# Predictive Control
# Lecture 2

**Bo Bernhardsson**

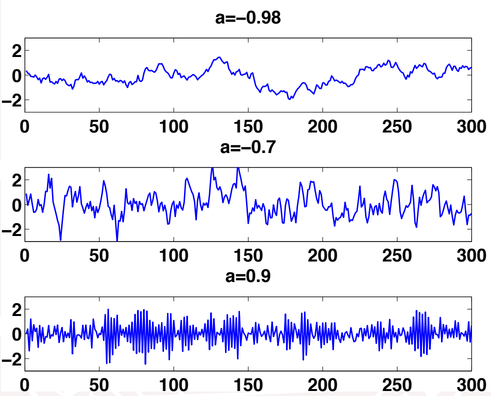
Department of Automatic Control
Lund University, Sweden

# Lecture 2

Today's lecture

- Stochastic Signal Models (Chapter 3)
- Recursive Identification (Chapter 4)

# Some Stochastic Signals with unit variance




The signals seem to have different properties

We should design different predictors/controllers for them

How do we describe spectral properties of stochastic signals?

# Stochastic Variables

- $\mu_x = E(x)$ = mean value of $x$
- $\sigma_x^2 = V(x) = E(x - \mu_x)^2$ = variance of $x$
- $Cov(x, y) = E(x - \mu_x)(y - \mu_y)$ the cross-covariance between $x$ and $y$
- Uncorrelated if $Cov(x, y) = 0$

If $x$ and $y$ are (column) vectors

- $P_x = V(x) = E(x - \mu_x)(x - \mu_x)^T$ = variance matrix
- $Cov(x, y) = E(x - \mu_x)(y - \mu_y)^T$ cross-covariance matrix

# White noise

A sequence of uncorrelated stochastic variables with

$$E(w_i) = 0, \qquad E(w_i^2) = \sigma^2, \qquad Cov(w_i, w_j) = 0, \;\; i \neq j$$

is called **white noise** with variance $\sigma^2$.

This is in some sense a "worst case" enemy, since old noise observations reveal no information at all about the future.

# Useful facts

If $y = Ax + b$ where $x$ is stochastic and $A$ and $b$ known, then

- $E(y) = AE(x) + b$
- $V(x) = AV(x)A^T$

Example: The system $x_{k+1} = Ax_k + e_k$ with $e$ white noise with variance matrix $R$ has

$$P_x(k+1) = Cov(x_{k+1}x_{k+1}^T) = \ldots = AP_x(k)A^T + R$$

# Autocovariance functions

For a stationary signal $y_k$ the covariance function is defined as

$$C_{yy}(\tau) = Cov(y(k+\tau)y(k)^T)$$

2 min-problem: Assume that

$$x_{k+1} = Ax_k + e_k$$

($e$ is white noise) and that $x$ is stationary. Why is $C_{xx}(\tau) = A^\tau V(x)$ for $\tau \geq 0$? What happens if $\tau < 0$? .

# Spectrum of stochastic signals

The **auto-spectrum** of $y$ is defined as $S_{yy}(z) = \sum_{k=-\infty}^{\infty} C_{yy}(k) z^{-k}$

The **power spectral density** at frequency $\omega = 2\pi f$ is given by

$$S_{yy}(e^{i\omega h})$$

Similar definitions for $S_{uu}(z)$ and $S_{yu}(z)$

If $Y(z) = G(z)U(z) + V(z)$ with uncorrelated $u$ and $v$ then

$$S_{yy}(z) = G(z)S_{uu}(z)G^T(z^{-1}) + S_{vv}(z)$$

# Example

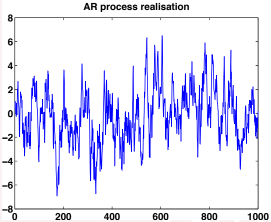
Assume $|a| < 1$, $h = 1$, and $w$ is white noise ($W(z) = 1$)

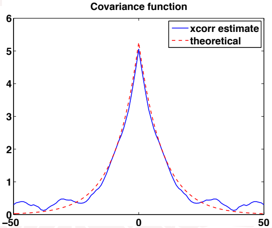
$$y_k = -a y_{k-1} + w_k \iff Y(z) = \frac{1}{1 + a z^{-1}} W(z) = \frac{z}{z + a} W(z)$$

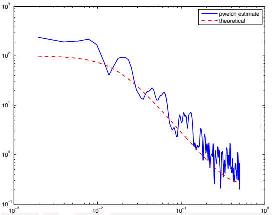
$$\begin{aligned} C_{yy}(z) &= \frac{1}{1 - a^2} (-a)^{|\tau|} \\ S_{yy}(z) &= \frac{1}{1 + az} \frac{1}{1 + az^{-1}} \\ S_{yy}(e^{i\omega}) &= \frac{1}{1 + a^2 + 2a \cos \omega} \end{aligned}$$

Matlab: `xcorr` and `pwelch`,

# Matlab simulation



# Example

```
figure(1);
nr=1000; % nr of samples
a=-0.9; h=1; w=randn(nr,1);
y = filter(1,[1 a],w);
plot(y,'linewidth',1.4)

figure(2);
[C,tau] = xcorr(y,50,'unbiased');
plot(tau,C,'linewidth',1.4); hold on;
plot(tau,1/(1-a^2)*(-a).^abs(tau),'r--','linewidth',1.4);

figure(3); nrpoints=512;
[Pxx,f]=pwelch(y,[],[],nrpoints,1/h);
loglog(f,Pxx,'linewidth',1.4); hold on;
loglog(f,1./(1+a^2+2*a*cos(2*pi*f)),'r','linewidth',1.4)
```

# Stationary State Covariance

Assume $A$ stable and $v$ and $e$ uncorrelated

$$\begin{aligned} x_{k+1} &= Ax_k + v_k \\ y_k &= Cx_k + e_k \end{aligned}$$

Stationary covariance $P_x = E(xx^T)$ can be found from

$$P_x = AP_xA^T + Q_v$$

Stationary covariance $P_y$ is then given by

$$P_y = CP_xC^T + Q_e$$

Matlab: `Px = lyap(A,Qv); Py=C*Px*C+Qe`

# Real-time Parameter Estimation

1. Introduction
2. Least squares and regression
3. Dynamical systems
4. Experimental conditions
5. Examples
6. Conclusions

# System Identification

- How to get models?
  - Physics (white-box)
  - Experiments (black-box)
  - Combination (grey-box)
- Experiment planning
- Choice of model structure
  - Transfer functions
  - Impulse response
  - State-space models
- Parameter estimation
  - System identification/Statistics/Inverse problems
- Validation

# Identification Techniques

- Nonparametric methods
  - Frequency response analysis
  - Transient response analysis
  - Correlation analysis
  - Spectrum analysis
- Parametric methods
  - Least Squares (LS)
  - Maximum Likelihood (ML)
  - Subspace-model identification
- Identification for control
- Model approximation/ Model Reduction

Related areas: Statistics, Numerical analysis, Econometrics, ...

