# OPTIMAL FILTERING

**LECTURE 1**

1. Formalities
2. Some background
3. Least-squares estimation
4. Conditional mean
5. Linear least-squares estimation
6. Example with Gaussian distribution
7. Computational complexity and displacement rank

KTH Electrical Engineering



# Optimal Filtering, Format and Requirements

Two versions of the course!

**Master course EQ2800 (6cr)**

- 7 weekly lectures, Mats Bengtsson, Magnus Jansson
- 7 sets of weekly homeworks. **Individually formulated solutions!**
- One project assignment, written report. Groups of 2 students.
- Preliminary grading: E=60%, D=65%, C=70%, B=80%, A=90% on the homeworks.
- Course web page: `www.kth.se/social/course/EQ2800`

KTH Electrical Engineering

# Optimal Filtering, Format and Requirements

**PhD course FEM3200 (10cr)**

- 9 weekly lectures, Mats Bengtsson, Magnus Jansson
- 8 sets of weekly homeworks. **Individually formulated solutions!**
- One project assignment, written report. Groups of 2 students.
- Group-wise peer grading of homework
- Individual presentation on selected topic
- Take home examination (72 hours).
- Grade: Pass/Fail
- Prel. requirements: $\geq 70\%$ on homeworks+ $\geq 50\%$ on final exam.
- Webpage: `www.kth.se/social/group/optimal-filtering-ph`

KTH Electrical Engineering



# OPTIMAL FILTERING

**BRIEF COURSE OUTLINE**

1. Some Basic Estimation Theory and Geometric Interpretation
2. Wiener Filters, Continuous Time and Discrete Time
3. Discrete Time Kalman Filters
4. Innovations Process
5. Stationary Kalman Filter, spectral properties
6. Smoothing (fixed-point, fixed-lag, fixed-time)
7. Non-linear filtering

.................................................................

8. Continuous Time Kalman Filters
9. Numerical and computational issues

KTH Electrical Engineering

# What?

Linear least-squares estimation of signals with finite dimensional state-space models

# Why?

Filtering – Prediction –Smoothing
problems in a wide range of engineering and scientific disciplines



# Who?

Babylonians $\Longrightarrow$ Galilei $\Longrightarrow$ Gauss ...

$$\left.\begin{array}{l}\text{Kolmogorov}\\ \text{Krein}\\ \text{Wold}\end{array}\right\}\begin{array}{l}\text{discrete time}\\ \text{stationary processes}\end{array}$$

# Who? (cont.)

$$\left.\begin{array}{l}\text{Wiener}\\ \text{(Hopf)}\end{array}\right\}\begin{array}{l}\text{continuous time}\\ \text{stationary processes}\end{array}$$

Kalman 1960
(Bucy, Stratonovich)
Swerling 1958

Kailath
B.D.O. Anderson
Wonham, Rissanen ...



# Swedish Work

Zachrisson

Åström

Mårtensson

Ljung

Lindquist

Hedelin

For a historical account see the reference:

**T. Kailath** "A view of three decades of linear filtering theory", IEEE Trans. on Information Theory 1974 page 146–.

# LEAST-SQUARES ESTIMATION

**Generic Problem**

Estimate $x$, given $\{y_0, \ldots, y_N\}$ so as to minimize

$$\mathrm{E}\{x - h(y_0, \ldots, y_N)\}^2$$



# LEAST-SQUARES ESTIMATION (CONT.)

**Why Least-Squares?**

- explicit in terms of conditional mean
- for Gaussian random variables, the LS estimate is a linear function of the observations
- linear LS estimates depend only on first and second order statistics of the random variables involved
- conditional expectation – connections with martingale theory – useful in signal detection

