \in \mathbb{C}$ S.T.

$$\text{RANK}\begin{bmatrix} A - sI & B \end{bmatrix} < n$$

(THE λ_i WHICH MAKE THE PBH TEST FAIL)

c) $\bar{x}_1(t+1) = \bar{A}_{11}\bar{x}_1(t) + \bar{A}_{12}\bar{x}_2(t) + \bar{B}_1 u(t)$

$\bar{x}_2(t+1) = \bar{A}_{22}\bar{x}_2(t)$

DYNAMICS OF $\bar{x}_2$ ARE NOT INFLUENCED BY THE INPUT

STABILIZABILITY

STABILIZABILITY

Σ IS SAID TO BE STABILIZABLE

$$|\lambda_i(\bar{A}_{22})| < 1 \quad \forall i$$

$$\Updownarrow$$

$$\exists \; K \; \text{S.T.} \; |\lambda_i(A+BK)| < 1 \quad \forall i$$

REPRESENTATION OF NON OBSERVABLE Σ

$$\text{RANK} \left[C^T \; A^T C \ldots (A^T)^{n-1} C^T \right] = \rho < n$$

$\exists$ T INVERTIBLE WITH

$$x = T\bar{x} \quad \text{S.T.}$$

$$\bar{A} = T^{-1} A T = \rho\updownarrow \begin{bmatrix} \bar{A}_{11} & 0 \\ \bar{A}_{21} & \bar{A}_{22} \end{bmatrix},$$

$$\bar{B} = T^{-1}B = \begin{bmatrix} \bar{B}_1 \\ \bar{B}_2 \end{bmatrix} \begin{matrix} \updownarrow \rho \\ \end{matrix},$$

$$\bar{C} = CT = [\underbrace{\bar{C}_1}_{\rho} \ \ 0]$$

AND ONE HAS

a) $(\bar{A}_{11}, \bar{C}_1)$ OBSERVABLE,

$(\bar{A}_{11}, \bar{B}_1, \bar{C}_1)$ OBSERVABLE SUBSYSTEM

b) $\lambda_i(\bar{A}_{22})$ ARE ALL THE $s \in \mathbb{C}$ S.T.

$$\text{RANK}\,[A^T - sI \ \ C^T] < n$$

c) $(\bar{A}_{22}, \bar{B}_2, 0)$ IS THE NON OBS. SUBSYSTEM.

$$\bar{x}_1(t+1) = \bar{A}_{11}x_1(t) + \bar{B}_1 u(t)$$

$$\bar{x}_2(t+1) = \bar{A}_{21}x_1(t) + \bar{A}_{22}x_2(t) + \bar{B}_2 u(t)$$

$y(t) = \bar{C}_1 x_1(t)$, $\bar{x}_2(t)$ DOES NOT INFLUENCE $y(t)$

< Note

DETECTABILITY

Σ IS SAID TO BE DETECTABLE

IF $|\lambda_i(\bar{A}_{22})| < 1 \quad \forall i$

$$\Updownarrow$$

$\exists K$ S.T.

$|\lambda_i(A+KC)| < 1 \quad \forall i$

A SIMPLE OBSERVATION

STABLE Σ $\Longrightarrow$ STABILIZABLE AND DETECTABLE Σ

(IN FACT $|\lambda_i(A)| < 1 \quad \forall i$)

ANOTHER SIMPLE OBSERVATION

Σ DETECTABLE Σ

(IN FACT $|\lambda_i(A)| < 1 \ \forall i$)

ANOTHER SIMPLE OBSERVATION

(A, C) OBS.
(A, B) CONTR. $\Longrightarrow$ STABILIZABLE AND DETECTABLE Σ

(IN FACT THERE ARE NO UNSTABLE λ_i IN THE NON OBS. AND NON CONTR. SUBSYSTEMS SINCE SUCH SUBSYSTEMS DO NOT EXIST!)

IS MAYBE STABILIZABILITY AND DETECTABILITY THE CONDITION WEAKER THAN STABILITY AND THAN (OBS. + CONTR.) WHICH IS KEY TO CHARACTERIZE THE ASYMPTOTICS OF $\hat{\Sigma}$?

GENERAL CONVERGENCE THEOREM

CONSIDER

$$x(t+1) = A x(t) + v(t), \quad \text{VAR } v(t) = Q$$

$$y(t) = C x(t) + w(t), \quad \text{VAR } w(t) = R > 0$$

$$v \perp w$$

IF $Q = BB^T$, ONE HAS

(A, C) DETECTABLE
(A, B) STABILIZABLE
$\iff$
1) $P(t) \to \bar{P} \ \forall P_0$
2) $\exists ! \ \bar{P} = \bar{P}^T \geq 0$ ARE SOLUTION
3) $\bar{P}$ IS STABILIZING

WE WILL PROVE $\Rightarrow$ IN SIX POINTS

POINT 1:

DETECTABILITY $\Rightarrow \forall P_0 \ \exists \ U = U^T \geq 0$ S.T.