# Least Squares and Regression

- Introduction
- The LS problem
- Interpretation
  - Geometric
  - Statistic
- Recursive Calculations
- Continuous time models

# Good Methods are adopted by Everybody

- Mathematics
- Statistics
- Numerical analysis
- Physics
- Economics
- Biology
- Medicine
- Control
- Signal processing

# THEORIA
# MOTVS CORPORVM
# COELESTIVM

IN

SECTIONIBVS CONICIS SOLEM AMBIENTIVM

AVCTORE

CAROLO FRIDERICO GAVSS.

HAMBVRGI SVMTIBVS FRID. PERTHES ET I. H. BESSER
1809.

# The Least Squares Method

The problem: The Orbit of Ceres

The problem solver: Karl Friedrich Gauss

The principle: “Therefore, that will be the most probable system of values of the unknown quantities, in which the sum of the squares of the differences between the observed and computed values, multiplied by numbers that measure the degree of precision, is a minimum.”

In conclusion, the principle that the sum of the squares of the differences between the observed and computed quantities must be a minimum, may be considered independently of the calculus of probabilities.

An observation: Other criteria could be used. “But of all these principles ours is the most simple; by the others we should be led into the most complicated calculations.”

# Mathematical Formulation—Regression Model

$$y(t) = \varphi_1(t)\theta_1 + \varphi_2(t)\theta_2 + \cdots + \varphi_n(t)\theta_n = \varphi(t)^T\theta$$

- $y$—observed data
- $\theta_i$—unknown parameters
- $\varphi_i$—known functions regression variables

Some notations

$$\begin{aligned}
\varphi^T(t) &= \big[\varphi_1(t)\ \ \varphi_2(t)\ldots\varphi_n(t)\big], \quad \theta^T = \big[\theta_1\ \ \theta_2\ldots\theta_n\big] \\
Y(t) &= \big[y(1)\ \ y(2)\ldots y(t)\big]^T, \quad E(t) = \big[\varepsilon(1)\ \ \varepsilon(2)\ldots\varepsilon(t)\big]^T \\
\Phi(t) &= \begin{pmatrix} \varphi^T(1) \\ \vdots \\ \varphi^T(t) \end{pmatrix}, \quad P(t) = \left(\sum_{i=1}^{t}\varphi(i)\varphi^T(i)\right)^{-1} = \big(\Phi^T(t)\Phi(t)\big)^{-1} \\
\varepsilon(i) &= y(i) - \hat{y}(i) = y(i) - \varphi^T(i)\theta
\end{aligned}$$

# Solving the LS Problem

Minimize with respect to $\theta$

$$V(\theta,t) = \frac{1}{2}\sum_{i=1}^{t}\varepsilon(i)^2 = \frac{1}{2}\sum_{i=1}^{t}\big(y(i)-\varphi^T(i)\theta\big)^2 = \frac{1}{2}\;\theta^T A\theta - b^T\theta + \frac{1}{2}\;c$$

where

$$A = \sum_{i=1}^{t}\varphi(i)\varphi^T(i), \quad b \quad = \sum_{i=1}^{t}\varphi(i)y(i), \quad c = \sum_{i=1}^{t}y^2(i)$$

The parameter $\hat{\theta}$ that minimizes the loss function are given by the $\mathbf{normal\ equations}$

$$A\hat{\theta} = b, \quad \Rightarrow \hat{\theta} = A^{-1}b = Pb$$

If the matrix $A$ is nonsingular, the minimum is unique

# An Example

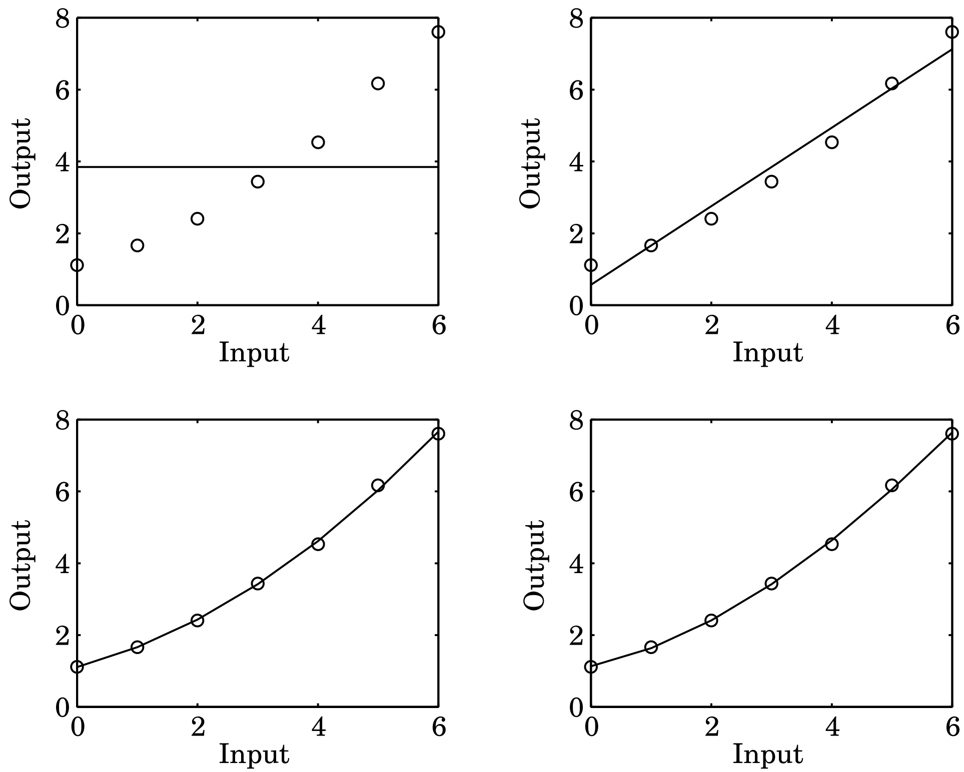
$$
\begin{aligned}
y(t) &= b_0 + b_1 u(t) + b_2 u^2(t) + e(t) \\
\sigma &= 0.1 \\
\varphi^T(T) &= \begin{bmatrix} 1 & u(t) & u^2(t) \end{bmatrix} \\
\theta^T &= [b_0 \quad b_1 \quad b_2]
\end{aligned}
$$

