Filtering and Identification

Day 1 - Lecture 2:
Random processes and Linear Least Squares

Michel Verhaegen

Division of Automatic Control - EE

Motivation for the course

State Space Models of the form:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + Ke(k) \quad e(k) \sim (0, R_e) \\ y(k) &= Cx(k) + Du(k) + e(k) \end{aligned}$$

are omni-present in the field of systems and control! This course is not only about using techniques/algorithms for systems and control problems, but $\cdots$ also about deriving them. Since,

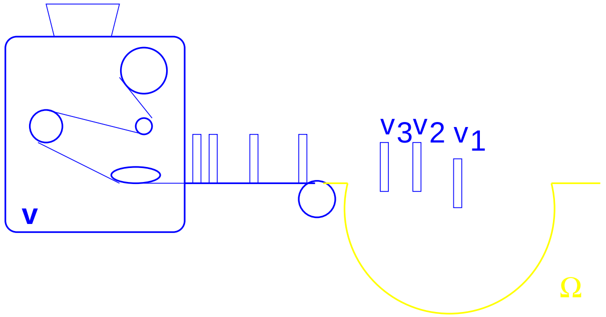
- the best use results when you know how the algorithms are derived!
- to prepare you in solving new problems in systems and control!

Overview

- **Characterization of random processes (RPs) in the time domain**
- Filtering a white noise sequence
- Deterministic least squares (LS) problems
- Statistical Properties LS estimates

Random variables, random signals/RPs

Random Variables (RVs)



$$\mathbf{v} \;\rightarrow\; \{v_1, v_2, \cdots\} = \Omega$$

(Sample Space)

Random Processes (RPs)

$$\{\mathbf{v_1}, \mathbf{v_2}, \cdots\}$$

$$\downarrow \quad \downarrow$$

$$\left\{\begin{matrix} v_{11} & v_{21} & \cdots \\ v_{12} & v_{22} & \\ \vdots & \vdots & \end{matrix}\right\}$$

Ergodicity

By the property of Ergodicity, a random process becomes like a random variable and is only characterized by a single time sequence $\{v(1), v(2), v(3), \cdots\}$.

Restriction to first and second moments

$$E[v] = \int_\Omega \alpha f_v(\alpha) d\alpha$$

$$\begin{aligned} E[vw] &= \int_\Omega \alpha\beta f_{vw}(\alpha,\beta) d\alpha d\beta \\ &= <\mathbf{v},\mathbf{w}> (in-product!) \end{aligned}$$

Independent, uncorrelated, orthogonal

RV1: $x(x_k)$ with p.d.f. $f_x(\alpha)$
RV2: $y(x_{k+1})$ with p.d.f. $f_y(\alpha)$ and $f_{xy}(\alpha, \beta)$

1 x, y Independent $\Leftrightarrow \; f_{xy}(\alpha, \beta) = f_x(\alpha) f_y(\beta)$

2 x, y Uncorrelated $\Leftrightarrow \; E[xy^*] = E[x]E[y^*]$

3 x, y Orthogonal $\Leftrightarrow \; E[xy^*] = 0$

Example

x, y jointly Gaussian with PDF

$$f_{xy}(\alpha,\beta) = Ae^{-\frac{1}{2}\left[\frac{(\alpha - m_x)}{\sigma_x} \quad \frac{(\beta - m_y)}{\sigma_y}\right]\begin{bmatrix} 1 & \rho_{xy} \\ \rho_{xy} & 1 \end{bmatrix}^{-1}\begin{bmatrix} \frac{(\alpha - m_x)}{\sigma_x} \\ \frac{(\beta - m_y)}{\sigma_y} \end{bmatrix}}$$

with $\rho_{xy} = \frac{E[(x - m_x)(y - m_y)^*]}{\sigma_x \sigma_y}$ and $A = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - \rho_{xy}^2}}$

Show that if $\rho_{xy} = 0$,

1 x, y are Independent

2 x, y are Uncorrelated

3 If in addition, $E[x] = 0$ or $E[y] = 0 \Rightarrow x, y$ are Orthogonal

Stationarity

Definition wide sense stationarity (WSS): A random process $\mathbf{x}(k) \in \mathbb{R}$ is WSS if the following **three** conditions are satisfied:

1. mean is constant, $m_x(k) = m_x$
2. auto-correlation function $R_x(k, \ell) = E[\mathbf{x}(k)\mathbf{x}(\ell)]$ only depends on the lag $k - \ell$
3. variance $E[(\mathbf{x}(k) - m_x)^2]$ is finite

White noise $n(k) \sim (0, \sigma_n^2)$

A zero-mean white noise sequence (ZMWN):
The random process $\mathbf{n}(k)$ is a ZMWN if it has mean zero and its auto-covariance (auto-correlation) function equals:

$$E[\mathbf{n}(k)\mathbf{n}(\ell)] = \begin{cases} \sigma_n^2 & \text{for } k = \ell \\ 0 & \text{otherwise} \end{cases}$$

Denoted as $R_n(\ell) = E[n(k)n(k-\ell)] = \sigma_n^2 \Delta(\ell)$, with $\Delta(\ell)$ the unit-pulse.

RPs in the time-domain

If the (real) RPs $x(k)$ and $y(k)$ are wide sense stationary (WSS), then these RPs are fully characterized in the time-domain by their means

$$E[x(k)] = m_x, \qquad E[y(k)] = m_y$$

and their auto-, cross-covariance functions:

$$\begin{aligned} C_x(\tau) &= E\Big[(x(k) - m_x)(x(k-\tau) - m_x)^T\Big] \\ C_{xy}(\tau) &= E\Big[(x(k) - m_x)(y(k-\tau) - m_y)^T\Big] \end{aligned}$$

RPs in the time-domain

An equivalent characterization is to replace the auto-, cross-covariance functions by the auto-, cross-correlation functions:

$$
\begin{aligned}
R_x(\tau) &= E\Big[x(k)x(k-\tau)^T\Big] = E\Big[x(k+\tau)x(k)^T\Big] \\
R_{xy}(\tau) &= E\Big[x(k)y(k-\tau)^T\Big]
\end{aligned}
$$

RPs in the time-domain

The numerical calculation may proceed via the assumption of ergodicity which enables to proof relationships like:

$$\Pr\left[\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}x(k)x(k-\tau)^T - R_x(\tau)\right] = 1$$

Overview

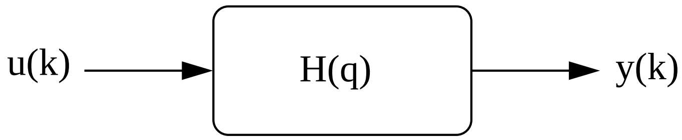
- Characterization of random processes (RPs) in the time domain
- **Filtering a white noise sequence**
- Deterministic least squares (LS) problems
- Statistical Properties LS estimates

Filtering of RPs

Let $u(k)$ be a zero-mean WSS RP, then we want to 'characterize' the RP $y(k)$ that is obtained by LTI filtering of $u(k)$

$$
\begin{aligned}
y(k) &= \sum_{p=-\infty}^{\infty} h(p)u(k-p) \\
&= \left(\sum_{p=-\infty}^{\infty} h(p)q^{-p} \right) u(k) \\
&= H(q)u(k)
\end{aligned}
$$

Calculation of the cross-covariance function



$$
\begin{aligned}
R_{yu}(\tau) &= E\Big[y(k)u(k-\tau)\Big] \\
&= E\Big[\sum_{p=-\infty}^{\infty} h(p)u(k-p)u(k-\tau)\Big] \\
&= \sum_{p=-\infty}^{\infty} h(p)E[u(k-p)u(k-\tau)] \\
&= \sum_{p=-\infty}^{\infty} h(p)R_u(\tau-p)
\end{aligned}
$$

$$
\boxed{R_{yu}(\tau) = h(\tau) * R_u(\tau)}
$$

Deconvolution Problem

Consider the input-output sequences $\{u(k), y(k)\}$ of the unknown LTI system $H(q)$, related as:

$$y(k) = \sum_{p=-\infty}^{\infty} h(p)u(k-p) + e(k) \quad e(k) \sim (0, R_e)$$

Given the covariance functions:

$$R_u(\tau), R_{yu}(\tau) \qquad \left(R_{yu}(\tau) = h(\tau) * R_u(\tau)\right)$$

Then the *deconvolution problem* is to determine $H(q) = \sum_{p=-\infty}^{\infty} h(p)q^{-p}$