DETECTABILITY $\Longrightarrow$ $\forall P_0 \; \exists \; U = U^T \geq 0$ S.T.

$$P(t) \leq U \quad \forall t$$

PROOF: DETECTABILITY IMPLIES THAT EXISTS $\tilde{K}$ S.T. $A - \tilde{K}C$ IS STABLE.

WE USE A PREDICTOR WITH GAIN $\tilde{K}$:

$$x(t+1|t) = Ax(t|t-1) + \tilde{K}\left(y(t) - Cx(t|t-1)\right)$$

$$\Downarrow$$

$$x(t+1) - x(t+1|t) = (A - \tilde{K}C)\left(x(t) - x(t|t-1)\right) + v(t) - \tilde{K}w(t)$$

SO THE ERROR IS THE STATE OF A STABLE SYSTEM FED WITH A STATIONARY INPUT $v(t) - \tilde{K}w(t)$

$$\Downarrow$$

$\tilde{P}(t) := \mathrm{VAR}\left(x(t+1) - x(t+1|t)\right)$ TENDS TO A FINITE MATRIX. SO THERE EXISTS U S.T.

$$\tilde{P}(t) \leq U$$

$$\Downarrow$$

$\tilde{P}(t) := VAR\left(x(t+1) - x(t+1|t)\right)$ TENDS TO A FINITE MATRIX. SO THERE EXISTS U S.T.

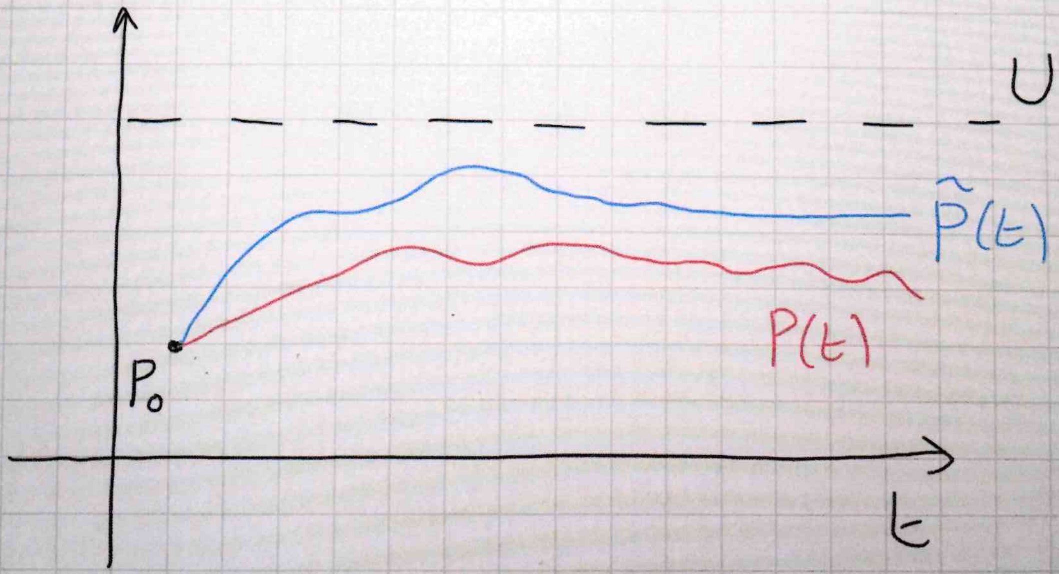
$$\tilde{P}(t) \leq U$$

WHERE $\tilde{P}(t)$ AND U DEPEND ON P_0.

FOR ANY P_0, THE DRE GENERATES VARIANCE ERRORS $P(t)$ S.T.

$$P(t) \leq \tilde{P}(t) \leq U \quad \forall t$$

BY THE OPTIMALITY OF THE KALMAN FILTER.



NOTE: WE CANNOT SAY THAT $P(t)$ CONVERGES TO SOMETHING AS $\widetilde{P}(t)$ DOES. $\widetilde{K}$ IS SUBOPTIMAL BUT STABILIZES BY CONSTRUCTION

$$\left(|\lambda_i(A-\widetilde{K}C)| < 1 \ \forall i \right).$$

$K(t)$ IS OPTIMAL BUT IT IS NOT KNOWN IF IT CONVERGES TO A $\bar{K}$ (WHICH COULD STABILIZE $A-\bar{K}C$).