Estimated models

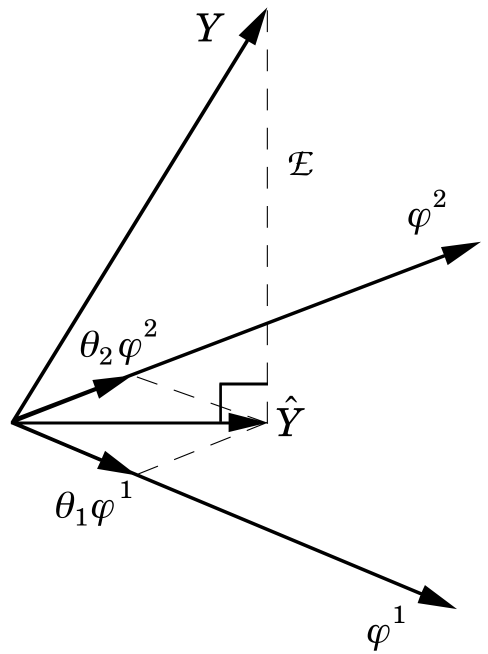
$$
\begin{aligned}
\text{Model } 1: \quad y(t) &= b_0 \\
\text{Model } 2: \quad y(t) &= b_0 + b_1 u \\
\text{Model } 3: \quad y(t) &= b_0 + b_1 u + b_2 u^2 \\
\text{Model } 4: \quad y(t) &= b_0 + b_1 u + b_2 u^2 + b_3 u^3
\end{aligned}
$$

# Example Continued

| Model | $\hat{b}_0$ | $\hat{b}_1$ | $\hat{b}_2$ | $\hat{b}_3$ | $V$ |
|---|---|---|---|---|---|
| 1 | 3.85 | | | | 34.46 |
| 2 | 0.57 | 1.09 | | | 1.01 |
| 3 | 1.11 | 0.45 | 0.11 | | 0.031 |
| 4 | 1.13 | 0.37 | 0.14 | $-0.003$ | 0.027 |



# Geometric Interpretation




$$E = Y - \varphi^1\theta_1 - \varphi^2\theta_2 - \cdots - \varphi^n\theta_n$$

Find the smallest possible $E$?

$$(\varphi^i)^T \left(y - \theta_1\varphi^1 - \theta_2\varphi^2 - \cdots - \theta_n\varphi^n\right) = 0$$

The normal equations!

# Statistical Interpretation

$$y(t) = \varphi^T(t)\theta^0 + e(t)$$

$\theta^0$ – "true" parameters

$e(t)$ – independent random variables with zero mean and variance $\sigma^2$

If $\Phi^T\Phi$ is nonsingular, then

$$E\hat{\theta} = \theta^0$$

$$\text{cov } \hat{\theta} = \sigma^2(\Phi^T\Phi)^{-1} = \sigma^2 P$$

$$s^2 = 2V(\hat{\theta}, t)/(t - n)$$

is an unbiased estimate of $\sigma^2$

$n$ – number of parameters in $\theta^0$

$t$ – number of data

# Recursive Least Squares

Idea:

- Want to avoid repeating all calculations if data new data arrives recursively
- Does there exist a recursive formula that expresses $\hat{\theta}(t)$ in terms of $\hat{\theta}(t-1)$?

# Recursive Least Squares

The LS estimate is given by

$$\begin{aligned}\hat{\theta}(t) &= P(t)\left(\sum_{i=1}^{t}\varphi(i)y(i)+\varphi(t)y(t)\right)\\ P(t) &= (\sum_{i=1}^{t}\varphi(i)\varphi^T(i))^{-1}, \quad P(t)^{-1}=P(t-1)^{-1}+\varphi(t)\varphi^T(t)\end{aligned}$$

But

$$\sum_{i=1}^{t-1}\varphi(i)y(i) \quad = P(t-1)^{-1}\hat{\theta}(t-1) = P(t)^{-1}\hat{\theta}(t-1)-\varphi(t)\varphi^T(t)\hat{\theta}(t-1)$$

The estimate at time $t$ can now be written as

$$\begin{aligned}\hat{\theta}(t) &= \hat{\theta}(t-1)-P(t)\varphi(t)\varphi^T(t)\hat{\theta}(t-1)+P(t)\varphi(t)y(t)\\ &= \hat{\theta}(t-1)+P(t)\varphi(t)\left(y(t)-\varphi^T(t)\hat{\theta}(t-1)\right)=\hat{\theta}(t-1)+K(t)\varepsilon(t)\end{aligned}$$

Want recursive equation for $P(t)$ not for $P(t)^{-1}$

# The Matrix Inversion Lemma

Let $A$, $C$, and $(C^{-1} + DA^{-1}B)$ be nonsingular square matrices. Then

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

Prove by direct substitution

Given $A^{-1}$ we can get the LHS inverse

What about the inverse on the RHS?

# Recursion for $P(t)$

The matrix inversion lemma gives

$$
\begin{aligned}
P(t) &= \left( \sum_{i=1}^{t} \varphi(i)\varphi^T(i) \right)^{-1} \\
&= \left( \sum_{i=1}^{t-1} \varphi(i)\varphi^T(i) + \varphi(t)\varphi^T(t) \right)^{-1} \\
&= \left( P(t-1)^{-1} + \varphi(t)\varphi^T(t) \right)^{-1} \\
&= P(t-1) - P(t-1)\varphi(t) \\
&\quad \times \left( I + \varphi^T(t)P(t-1)\varphi(t) \right)^{-1} \varphi^T(t)P(t-1)
\end{aligned}
$$

Hence

$$
\begin{aligned}
K(t) &= P(t)\varphi(t) \\
&= P(t-1)\varphi(t) \left( I + \varphi^T(t)P(t-1)\varphi(t) \right)^{-1}
\end{aligned}
$$