Overview

- Characterization of random processes (RPs) in the time domain
- Filtering a white noise sequence
- **Deterministic least squares (LS) problems**
- Statistical Properties LS estimates

Deterministic least squares problem

Given $F \in \mathbb{R}^{N \times n}, y \in \mathbb{R}^N$, the problem is:

$$\min_x \epsilon^T \epsilon \quad \text{subject to:} \quad y = Fx + \epsilon$$

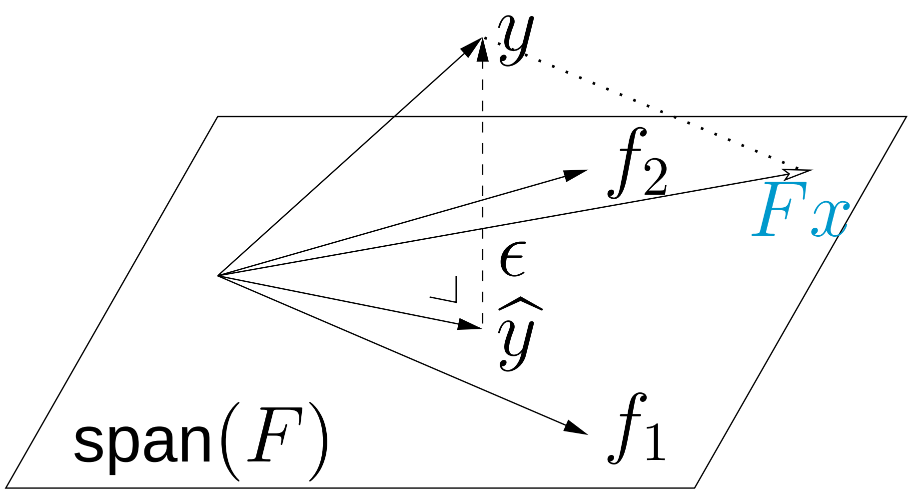
The argument that minimizes this problem is the least squares solution and is denoted as, $\widehat{x}$.

For all $x \in \mathbb{R}^n$, it satisfies,

$$\|y - F\widehat{x}\|_2^2 \leq \|y - Fx\|_2^2$$

Deterministic least squares problem

$$\|y - F\widehat{x}\|_2^2 \leq \|y - Fx\|_2^2$$



where $\widehat{y} = F\widehat{x} = \begin{bmatrix} f_1 & f_2 \end{bmatrix} \widehat{x}$

The classical solution

Lemma: Let the matrix F in

$$\min_x \epsilon^T \epsilon \quad \text{subject to:} \quad y = Fx + \epsilon$$

have full column rank, then the solution $\widehat{x}$ is:

$$\widehat{x} = \left(F^T F\right)^{-1} F^T y$$

This follows from the normal equations:

$$F^T F \widehat{x} \overset{\downarrow}{=} F^T y$$

Proof of the classical solution

Via the completion of squares. For all x and $\hat{x}$ satisfying:

$$(F^T F)\hat{x} = F^T y$$

we can write the least squares cost function as:

$$\begin{aligned}\|y - Fx\|_2^2 &= (y - Fx)^T(y - Fx) \\ &= y^T y - x^T F^T y - y^T F x + x^T F^T F x \\ &= y^T y - y^T F\hat{x} + (x - \hat{x})^T F^T F (x - \hat{x})\end{aligned}$$

Therefore,

$$\text{arg min}_x \|y - Fx\|_2^2 = \hat{x}$$

Proof of the classical solution

$$\|y - Fx\|_2^2 = \begin{bmatrix} 1 & x^T \end{bmatrix} \underbrace{\begin{bmatrix} y^T y & -y^T F \\ -F^T y & F^T F \end{bmatrix}}_{M} \begin{bmatrix} 1 \\ x \end{bmatrix}$$

$$M = \begin{bmatrix} I & -\widehat{x}^T \\ 0 & I \end{bmatrix} \begin{bmatrix} y^T y - y^T F \widehat{x} & \\ 0 & F^T F \end{bmatrix} \begin{bmatrix} I & 0 \\ -\widehat{x} & I \end{bmatrix},$$

for $\widehat{x}$ satisfying,

$$F^T F \widehat{x} = F^T y.$$

Proof of the classical solution (Ct'd)