POINT 2:

GIVEN TWO INITIAL CONDITIONS

$$P'(1) = M \geq 0, \quad P''(1) = N \geq 0,$$

THEN

$$M \geq N \Rightarrow P'(t) \geq P''(t) \quad \forall t$$

DRE SOLUTIONS WITH I.C. $P'(1)$, $P''(1)$, RESPECTIVELY

PROOF: USEFUL TO WRITE THE R.H.S. OF THE DRE USING

PROOF: USEFUL TO WRITE THE R.H.S. OF THE DRE USING

$$\Delta(P,K) = (A-KC)P(A-KC)^T + Q + KRK^T \quad \circledast$$

Δ: 2 MATRICES $\rightarrow$ 1 MATRIX

IF WE USE

$K = K(P) = APC^T(CPC^T+R)^{-1}$,

ONE OBTAINS THE DRE

$$\Pi(P) = \Delta(P, K(P))$$

Π = DRE: 1 MATRIX $\rightarrow$ 1 MATRIX

$P(t+1) = \Pi(P(t))$ DRE

$P = \Pi(P)$ ARE

LET US NOW PROVE TWO INEQUALITIES

a) $M \geq N \Rightarrow \underbrace{\Delta(M, K(M))}_{=\Pi(M)} \geq \underbrace{\Delta(N, K(M))}_{\neq \Pi(N)}$

THIS DERIVES FROM $\circledast$. WE CAN REPLACE THERE $(M, K(M))$ AND THEN $(N, K(M))$ AND ONE EASILY SEES

< Note

Fine

$(N, K(M))$ AND ONE EASILY SEES THAT $\Delta(M, K(M)) \geq \Delta(N, K(M))$ REDUCES TO

$$(A-KC)M(A-KC)^T \geq (A-KC)N(A-KC)^T$$

$(K = K(M)$ EVERYWHERE$)$

WHICH HOLDS TRUE IF $M \geq N$

$(\text{SINCE } M \geq N \Leftrightarrow x^T M x \geq x^T N x \ \forall x)$

b) $\underbrace{\Delta(N, K(M))}_{\neq \Pi(N)} \geq \underbrace{\Delta(N, K(N))}_{= \Pi(N)}$

THIS INEQUALITY HOLDS SINCE EACH Δ DESRIBES THE ERROR COVARIANCES OF TWO FILTERS WITH $P(1) = N$ AND ON THE RIGHT THE OPTIMAL KALMAN GAIN $K(N)$ IS USED.

b) $\underbrace{\Delta(N, K(M))}_{\neq \Pi(N)} \geq \underbrace{\Delta(N, K(N))}_{= \Pi(N)}$

THIS INEQUALITY HOLDS SINCE EACH Δ DESRIBES THE ERROR COVARIANCES OF TWO FILTERS WITH $P(1) = N$ AND ON THE RIGHT THE OPTIMAL KALMAN GAIN $K(N)$ IS USED.

COMBINING a) AND b) :

$$\Delta(M, K(M)) \underset{a}{\geq} \Delta(N, K(M)) \underset{b}{\geq} \Delta(N, K(N))$$

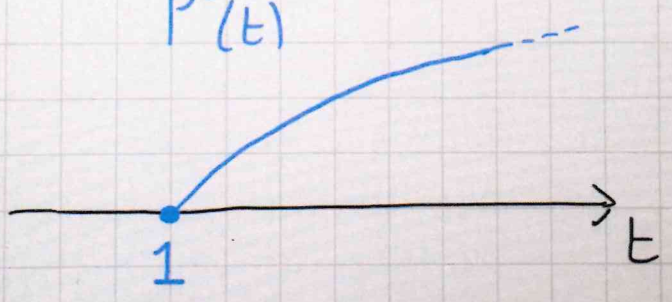
AND THIS CONCLUDES THE PROOF $\blacksquare$

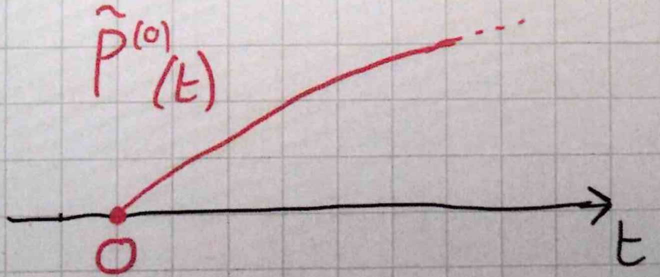
POINT 3:

THE DRE SOLUTION WITH INITIAL CONDITION $P(1)=0$ IS MONOTONICALLY INCREASING IN MATRIX SENSE, I.E.

$$P^{(0)}(t+1) \geq P^{\circ}(t)$$

PROOF: THE MAIN IDEA IS TO INTRODUCE TWO FILTERS WITH A TIME DELAY AND THEN TO EXPLOIT POINT 2





SO, CONSIDER A FILTER WITH NULL I.C. AT $t_0 = 0$ IN PLACE OF $t_0 = 1$.