# Recursive Least-Squares RLS

$$
\begin{aligned}
\hat{\theta}(t) &= \hat{\theta}(t-1) + K(t)\big(y(t) - \varphi^T(t)\hat{\theta}(t-1)\big) \\
K(t) &= P(t)\varphi(t) \\
&= P(t-1)\varphi(t)\big(I + \varphi^T(t)P(t-1)\varphi(t)\big)^{-1} \\
P(t) &= P(t-1) - P(t-1)\varphi(t) \\
&\quad \times \big(I + \varphi^T(t)P(t-1)\varphi(t)\big)^{-1}\varphi^T(t)P(t-1) \\
&= \big(I - K(t)\varphi^T(t)\big)P(t-1)
\end{aligned}
$$

- Intuitive interpretation
- Kalman filter
- Interpretation of $\theta$ and $P$
- Initial values ($P(0) = r \cdot I$)

## Example: Recursive estimate of a constant

Consider the following noisy observation of a constant parameter

$$y_k = \theta + v_k, \quad \mathcal{E}\{v_k\} = 0, \quad \mathcal{E}\{v_i v_j\} = \sigma^2 \delta_{ij}$$

Use linear regression of $y_k = \phi_k \theta + v_k$ with $\phi_k = 1$ for all $k$.

A least-squares estimate is found as the sample average

$$\widehat{\theta}_k = \frac{1}{k} \sum_{i=1}^{k} y_i$$

Include previously made summations in some state $\widehat{\theta}_k$, which is updated when new data arrive!!

A feasible recursive state equation

$$\begin{aligned} \widehat{\theta}_k &= \widehat{\theta}_{k-1} + \frac{1}{k}(y_k - \widehat{\theta}_{k-1}) \\ p_k &= \sigma^2 (\sum_{i=1}^{k} \phi_i \phi_i^T)^{-1} \approx \mathcal{E}\{(\widehat{\theta} - \theta)(\widehat{\theta} - \theta)^T\} \end{aligned}$$

**Example: Recursive estimate of a constant** The parameter variance estimate can be expressed as a state vector with updating in each recursion according to

$$p_k^{-1} = p_{k-1}^{-1} + \frac{1}{\sigma^2}$$

or

$$p_k = \frac{\sigma^2 p_{k-1}}{\sigma^2 + p_{k-1}}$$

where it can be noticed that $p_k \to 0$ as $k \to \infty$. $\square$

## Some properties of recursive least-squares estimation

For convergence analysis, consider the following quadratic function of the parameter error

$$Q(\widehat{\theta}_k) = \frac{1}{2}(\widehat{\theta}_k - \theta)^T P_k^{-1}(\widehat{\theta}_k - \theta) = \frac{1}{2}\widetilde{\theta}_k^T P_k^{-1}\widetilde{\theta}_k$$

This function develops in each recursion according to

$$\begin{aligned}
Q(\widehat{\theta}_k) - Q(\widehat{\theta}_{k-1}) &= \frac{1}{2}\widetilde{\theta}_k^T P_k^{-1}\widetilde{\theta}_k - \frac{1}{2}\widetilde{\theta}_{k-1}^T P_{k-1}^{-1}\widetilde{\theta}_{k-1} \\
&= \frac{1}{2}\widetilde{\theta}_{k-1}^T (P_k^{-1} - P_{k-1}^{-1})\widetilde{\theta}_{k-1} + \widetilde{\theta}_{k-1}^T \phi_k \varepsilon_k + \frac{1}{2}\phi_k^T P_k \phi \\
&= \frac{1}{2}(\widetilde{\theta}_{k-1}^T \phi_k + \varepsilon_k)^2 + \frac{1}{2}(-1 + \phi_k^T P_k \phi_k)\varepsilon_k^2 \\
&= \frac{1}{2}(\widetilde{\theta}_{k-1}^T \phi_k + \varepsilon_k)^2 - \frac{1}{2}\frac{\phi_k^T P_{k-1}\phi_k}{1 + \phi_k^T P_{k-1}\phi_k}\varepsilon_k^2
\end{aligned}$$

Under the assumption $y_k = \phi_k^T \theta + v_k$ so that $\varepsilon_k = -\widetilde{\theta}_{k-1}^T \phi_k + v_k$

$$Q(\widehat{\theta}_k) - Q(\widehat{\theta}_{k-1}) = \frac{1}{2}v_k^2 - \frac{1}{2}\frac{\phi_k^T P_{k-1}\phi_k}{1 + \phi_k^T P_{k-1}\phi_k}\varepsilon_k^2$$

# Time-varying Parameters

Loss function with discounting

$$V(\theta, t) = \frac{1}{2} \sum_{i=1}^{t} \lambda^{t-i} \left( y(i) - \varphi^T(i)\theta \right)^2$$

The LS estimate then becomes

$$\begin{aligned}
\hat{\theta}(t) &= \hat{\theta}(t-1) + K(t)\big(y(t) - \varphi^T(t)\hat{\theta}(t-1)\big) \\
K(t) &= P(t)\varphi(t) \\
&= P(t-1)\varphi(t)\left(\lambda I + \varphi^T(t)P(t-1)\varphi(t)\right)^{-1} \\
P(t) &= \left(I - K(t)\varphi^T(t)\right)P(t-1)\,/\,\lambda
\end{aligned}$$

# Forgetting Factor

$$0 < \lambda \le 1$$

Equivalent time constant

$$e^{-h/T} = \lambda, \quad T = -\frac{h}{\log \lambda} \approx \frac{h}{1-\lambda}$$

**Rule of thumb:** Memory decay to 10% after

$$N = \frac{2}{1-\lambda}$$

- $+$ Forgets old information
- $+$ Adapts quickly when the process changes
- $-$ The estimates get noise sensitive
- $-$ The $P$ matrix may grow (Wind-up)

# Continuous Time Models

Regression model

$$y(t) = \varphi_1(t)\theta_1 + \varphi_2(t)\theta_2 + \cdots + \varphi_n(t)\theta_n = \varphi(t)^T\theta$$

Loss function with forgetting

$$V(\theta) = \int_0^t e^{\alpha(t-s)} \left(y(s) - \varphi^T(s)\theta\right)^2 ds$$

The normal equations

$$\int_0^t e^{-\alpha(t-s)}\varphi(s)\varphi^T(s)ds\hat{\theta}(t) = \int_0^t e^{-\alpha(t-s)}\varphi(s)y(s)ds$$

Estimate is unique if the matrix

$$R(t) = \int_0^t e^{-\alpha(t-s)}\varphi(s)\varphi^T(s)\,ds$$



is positive definite.