# LEAST-SQUARES ESTIMATION (CONT.)

**Theorem:**

Let $X$ and $Y$ be two jointly distributed random variables. The least-squares (minimum variance) estimator $\hat{X}$ of $X$ given $Y$ is

$$\hat{X}(Y) = \mathrm{E}_X\{X|Y\}$$

Proof:

$$\mathrm{E}\{(x - h(y))^2\} = \int\int f_{XY}(x, y)(x - h(y))^2 dx dy$$
$$= \int \underbrace{f_Y(y)}_{>0} dy \underbrace{\int f_{X|Y}(x|y)(x - h(y))^2 dx}_{\text{minimize}}$$



$$\int f_{X|Y}(x|y)(x-h(y))^2 dx = \int x^2 f_{X|Y}(x|y)dx - 2h(y)\int x f_{X|Y}(x|y)dx + h^2(y)$$

$$= \left(h(y) - \int x f_{X|Y}(x|y)dx\right)^2 + \int x^2 f_{X|Y}(x|y)dx - \left(\int x f_{X|Y}(x|y)dx\right)^2$$

Minimum achieved when

$$h(y) = \int x f_{X|Y}(x|y)dx = \mathrm{E}_X\{X|Y\} = \hat{X}(Y)$$

Remark: Note that the minimum *conditional* variance estimate is given by

$$\hat{x} = \mathrm{E}\{X|Y = y\} \quad \text{(not a random variable)}$$

# LEAST-SQUARES ESTIMATION (CONT.)

Note that the minimum conditional mean square error is

$$\mathrm{E}_X\{(x-\hat{x})^2|Y=y\} = \int x^2 f_{X|Y}(x|y)dx - \left(\int x f_{X|Y}(x|y)dx\right)^2$$
$$= \mathrm{E}_X\{x^2|Y=y\} - \hat{x}^2$$

In general this is not the same as the minimum (unconditional) mean square error.

$$\mathrm{E}_{X,Y}\{(X-\hat{X}(Y))^2\}$$

However, these are the same in the Gaussian case (show this!).



# UNBIASEDNESS

Unbiased *estimate*

$$\mathrm{E}_{X|Y}\{x-\hat{x}\} = \underbrace{\mathrm{E}_{X|Y}\{x|Y=y\}}_{\hat{x}} - \hat{x} = 0$$

Unbiased *estimator*

$$\mathrm{E}_{X,Y}\{X-\hat{X}(Y)\} = \mathrm{E}_X\{X\} - \underbrace{\mathrm{E}_Y\{\mathrm{E}_X\{X|Y\}}_{\mathrm{E}_X\{X\}} = 0$$

## LEAST-SQUARES ESTIMATION (CONT.)

Problems with this explicit solution are that

- $\hat{x}$ is often a complicated function of $\{y_0, \ldots, y_n\}$
- computation of $\hat{x}$ requires knowledge of the *joint probability density function* $f_{X|Y}(x|Y = y_0, \ldots, y_N)$ which in general is unknown



## LINEAR LEAST-SQUARES ESTIMATION

If our estimate of $x$ is restricted to be a linear function of the observations, it turns out that $\hat{x}$ depends only on

$$\mathrm{E}\{y\} - \text{the mean value}$$
$$\mathrm{E}\{(y - \mathrm{E}\{y\})(y - \mathrm{E}\{y\})^T\} - \text{the covariance}$$

*Tremendous simplification* in general and the major reason for focusing on *linear* least-squares estimation.

Also, if $\{x, y_0, \ldots, y_N\}$ are *jointly Gaussian*, the conditional mean is a linear function of the observations.

# LINEAR LEAST-SQUARES ESTIMATION

Let $X$ and $Y$ be two real-valued jointly Gaussian random variables with

$$\text{mean} \quad \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} \quad \text{and covariance matrix} \quad \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$$

i.e.

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim N\left( \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \right)$$

then the conditional pdf (probability density function) $f_{X|Y}(x|y)$ is Gaussian with

mean $\bar{x} + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \bar{y})$ and covariance matrix $\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}$



# LLSE (CONT.)

$$\Longrightarrow \quad \hat{x} = \mathrm{E}\{X|Y = y\} = \bar{x} + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \bar{y}) \quad \text{Show this!}$$

Note that $\hat{x}$ is a linear (affine) in the observations $y$.

Let us constrain $\hat{X}$ to be a linear (affine) function of $Y$.

$$\hat{X} = hY + g$$

and find $h, g$ so as to minimize the mean square error (mse).