$$\|y - Fx\|_2^2 = \begin{bmatrix} 1 & x^T \end{bmatrix} \begin{bmatrix} y^T y & -y^T F \\ -F^T y & F^T F \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix}$$

$$= \begin{bmatrix} 1 & x^T \end{bmatrix} \begin{bmatrix} I & -\widehat{x}^T \\ 0 & I \end{bmatrix} \begin{bmatrix} y^T y - y^T F\widehat{x} & \\ 0 & F^T F \end{bmatrix} \begin{bmatrix} I & 0 \\ -\widehat{x} & I \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix}$$

$$= \begin{bmatrix} 1 & (x - \widehat{x})^T \end{bmatrix} \begin{bmatrix} y^T y - y^T F\widehat{x} & \\ 0 & F^T F \end{bmatrix} \begin{bmatrix} 1 \\ x - \widehat{x} \end{bmatrix}$$

for $\widehat{x}$ satisfying $F^T F\widehat{x} = F^T y$.

$$\|y - Fx\|_2^2 = (y^T y - y^T F\widehat{x}) + (x - \widehat{x})^T F^T F(x - \widehat{x}).$$

Overview

- Characterization of random processes (RPs) in the time domain
- Filtering a white noise sequence
- Deterministic least squares (LS) problems
- **Statistical Properties LS estimates**

“Measurement Errors” $\epsilon \sim (0, I)$

Given $F \in \mathbb{R}^{N \times n}, y \in \mathbb{R}^N$, the problem is:

$$\min_x \epsilon^T \epsilon \quad \text{subject to:} \quad y = Fx + \epsilon$$

- F is a known full column rank matrix.
- x an unknown, deterministic vector.
- ϵ is a zero-mean random vector with $E[\epsilon\epsilon^T] = I$

Linear Estimators for the least squares problem

Least squares solution

$$\widehat{x} = \left(F^T F\right)^{-1} F^T y$$

is a linear estimator: it is linear in y.

Definition: *Linear estimator for* x given y has the form:

$$\widetilde{x} = My$$

with $M \in \mathbb{R}^{n \times N}$

Unbiased and minimum variance

The linear estimator $\widehat{x} = \widehat{M}y$ is unbiased if

$$E[\widehat{x} - x] = 0$$

The linear estimator $\widehat{x} = \widehat{M}y$ is called the minimum variance estimator if

$$E\left[(\widehat{x} - x)(\widehat{x} - x)^T\right] \leq E\left[(\widetilde{x} - x)(\widetilde{x} - x)^T\right]$$

for all linear estimators $\widetilde{x} = My$.

The Gauss-Markov theorem

The least squares solution

$$\widehat{x} = \widehat{M} y$$

with

$$\widehat{M} = \left(F^T F\right)^{-1} F^T$$

is an unbiased minimum variance estimate (UMVE).

Proof of the Gauss-Markov theorem

Linear Estimator which is UNBIASED

$$\widetilde{x} = My = MFx + M\epsilon \Rightarrow \widetilde{x} - x = (MF - I)x + M\epsilon$$

Consider the mean of $\widetilde{x} - x$

$$E[\widetilde{x} - x] = (MF - I_n)x + ME[\epsilon] = (MF - I)x$$

The linear estimator is unbiased provided,

$$E[\widetilde{x} - x] = 0 \Leftrightarrow \boxed{MF = I}$$

The least squares estimator $M = \left(F^T F\right)^{-1} F^T$ clearly satisfies $MF = I_n$

Proof of Minimum Variance Property

Recall

$$\widetilde{x} - x = (MF - I)x + M\epsilon = M\epsilon$$

Then, the covariance matrix of the Unbiased linear estimate $\widetilde{x} = My$ with M satisfying $MF = I$:

$$\begin{aligned} E\left[(\widetilde{x} - x)(\widetilde{x} - x)^T\right] &= ME[\epsilon\epsilon^T]M^T \\ &= MM^T \end{aligned}$$

For the least squares solution $\hat{x}(\hat{M} = (F^TF)^{-1}F^T)$, its covariance matrix equals,

$$\begin{aligned} E\left[(\widehat{x} - x)(\widehat{x} - x)^T\right] &= (F^TF)^{-1}F^TF(F^TF)^{-1} \\ &= (F^TF)^{-1} \end{aligned}$$