# Recursive Equations for Continuous Time Models

Regression model

$$y(t) = \varphi(t)^T \theta$$

Recursive least equations

$$\begin{aligned}
\frac{d\hat{\theta}}{dt} &= P(t)\varphi(t)e(t) \\
e(t) &= y(t) - \varphi^T(t)\hat{\theta}(t) \\
\frac{dP(t)}{dt} &= \alpha P(t) - P(t)\varphi(t)\varphi^T(t)P(t) \\
\frac{dR}{dt} &= -\alpha R + \varphi\varphi^T
\end{aligned}$$

# Real-time Parameter Estimation

1. Introduction
2. Least squares and regression
3. Dynamical systems
4. Experimental conditions
5. Examples
6. Conclusions

# Estimating Parameters in Dynamical Systems

Basic idea: Rewrite the equations as a regression model!

- Dynamical systems
  - FIR models
  - ARMA models
  - Continuous time models
  - Nonlinear models
- Experimental conditions
  - Excitation
  - Closed loop identification

# Finite Impulse Response (FIR) Models

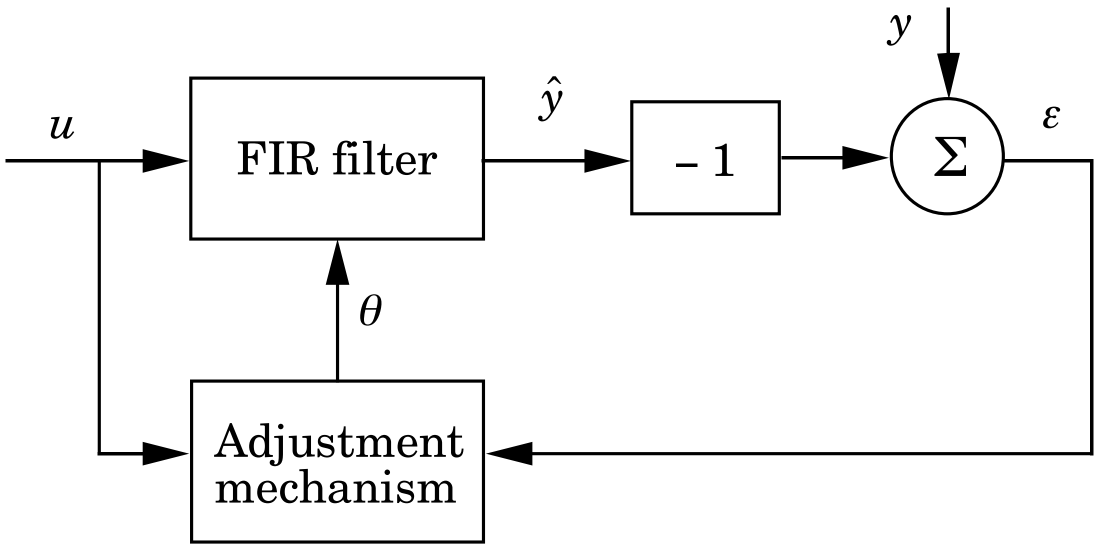
$$y(t) = b_1 u(t-1) + b_2 u(t-2) + \cdots + b_n u(t-n)$$

or

$$y(t) = \varphi^T(t-1)\theta, \quad \theta = [b_1 \ldots b_n]^T$$
$$\varphi^T(t-1) = [u(t-1) \ldots u(t-n)]$$

A regression model!

$$\hat{y}(t) = \hat{b}_1(t-1)u(t-1) + \cdots + \hat{b}_n(t-1)u(t-n)$$



# Pulse Transfer Function Models

$$y(t) + a_1 y(t-1) + \cdots + a_n y(t-n) =$$
$$b_1 u(t-1) + \cdots + b_n u(t-n)$$

Write as

$$y(t) = \varphi(t-1)^T \theta$$

where

$$\varphi(t-1)$$
$$= [-y(t-1) \ldots - y(t-n)\, u(t-1) \ldots u(t-n)]^T$$
$$\theta = [a_1 \ldots a_n\, b_1 \ldots b_n]^T$$

- Autoregression!
- Equation error

# Transfer Function Models

$$\frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} + \cdots + a_n y = b_1 \frac{d^{n-1} u}{dt^{n-1}} + \cdots + b_n u$$

as

$$A(p)y(t) = B(p)u(t)$$

Introduce

$$A(p)y_f(t) = B(p)u_f(t), \quad y_f(t) = F(p)y(t), \quad u_f(t) = F(p)u(t)$$

and $F(p)$ has pole excess greater than $n$

$$\begin{aligned} \theta &= \left[a_1 \ldots a_n \;\; b_1 \ldots b_n\right]^T \\ \varphi_f(t) &= \left[-p^{n-1} y_f \ldots - y_f \;\; p^{n-1} u_f \ldots u_f\right] \\ &= \left[-p^{n-1} F(p) y \ldots - F(p) y \qquad p^{n-1} F(p) u \ldots F(p) u\right] \end{aligned}$$

Hence, a regression model $y_f(t) = \varphi_f^T(t)\theta$

# Nonlinear Models

Consider the model

$$y(t) + ay(t-1) = b_1 u(t-1) + b_2 u^2(t-1)$$

Introduce

$$\theta = [a \;\; b_1 \;\; b_2]^T$$

and

$$\varphi^T(t) = \left[-y(t) \;\; u(t) \;\; u^2(t)\right]$$

Hence

$$y(t) = \varphi^T(t-1)\theta$$

Autoregression!

Linearity in the parameters

# Experimental Conditions

- Excitation
- Closed loop identification
- Model structure

# Persistent Excitation

The matrix $\sum_{k=n+1}^{t} \varphi(k)\varphi^T(k)$ is given by

$$\begin{bmatrix} \sum_{n+1}^{t} u^2(k-1) & \dots & \sum_{n+1}^{t} u(k-1)u(k-n) \\ \vdots & & \\ \sum_{n+1}^{t} u(k-1)u(k-n) & \dots & \sum_{n+1}^{t} u^2(k-n) \end{bmatrix}$$

$$C_n = \lim_{t\to\infty} \frac{1}{t}\Phi^T\Phi, \quad C_n = \begin{bmatrix} c(0) & c(1) & \dots & c(n-1) \\ c(1) & c(0) & \dots & c(n-2) \\ \vdots & & & \\ c(n-1) & c(n-2) & \dots & c(0) \end{bmatrix}$$

$$c(k) = \lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{t} u(i)u(i-k)$$

A signal $u$ is called *persistently exciting* (PE) of order $n$ if the matrix $C_n$ is positive definite.