Scalar case:

$$\begin{aligned} \mathrm{E}\{(X - hY - g)^2\} = \mathrm{E}\{X^2\} + h^2\mathrm{E}\{Y^2\} + g^2 \\ - 2g\bar{x} + 2g\bar{y}h - 2h\mathrm{E}\{XY\} \end{aligned}$$

# LLSE (cont.)

Differentiate $\mathrm{E}\{X^2\} + h^2\mathrm{E}\{Y^2\} + g^2 - 2g\bar{x} + 2g\bar{y}h - 2h\mathrm{E}\{XY\}$

$$\Longrightarrow \quad g = \bar{x} - h\bar{y} \qquad \mathrm{E}\{Y^2\}h = \mathrm{E}\{XY\} - g\bar{y}$$

$$\Longrightarrow \quad h = \frac{\rho}{\sigma_y^2} \quad \text{and} \quad g = \bar{x} - \rho\frac{\bar{y}}{\sigma_y^2}$$

where

$$\rho = \mathrm{E}\{(X-\bar{x})(Y-\bar{y})\} = \mathrm{E}\{XY\} - \bar{x}\bar{y}$$
$$\sigma_y^2 = \mathrm{E}\{(Y-\bar{y})^2\} = \mathrm{E}\{Y^2\} - \bar{y}^2$$

Thus, the linear least-squares estimate of $X$ given $Y$ is

$$\hat{X} = \bar{x} + \frac{\rho}{\sigma_y^2}(Y - \bar{y})$$

Depends only on first and second order statistics!



# LLSE (cont.)

Vector case is more cumbersome but straight forward (see *Linear Estimation*). We obtain

$$\Longrightarrow \quad h\Sigma_{yy} = \Sigma_{xy} \quad g = \bar{x} - h\bar{y}$$

$$\boxed{\hat{X} = \bar{x} + \Sigma_{xy}\Sigma_{yy}^{-1}(Y - \bar{y})}$$

# LLSE (cont.)

**Reduction to the zero mean case**

Replace $X$ and $Y$ by the zero mean random variables $X - \bar{x}$ and $Y - \bar{y}$.

The linear least-squares estimate reduces to

$$\hat{X} = \Sigma_{xy}\Sigma_{yy}^{-1}Y$$

Of course, by substituting the original random variable we retain the original expression.

Thus, we can assume *zero mean random variables without loss of generality*.



# GENERIC SOLUTION

If $\hat{x}_N = \sum_{i=0}^{N} a_{N,i}y_{N-i}$ then the coefficients $A_N = [a_{N,0} \cdots a_{N,N}]$ are the solution of the linear equations

$$A_N\Sigma_{yy} = \Sigma_{xy}$$

Standard problem *but*

1. takes $O(N^3)$ operations (additions and multiplications) to compute and $N$ may be large!
2. often we want to find solutions *recursively*

$$A_N \to A_{N+1} \quad \text{and} \quad \hat{x}_N \to \hat{x}_{N+1}$$

Therefore we need to impose *more structure* on the random variables $\{y_0, \ldots, y_N\}$ to avoid these problems.

# SUMMARY OF RESULTS

Two main classes of assumptions

1. Stationarity or near-stationarity (index $r$)
2. Finite dimensionality (the number of states, $n$, of the stochastic process is finite)

| | |
|---|---|
| General non-stationary process | $O(N^3)$ |
| Stationary processes | $O(N^2)$ |
| $r$-stationarity ($1 \leq r \leq N$) | $O(rN^2)$ |
| General $n$-state model | $O(Nn^3)$ |
| Stationary $n$-state model | $O(Nn^2)$ |
| $r$-stationary $n$-state model | $O(rNn^2)$ |



# DISPLACEMENT RANK

A measure of the non-stationarity of a stochastic process is provided by the *displacement rank*, $r$.

Let $\Sigma_{yy} = \mathrm{E}\{(Y - \bar{y})(Y - \bar{y})^T\}$ be the covariance matrix of $\{y\}$ and

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

The displacement rank is defined as

$$r \triangleq \mathrm{rank}(\Sigma_{yy} - Z\Sigma_{yy}Z^*)$$

What is the displacement rank of a stationary process?

# APPLICATIONS

- Adaptive filtering
- Target tracking
- GPS, navigation
- Change detection
- ⋮