Proof of Minimum Variance Property (C'td)

We need to show that for M satisfying $MF = I$, it holds that

$$MM^T \geq (F^T F)^{-1}?$$

Alternatively,

$$MM^T - (F^T F)^{-1} \geq 0$$

The matrix $MM^T - (F^T F)^{-1}$ is equals to,

$$MM^T - MF(F^T F)^{-1} F^T M^T = M(I - F(F^T F)^{-1} F^T) M^T \geq 0?$$

Proof of Minimum Variance Property (C'td)

The inequality is shown provided that,

$$\forall x : x^T M(I - F(F^T F)^{-1} F^T) M^T x \geq 0$$

Let us call $M^T x$ the vector y. Then the above inequality holds, provided that,

$$\forall y : y^T (I - F(F^T F)^{-1} F^T) y \geq 0?$$

or,

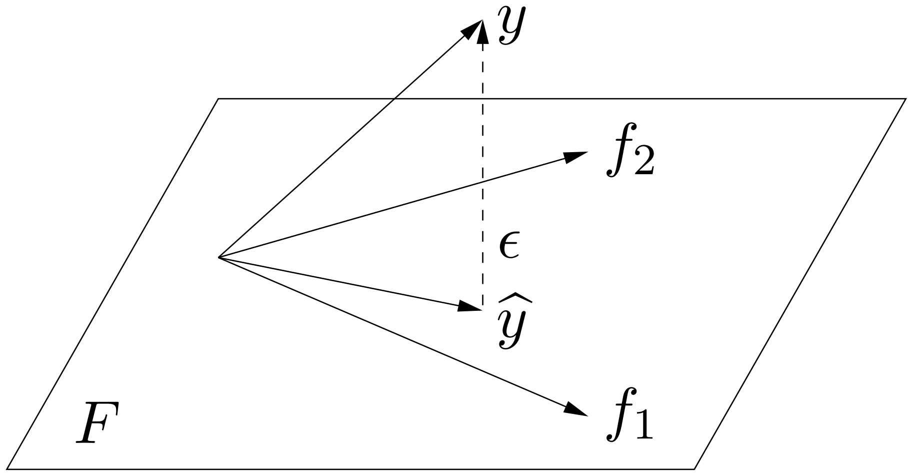
$$\forall y : y^T y - y^T F(F^T F)^{-1} F^T F \underbrace{(F^T F)^{-1} F^T y}_{?} \geq 0?$$

Proof of Minimum Variance Property (C'td)

Recall,

$$y^T y - y^T F(F^T F)^{-1} F^T \underbrace{F(F^T F)^{-1} F^T y}_{F\hat{x} = \hat{y}} \geq 0?$$

Here $F(F^T F)^{-1} F^T y$ equals the orthogonal projection $\hat{y}$ of y onto the column space of the matrix F.



The inequality follows from Pythagoras!

The weighted least squares problem

More general problem:

$$\min_x \epsilon^T \epsilon \quad \text{subject to} \quad y = Fx + L\epsilon,$$

with $L \in^{m \times m}$ a nonsingular matrix.

- ϵ: zero mean, covariance matrix $C_\epsilon = I_m$
- $\mu = L\epsilon$: zero-mean, covariance matrix $C_\mu = LL^T$

The weighted least squares problem

$$\min_x \epsilon^T \epsilon \quad \text{subject to} \quad L^{-1}y = L^{-1}Fx + \epsilon$$