# Another Characterization

A signal $u$ is persistently exciting of order $n$ if and only if

$$U = \lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^{t} \left(A(q)u(k)\right)^2 > 0$$

for all nonzero polynomials $A$ of degree $n-1$ or less.

**Proof** Let the polynomial $A$ be

$$A(q) = a_0 q^{n-1} + a_1 q^{n-2} + \cdots + a_{n-1}$$

A straightforward calculation gives

$$\begin{aligned} U &= \lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^{t} \left(a_0 u(k+n-1) + \cdots + a_{n-1}u(k)\right)^2 \\ &= a^T C_n a \end{aligned}$$

# Examples

A signal $u$ is called *persistently exciting* (PE) of order $n$ if the matrix $C_n$ is positive definite. An equivalent condition

- A step is PE of order 1

$$(q-1)u(t) = 0$$

- A sinusoid is PE of order 2

$$(q^2 - 2q\cos\omega h + 1)u(t) = 0$$

- White noise
- PRBS
- Physical meaning
- Mathematical meaning

# Loss of Identifiability due to Feedback

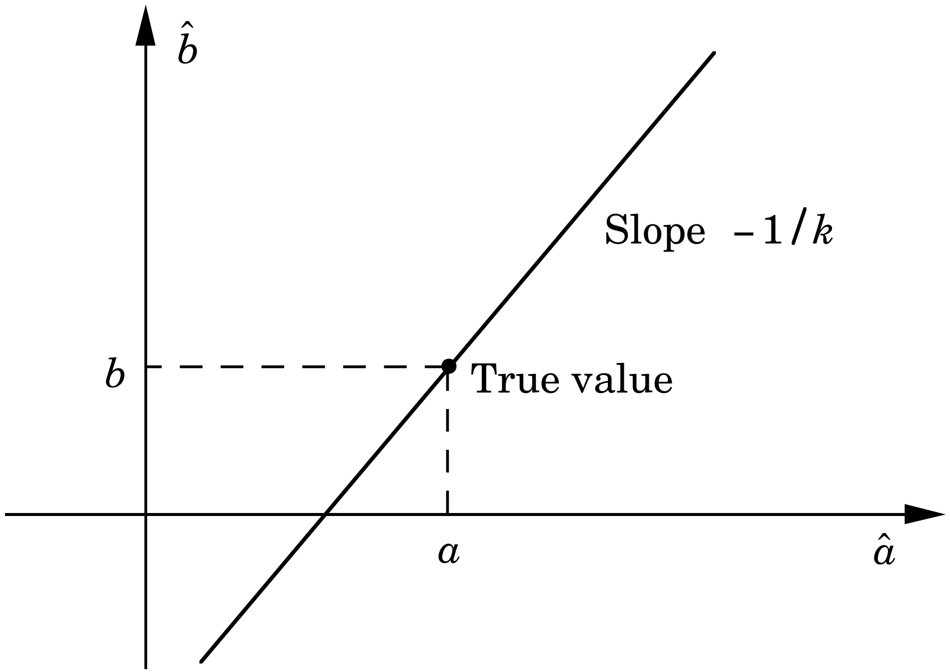
$$y(t) = ay(t-1) + bu(t-1) + e(t), \quad u(t) = -ky(t)$$

Multiply by $\alpha$ and add, then

$$y(t) = (a + \alpha k)y(t-1) + (b + \alpha)u(t-1) + e(t)$$

Same I/O relation for all $\hat{a}$ and $\hat{b}$ such that

$$\hat{a} = a + \alpha k, \quad \hat{b} = b + \alpha$$



# Real-time Parameter Estimation

1. Introduction
2. Least squares and regression
3. Dynamical systems
4. Experimental conditions
5. Examples
6. Conclusions