$P(0) = 0$ AND $\tilde{P}^{(0)}(t)$ IS THE RELATED DRE SOLUTION. ONE HAS

$$\tilde{P}^{(0)}(t) = P^{(0)}(t+1) \quad \alpha$$

IT IS OBVIOUS THAT

$$\tilde{P}^{(0)}(1) \geq P^{(0)}(1) \quad \text{SINCE } P^{(0)}(1) = 0$$

BUT THEN, USING POINT 2,

$$\begin{array}{ccc} \Pi\left(\tilde{P}^{(0)}(t)\right) & \geq & \Pi\left(P^{(0)}(t)\right) \quad \forall t \\ ||| & & ||| \\ \tilde{P}^{(0)}(t+1) & \geq & P^{(0)}(t+1) \end{array}$$

$$\Big\Downarrow \text{ USING } \alpha$$

$$P^{(0)}(t+2) \geq P^{(0)}(t+1)$$

POINT 4:

THE $P^{(0)}(t)$ SEEN BEFORE CONVERGES TO A MATRIX $\bar{P}$, I.E.

$$\lim_{t \to +\infty} P^{(0)}(t) = \bar{P}$$

PROOF: IT DERIVES FROM THE FACT THAT $P^{(0)}(t)$ IS MONOTONICALLY NON DECREASING AND BOUNDED BY U.

IN FACT:

LET $P_i = P_i^T \geq 0$ A SEQUENCE OF MATRICES S.T.

$$P_1 \leq P_2 \leq \dots \leq U.$$

LET US PROVE THAT

$$\lim_{i \to +\infty} P_i = \bar{P}$$

BY ASSUMPTION ONE HAS

BY ASSUMPTION ONE HAS

$$x^T P_i x \leq x^T P_{i+1} x \leq x^T U x \qquad \forall x, i$$

SO, $\{ x^T P_i x \}$ IS A SEQUENCE OF BOUNDED AND MONOTONICALLY NON DECREASING SCALARS. SO, IT CONVERGES. γ

LET

$$e_j = [0 \ 0 \ \ldots \ \underbrace{1}_{j\text{-TH ELEMENT}} \ 0 \ldots 0]^T$$

IT IS EASY TO SEE THAT

$$\underbrace{2\, e_a P_i e_b}_{[P_i]_{ab}} = \underbrace{(e_a + e_b)^T P_i (e_a + e_b)}_{\substack{\text{CONVERGES} \\ \text{BY } \gamma \text{ WITH } x = e_a + e_b}}$$

$$- \underbrace{e_a^T P_i e_a} - \underbrace{e_b^T P_i e_b}$$

IT IS EASY TO SEE THAT

$$\underbrace{2\, e_a P_i e_b}_{[P_i]_{ab}} = \underbrace{(e_a + e_b)^T P_i (e_a + e_b)}_{\substack{\text{CONVERGES} \\ \text{BY } \gamma \text{ WITH } x = e_a + e_b}}$$

$$- \underbrace{e_a^T P_i e_a}_{\substack{\text{CONVERGES} \\ \text{BY } \gamma \text{ WITH} \\ x = e_a}} - \underbrace{e_b^T P_i e_b}_{\substack{\text{CONVERGES} \\ \text{BY } \gamma \text{ WITH} \\ x = e_b}}$$

SO

$\lim_{i \to \infty} [P_i]_{ab}$ CONVERGES TOWARDS

A LIMIT IDENTIFYING

$$[\bar{P}]_{ab}$$

$$\Downarrow$$

$$P_i \longrightarrow \bar{P}$$

$\blacksquare$

COMMENT AFTER POINTS 1-4

WE HAVE JUST USED DETECTABILITY.

WE HAVE PROVED THAT

(A, C) DETECTABLE $\Longrightarrow$

1) IF $P(1)=0$, DRE SOLUTION $P(t)$ CONVERGES TO A MATRIX WE DENOTE WITH $\bar{P} \geq 0$

2) SO, THERE EXISTS AT LEAST A SOLUTION TO THE ARE $(\bar{P} \geq 0)$

3) ANY SOLUTION OF THE DRE (WITH ANY $P(1)$) IS BOUNDED

NOTE ON EIGENVECTORS AND EIGENVALUES

NOTE ON EIGENVECTORS AND EIGENVALUES

GIVEN $A \in \mathbb{R}^{n \times n}$, $v \in \mathbb{C}^n$, IF

$$Av = \lambda v, \quad \lambda \in \mathbb{C},$$

THEN v IS A RIGHT EIGENVECTOR OF A. INSTEAD $u \in \mathbb{C}^n$ S.T.

$$u^T A = \lambda u^T, \quad \lambda \in \mathbb{C}$$

IS A LEFT EIGENVECTOR

(ALSO SATISFYING

$$A^T u = \lambda u)$$

RIGHT AND LEFT EIGENVECTORS ARE IN GENERAL DIFFERENT BUT SHARE THE SAME EIGENVALUES.

POINT 5:

–

POINT 5:

$\bar{P}$ (AS GIVEN BY POINT 4)
IS STABILIZING

PROOF: $\bar{P}$ SATISFIES THE ARE, I.E.

$$\bar{P} = (A-\bar{K}C)\bar{P}(A-\bar{K}C)^T + \bar{K}R\bar{K}^T + Q \qquad (\beta)$$

WHERE $\bar{K}$ IS THE GAIN ASSOCIATED WITH $\bar{P}$, I.E.

$$\bar{K} = A\bar{P}C^T(C\bar{P}C^T+R)^{-1}.$$

FOR SAKE OF CONTRADICTION, ASSUME $A-\bar{K}C$ UNSTABLE.
THUS, THERE IS A LEFT EIGENVECTOR x S.T.

$$(A-\bar{K}C)^T x = \lambda x, \quad x \neq 0, \quad |\lambda| \geq 1 \qquad (\gamma)$$

FROM β, IF x^* IS THE CONJUGATE TRANSPOSE OF x:

$$x^* \bar{P} x = \underbrace{\overbrace{x^*(A-\bar{K}C)}^{\lambda^* x^*}\,\bar{P}\,\overbrace{(A-\bar{K}C)^T x}^{\lambda x}}_{|\lambda|^2 x^* \bar{P} x \text{ BY } \gamma}$$

$$+ \, x^* K R K^T x + x^* Q x$$

SO THAT

$$\underbrace{(1-|\lambda|^2)}_{\leq 0}\,\underbrace{x^* \bar{P} x}_{\geq 0} = \underbrace{x^* \bar{K} R \bar{K}^T x}_{\geq 0} + \underbrace{x^* Q x}_{\geq 0}$$

$$\Downarrow$$

$$\text{"} \leq 0 = \geq 0 \text{"}$$

$$\Downarrow$$

$$\begin{cases} x^* K R \bar{K}^T x = 0 \\ x^* Q x = 0 \end{cases}$$

$$\Downarrow$$

$$\Downarrow$$

a) $\bar{K}^T x = 0$ (SINCE $R > 0$)

WHICH, COMBINED WITH γ,

IMPLIES $A^T x = \lambda x$

b) $Q x = 0$

$$\Downarrow$$

$$\begin{bmatrix} A^T - \lambda I \\ Q \end{bmatrix} x = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Downarrow$$

$\begin{bmatrix} A - \lambda I & Q \end{bmatrix}$ DOES NOT

HAVE FULL RANK FOR

λ WITH $|\lambda| \geq 1$ BUT

THIS CONTRADICTS

STABILIZABILITY

POINT 6:

$$\lim_{t \to +\infty} P(t) = \bar{P} \quad \forall P(1)$$

WITH THE SAME $\bar{P}$ DEFINED ABOVE SETTING $P(1)=0$.

PROOF:

WE WILL USE DIFFERENT YET EQUIVALENT DRE EXPRESSIONS:

$$P(t+1) = (A - K(t)C)P(t)A^T + Q \qquad (1)$$

AND

$$P(t+1) = (A-K(t)C)P(t)(A-K(t)C)^T + K(t)RK^T(t) + Q \qquad (2)$$

NOW, FIX $P(1) = P_0 > 0$ AND LET

- $\psi(t) = (A-K(t-1)C)(A-K(t-2)C)\dots(A-K(1)C)$
- $\rho > 0$ S.T. $P_0 \geq \rho I$

USING THESE DEFINITIONS IN (2) WE OBTAIN:

PROOF:

WE WILL USE DIFFERENT YET EQUIVALENT DRE EXPRESSIONS:

$$P(t+1) = (A - K(t)C)P(t)A^T + Q \qquad (1)$$

AND

$$P(t+1) = (A-K(t)C)P(t)(A-K(t)C)^T + K(t)RK^T(t) + Q \qquad (2)$$

NOW, FIX $P(1) = P_0 > 0$ AND LET

- $\psi(t) = (A-K(t-1)C)(A-K(t-2)C) \dots (A-K(1)C)$
- $\rho > 0$ S.T. $P_0 \geq \rho I$

USING THESE DEFINITIONS IN (2)

WE OBTAIN:

$$P(t) = \psi(t)P_0\psi^T(t) + [\geq 0]$$

$$\geq \psi(t)P_0\psi^T(t) \geq \rho\,\psi(t)\psi^T(t).$$

WE KNOW THAT THERE IS U S.T.

$$P(t) \leq U \quad \forall t$$

SO,

$$U \geq P(t) \geq \rho\, \psi(t)\, \psi^T(t)$$

AND TAKING THE TRACE

$$\infty > \text{TRACE}[U] \geq \rho \sum_{ij} \psi_{ij}^2(t)$$

$$\Downarrow$$

$\psi(t)$ IS UNIFORMLY BOUNDED IN t.