# Examples

Model

$$y(t) + ay(t-1) = bu(t-1) + e(t)$$

Parameters

$$a = -0.9$$

$$b = 0.5$$

$$\sigma = 0.5$$

$$\hat{\theta}(0) = 0$$

$$P(0) = 100\cdot$$

$$\hat{\theta} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}$$

$$\varphi(t-1) = \begin{pmatrix} -y(t-1) & u(t-1) \end{pmatrix}$$

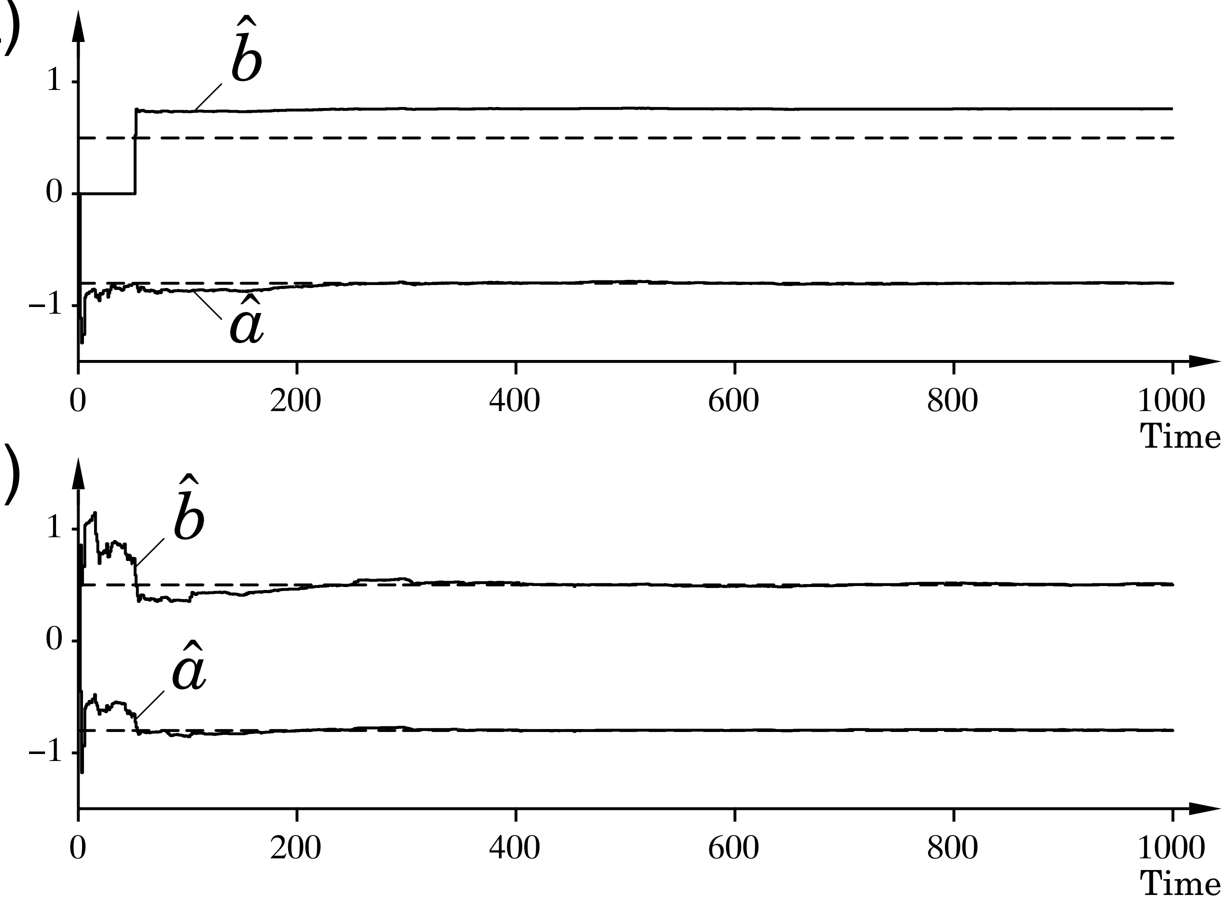
# Excitation

Input:

- Unit pulse at $t = 50$
- Square wave of unit amplitude and period 100

(a)

(b)