NOW, WE USE 1:

$$\begin{aligned} P(t+1) &= (A-K(t)C)P(t)A^T + Q \\ &= P^T(t+1) \\ &= A\,P(t)\,(A-K(t)C)^T + Q. \end{aligned}$$

RECALL THAT

$$\bar{P} = (A-\bar{K}C)\bar{P}A^T + Q.$$

NOW WE HAVE:

NOW WE HAVE:

$$P(t+1)-\bar{P} = AP(t)(A-K(t)C)^T - (A-\bar{K}C)\bar{P}A^T$$
$$= (A-\bar{K}C)(P(t)-\bar{P})(A-K(t)C)^T$$
$$\underbrace{+\bar{K}CP(t)(A-K(t)C)^T - (A-\bar{K}C)\bar{P}C^TK^T(t)}_{=0!\ \text{TEDIOUS CALCULATIONS}}$$

$= 0!$ TEDIOUS CALCULATIONS

USING

$$K(t) = A\left(P(t)-P(t)C^T(CP(t)C^T+R)^{-1}CP(t)\right)C^TR^{-1},$$
$$\bar{K} = \text{-------}$$

$$\Downarrow$$

$$P(t+1)-\bar{P} = (A-\bar{K}C)(P(t)-\bar{P})(A-K(t)C)^T$$

$$\Downarrow$$

$$P(t)-\bar{P} = \underbrace{(A-\bar{K}C)^{t-1}}_{\substack{\text{STABLE,}\\ \text{SINCE } \bar{P}\\ \text{IS STABILIZING}}}(P_0-\bar{P})\underbrace{(A-K(1)C)^T \cdots (A-K(t-1)C)^T}_{=\Psi^T(t)\ \text{BOUNDED IN } t}$$

$$\Downarrow$$

$$\lim_{t \to +\infty} P(t) = \bar{P} \quad \text{IF } P(1) = P_0 > 0$$

BUT IF $P_0 \geq 0$?

LET NOW $P_0 \geq 0$ AND $P_0 \neq 0$

$\exists \ \lambda > 0$ S.T.

$$\underbrace{\underbrace{0}_{\substack{\text{WITH}\\ \text{SUCH I.C.}\\ \text{WE CONVERGE}\\ \text{TO } \bar{P},\\ \tilde{P}_0(t) \to \bar{P}}} \quad \leq \quad \underbrace{P(1) = P_0}_{\substack{\text{LEADS}\\ \text{TO}\\ P(t)}} \quad \leq \quad \underbrace{\lambda I}_{\substack{\text{WE JUST}\\ \text{PROVED}\\ \text{CONVERGENCE}\\ \text{TO } \bar{P}\\ \tilde{P}_\lambda(t) \to \bar{P}}}}$$

BY POINT 2, SUCH ORDERING OF I.C. $0 \leq P_0 \leq \lambda I$ PERSISTS IN TIME, I.E.

$$\tilde{P}_0(t) \leq P(t) \leq \tilde{P}_\lambda(t)$$

AND SO ALSO THE $P(t)$ WILL CONVERGE TO $\bar{P}$ AS FORMALLY PROVED BELOW.

$\forall x$ WE HAVE

$$\begin{tikzcd}
x^T \tilde{P}_0(t) x \arrow[d, "t \to +\infty"'] & \leq & x^T P(t) x \arrow[d] & \leq & x^T \tilde{P}_\lambda(t) x \arrow[d, "t \to +\infty"] \\
x^T \bar{P} x & \Longrightarrow & \begin{array}{c}\text{MUST} \\ \text{CONVERGE} \\ \text{TO} \\ x^T \bar{P} x\end{array} & \Longleftarrow & x^T \bar{P} x
\end{tikzcd}$$

$(*)$

NOW, AS ALREADY SEEN

$$e_i^T P(t) e_j = (i,j)\text{-TH ELEMENT OF } P(t)$$

$$= \frac{(e_i + e_j)^T P(t) (e_i + e_j)}{2} - \frac{e_i^T P(t) e_i}{2} - \frac{e_j^T P(t) e_j}{2}$$

$t \to +\infty \downarrow \qquad t \to +\infty \downarrow \qquad t \to +\infty \downarrow$

$$e_i^T P(t)\, e_j = (i,j)\text{-TH ELEMENT OF } P(t)$$

$$= \frac{(e_i+e_j)^T P(t)(e_i+e_j)}{2} - \frac{e_i^T P(t) e_i}{2} - \frac{e_j^T P(t) e_j}{2}$$

$$\frac{(e_i+e_j)^T P(t)(e_i+e_j)}{2} \xrightarrow{t \to +\infty} \frac{1}{2}(e_i+e_j)^T \bar{P}(e_i+e_j)$$

BY $\circledast$ WITH $x = e_i + e_j$

$$\frac{e_i^T P(t) e_i}{2} \xrightarrow{t \to +\infty} \frac{1}{2} e_i^T \bar{P} e_i$$

BY $\circledast$ WITH $x = e_i$

$$\frac{e_j^T P(t) e_j}{2} \xrightarrow{t \to +\infty} \frac{1}{2} e_j^T \bar{P} e_j$$

BY $\circledast$ WITH $x = e_j$

THUS,

$$[P(t)]_{ij} \longrightarrow [\bar{P}]_{ij} \quad \text{SINCE}$$

$$\frac{1}{2}(e_i+e_j)^T \bar{P}(e_i+e_j) - \frac{1}{2} e_i^T \bar{P} e_i - \frac{1}{2} e_j^T \bar{P} e_j$$

$$\equiv$$

$$[\bar{P}]_{ij}$$

$\blacksquare$

CONCLUDING REMARKS

- DETECTABILITY AND STABILIZABILITY ARE ALSO NECESSARY.
  E.G. IF Σ IS NOT DETECTABLE, SOME UNSTABLE MODES ARE NOT SEEN IN THE OUTPUT. THEN, AN I.C. WHICH MAKES THOSE MODES ACTIVE WILL MAKE VARIANCE OF PREDICTION ERROR $P(t)$ GROW TO INFINITY.
  (EVEN IF THE I.C. OF THE NON OBS. SYSTEM WERE KNOWN, THE ERROR COULD DIVERGE IF THE TRANSITION NOISE EXCITES THE UNSTABLE PART)
- IF $v(t)$ AND $w(t)$ ARE CORRELATED, THE NECESSARY AND SUFFICIENT CONDITIONS BECOME

- IF $v(t)$ AND $w(t)$ ARE CORRELATED, THE NECESSARY AND SUFFICIENT CONDITIONS BECOME

(F, C) DETECTABLE

$(F, \tilde{Q}^{1/2})$ STABILIZABLE

WITH $F, \tilde{Q}$ DEFINED SOME LECTURES AGO (THEY ARE OUTCOMES FROM THE WHITENING OPERATION).

- THEOREM:

LET (A, B) BE STABILIZABLE.

THEN

(A, C) DETECTABLE $\iff$ $\forall P_0, \exists U$ S.T. $P(t) \leq U \ \forall t$

MEANING: POINT 1 SHOWED THAT

MEANING: POINT 1 SHOWED THAT DETECTABILITY IS SUFFICIENT FOR EXISTENCE OF U.
STABILIZABILITY MAKES IT ALSO NECESSARY SINCE IT GUARANTEES THAT THE TRANSITION NOISE CAN EXCITE THE UNSTABLE MODES OF THE NON OBSERVABLE SUBSYSTEM (IF STABILIZABILITY HOLDS, THE UNREACHABLE MODES INCLUDE ONLY STABLE MODES)

- THEOREM:

(A, C) DETECTABLE $\Longrightarrow$ $\exists$ AT LEAST ONE ARE SOLUTION (THE $\bar{P}$ OBTAINED WITH $P_0 = 0$!)

HOLDS, THE UNREACHABLE MODES INCLUDE ONLY STABLE MODES)

- THEOREM:

(A, C) DETECTABLE $\Longrightarrow$ $\exists$ AT LEAST ONE ARE SOLUTION (THE $\bar{P}$ OBTAINED WITH $P_0 = 0$!)

- THEOREM:

(A, B) STABILIZABLE $\Longrightarrow$ THERE IS AT MOST ONE $\bar{P} = \bar{P}^T \geq 0$ ARE SOLUTION (NECESSARILY STABILIZING)

KALMAN SMOOTHING FILTER

MEASUREMENTS ON A FIXED INTERVAL

$$y(1), y(2), \ldots, y(N)$$

AND WE WANT

$$\hat{E}[x(t) \mid y(1), \ldots, y(N)] = \hat{x}(t \mid N)$$

WHERE STILL WE HAVE

$$x(t+1) = Ax(t) + v(t), \quad v \sim (0, Q)$$

$$y(t) = Cx(t) + w(t), \quad w \sim (0, R)$$

$$v(t) \perp w(t)$$

THE KALMAN FILTER GIVES $\hat{x}(t|t)$ AND $\hat{x}(t+1|t)$ USING FORWARD RECURSIONS

FORWARD RECURSIONS

$$\xrightarrow[t]{\qquad\qquad\qquad}$$

THEOREM:

ONE HAS

$$\hat{x}(t|N) = \hat{x}(t|t) + R_t\left(\hat{x}(t+1|N) - \hat{x}(t+1|t)\right)$$

$$R_t = P(t|t)A^T P^{-1}(t+1|t)$$

BACKWARD RECURSIONS

$$\xleftarrow[t]{\qquad\qquad\qquad}$$

PROOF:

LET US COMPUTE TWO PROJECTIONS

1

$$\hat{E}\left[x(t) \mid \underbrace{\{y(k)\}_{k=1}^{t}}, \underbrace{x(t+1)-\hat{x}(t+1|t)}, \underbrace{\{v(k), w(k)\}_{k=t+1}^{N}}\right]$$