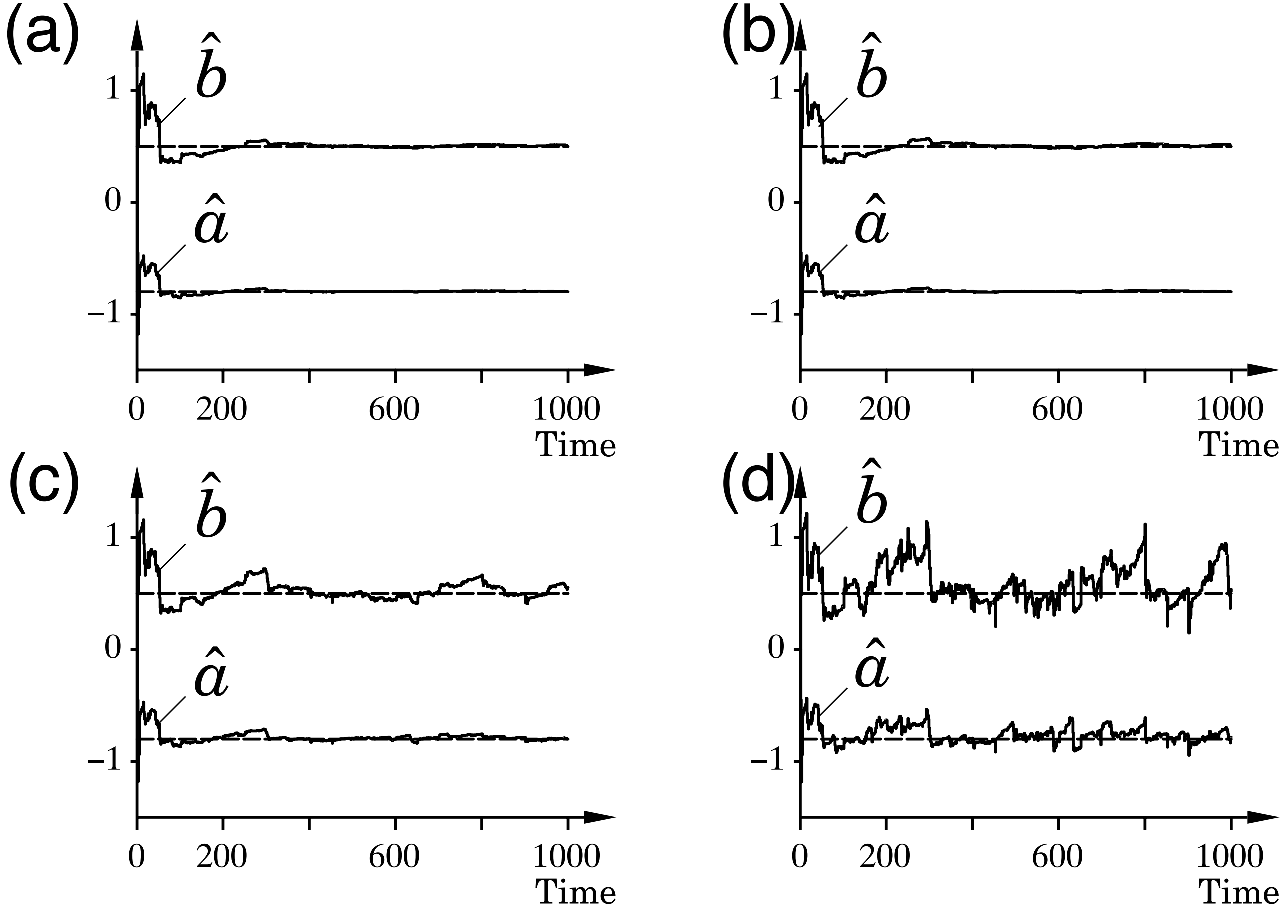


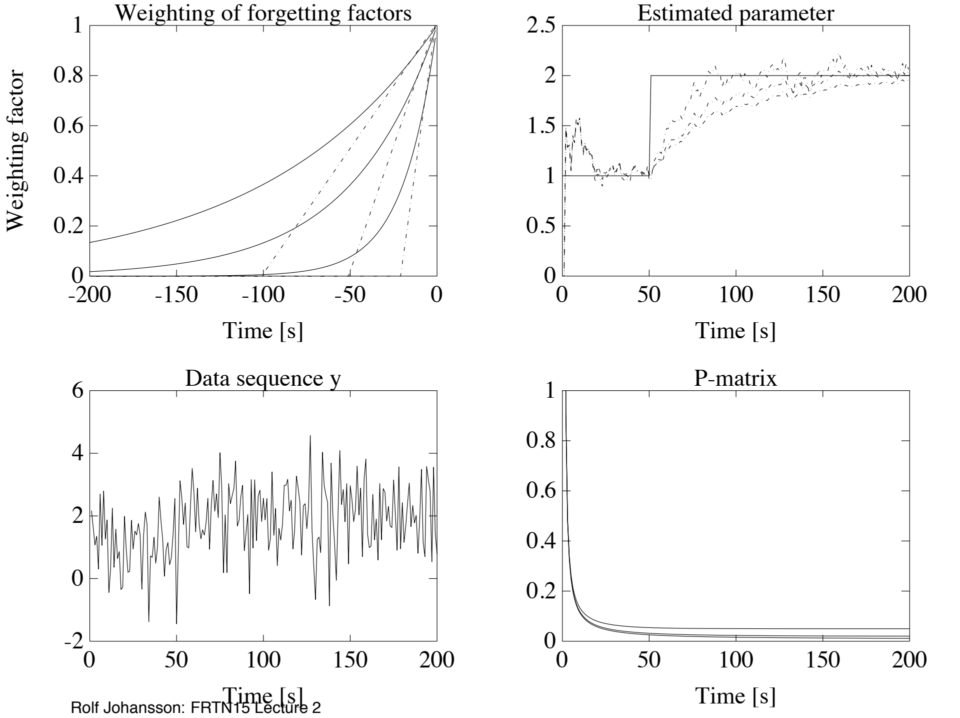
# Forgetting Factor

Recall

$$T \approx \frac{h}{1-\lambda}, \quad N = \frac{2}{1-\lambda}$$

Parameters: $\lambda = 1$, $\lambda = 0,999$, $\lambda = 0.99$, $\lambda = 0.95$



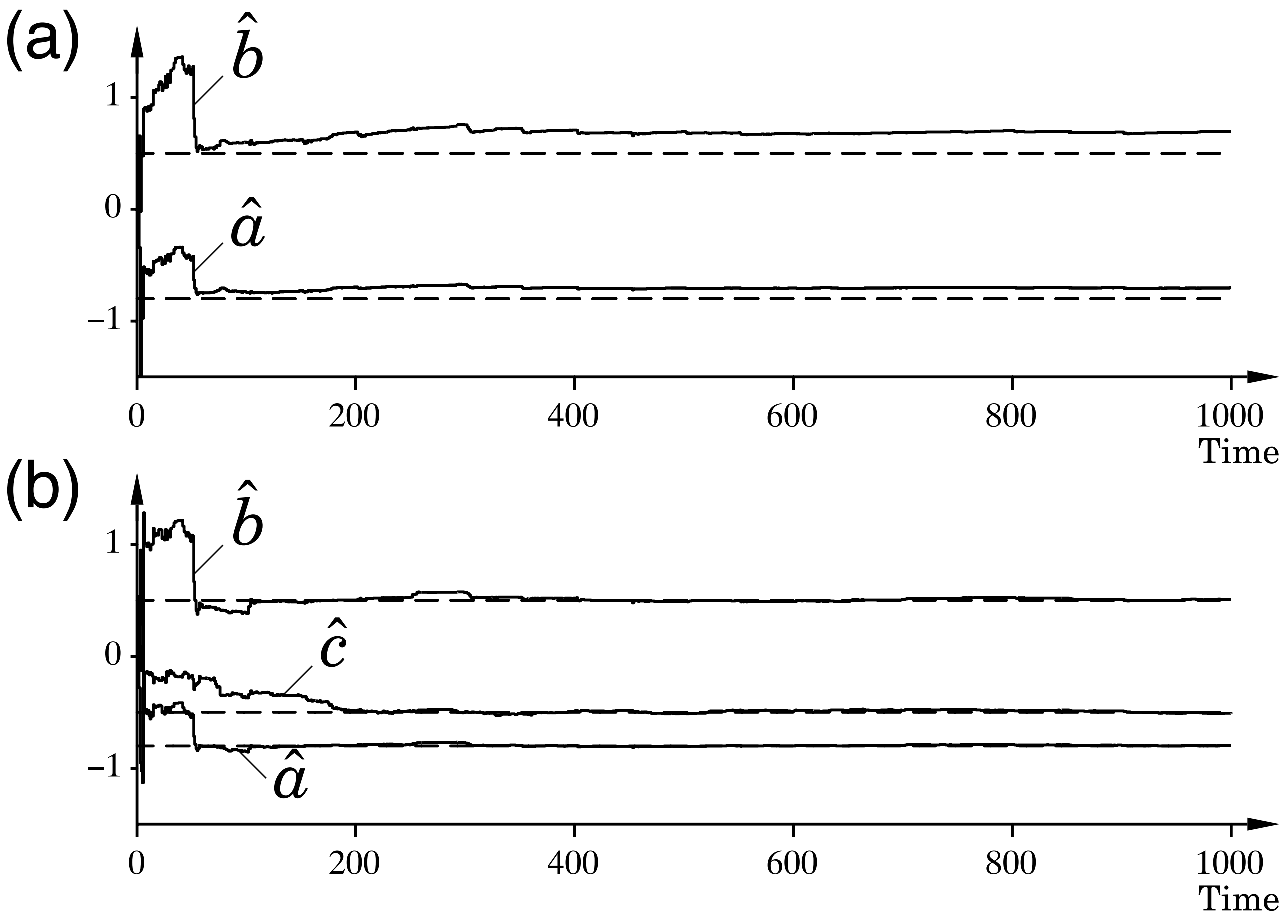


Weighting of forgetting factors
Weighting factor
1
0.8
0.6
0.4
0.2
0
-200
-150
-100
-50
0
Time [s]
Estimated parameter
2.5
2
1.5
1
0.5
0
0
50
100
150
200
Time [s]
Data sequence y
6
4
2
0
-2
0
50
100
150
200
Time [s]
P-matrix
1
0.8
0.6
0.4
0.2
0
0
50
100
150
200
Time [s]

# Colored Noise

Process model and Estimator model

$$
\begin{aligned}
y(t) - 0.8y(t-1) &= 0.5u(t-1) + e(t) - 0.5e(t-1) \\
y(t) + ay(t-1) &= bu(t-1) + e(t)
\end{aligned}
$$



# Conclusions

What you should remember:

- The least squares method
- The normal equations
- The recursive equations
- The matrix inversion lemma

What you should master:

- The recursive equations
- The role of excitation

Role in adaptive control

- Recursive estimation is a key part of adaptive control
- Recursive least squares is a useful method

# Innovations Model

It is also possible to represent the output process $y$ with only one noise source $w$.

One can find $K$ so that $y$ can be represented as

$$\begin{aligned} \hat{x}_{k+1} &= A\hat{x}_k + Kw_k \\ y_k &= C\hat{x}_k + w_k \end{aligned}$$

(This is obtained from the optimal Kalman predictor in Ch7)