THREE UNCORRELATED BLOCKS

$$= \underbrace{\hat{E}\left[x(t) \mid \{y(k)\}_{k=1}^{t}\right]} + \underbrace{\hat{E}\left[x(t) \mid x(t+1)-\hat{x}(t+1|t)\right]}$$

$$= \underbrace{\hat{E}\left[x(t) \mid \{y(k)\}_{k=1}^{t}\right]}_{\hat{x}(t|t)} + \underbrace{\hat{E}\left[x(t) \mid x(t+1) - \hat{x}(t+1|t)\right]}_{\circledast \text{ SEE BELOW!}}$$

$$+ \underbrace{E\left[x(t) \mid \{v(k), w(k)\}_{k=t+1}^{N}\right]}_{0}$$

LET $m(t+1) := x(t+1) - \hat{x}(t+1|t)$,

$$\Downarrow$$

$$\mathrm{VAR}[m(t+1)] = P(t+1|t)$$

THEN $\circledast$

$$\equiv$$

$$\mathrm{COV}(x(t), m(t+1)) \, \mathrm{VAR}^{-1}(m(t+1)) \, m(t+1)$$

$$= \mathrm{COV}\left(x(t) - \hat{x}(t|t) + \hat{x}(t|t), \; A(\underbrace{x(t) - \hat{x}(t|t)}_{\perp \hat{x}(t|t)}) + \underbrace{w(t)}_{\perp \text{ TO ALL THE LHS}}\right)$$

$$\cdot P^{-1}(t+1|t) \, m(t+1)$$

$$= P(t|t) A^{T} P^{-1}(t+1|t) \, m(t+1)$$

SO, LETTING

SO, LETTING

$$S = \text{SPAN}\left\{ \{y(k)\}_{k=1}^{t},\ m(t+1),\ \{v(k), w(k)\}_{k=t+1}^{N} \right\}$$

WE HAVE OBTAINED

$$\hat{E}\left[x(t) \mid S\right] = \hat{x}(t|t) + P(t|t)A^{T} P^{-1}(t+1|t)\, m(t+1)$$

② THE SECOND PROJECTION (FINAL STEP)

$$A = \text{SPAN}\{ y(1), \dots, y(N) \}$$

$$\Downarrow$$

$$A \subseteq S \quad \text{(SEE NOTE AT THE END)}$$

$$\Downarrow$$

$$\hat{x}(t|N) = \hat{E}\left[x(t) \mid A\right] = \hat{E}\left[E\left[x(t) \mid S\right] \mid A \right]$$

$$\overset{\parallel}{\hat{E}[x(t)|A]} = \hat{E}\underbrace{\Big[\, E[x(t)|S] \,\Big|A\Big]}$$

$$\equiv$$

$$\underbrace{\hat{x}(t|t)}_{\substack{\text{NOT AFFECTED}\\ \text{BY THE PROJECTION}\\ \text{ONTO A SINCE IT}\\ \text{ALREADY BELONGS}\\ \text{TO A}}} + P(t|t)\,A^T P^{-1}(t+1|t)\Big(\underbrace{\hat{x}(t+1|N)}_{\substack{x(t+1)\\ \text{PROJECTED}\\ \text{ONTO A}}} - \underbrace{\hat{x}(t+1|t)}_{\substack{\text{NOT}\\ \text{AFFECTED}\\ \text{BY THE}\\ \text{PROJ.}\\ \text{ONTO}\\ \text{A}}}\Big)$$

$$\begin{matrix}\hat{x}(t|t),\\ \hat{x}(t+1|t)\end{matrix} \in \text{SPAN}\{y(1),\dots,y(t)\}$$

WHY $A \subseteq S$?

$$S = \text{SPAN}\Big\{ \underbrace{\{y(k)\}_{k=1}^{t},\ x(t+1) - \overbrace{\hat{x}(t+1|t)}^{\in\,\text{SPAN}\{y(1),\dots,y(t)\}}}_{\substack{\text{ABLE TO GENERATE}\\ \text{ALL}}},\ \{v(k), w(k)\}_{k=t+1}^{N} \Big\}$$

$$\underbrace{\text{SPAN}\{y(1),\dots,y(t),x(t+1)\}}$$

ONTO A SINCE IT ALREADY BELONGS TO A

PROJ. ONTO A

$$\hat{x}(t|t),\ \hat{x}(t+1|t) \in \text{SPAN}\{y(1), \dots, y(t)\}$$

WHY $A \subseteq S$?

$$S = \text{SPAN}\Big\{ \underbrace{\underbrace{\{y(k)\}_{k=1}^{t},\ x(t+1) - \overset{\in \text{SPAN}\{y(1),\dots,y(t)\}}{\hat{x}(t+1|t)}}_{\text{ABLE TO GENERATE ALL } \text{SPAN}\{y(1),\dots,y(t),x(t+1)\}},\ \{v(k), w(k)\}_{k=t+1}^{N}}_{\text{ABLE TO GENERATE ALL } \text{SPAN}\{y(1),\dots,y(N)\} = A} \Big\}$$

ABLE TO GENERATE ALL SPAN$\{y(1), \dots, y(t), x(t+1)\}$

ABLE TO GENERATE ALL SPAN$\{y(1), \dots, y(N)\} = A$

$$\Downarrow$$

$$A \subseteq S$$