taking $W = (LL^T)^{-1}$, we get the weighted least squares problem

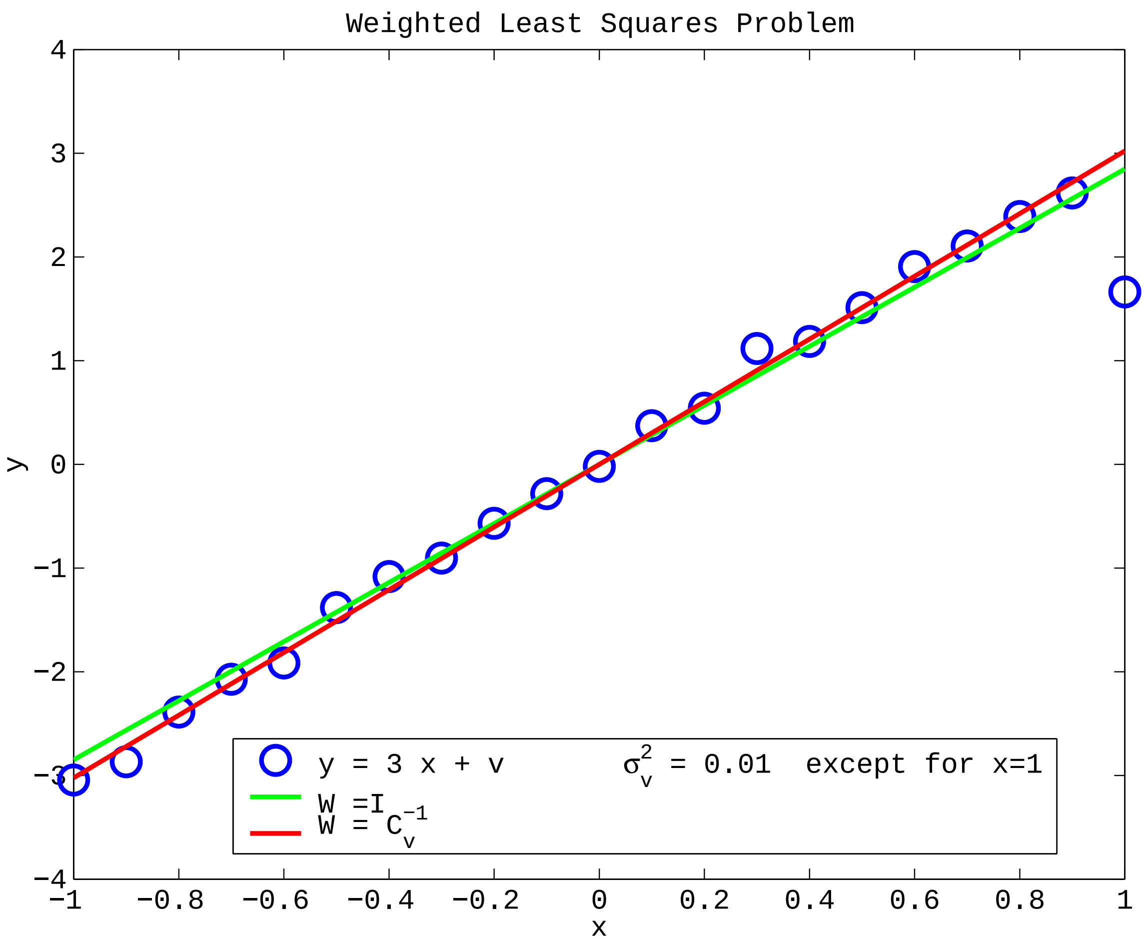
$$\min_x (Fx - y)^T W (Fx - y).$$

Solution:

$$\widehat{x} = (F^T W F)^{-1} F^T W y,$$

(by completion of the squares. . .)

Illustration WLS problem



Back to the numerics: The QR factorization

The QR-Theorem: Let $A \in \mathbb{R}^{m \times n}$ $(m \geq n)$, then there exists an orthogonal matrix $Q \in \mathbb{R}^{m \times m}$ that can be partitioned as:

$$Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \quad Q_1 \in \mathbb{R}^{m \times n}$$

such that,

$$Q^T A = \begin{bmatrix} R \\ 0 \end{bmatrix} \quad \text{with} \quad R \in \mathbb{R}^{n \times n} \quad \text{and} \quad R \ \text{upper-triangular}$$

[or the matrix A is factorized as $Q_1 R$.]

Comparison with SVD

Computational complexity of $m \times n$ matrix	
full-SVD (Golub-Reinsch): $2n^2(7m + 4n)$	full-QR (Householder): $2n^2(m - \frac{n}{3})$
Non-Iterative calculations	
SVD: No	QR: Yes
Rank Revealing:	
SVD: Yes	QR: No

rrqrdemo.m

QR Solution to LS problem

LSQR-Theorem: Consider the LS problem $\min_x \|y - Fx\|_2$ and consider the following QR factorization of F and the application of Q^T to y as,

$$F = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R \\ 0 \end{bmatrix} \quad \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} y = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

Consider the matrix F to have full column rank, then the LS solution $\widehat{x}$ and the LS residual satisfy:

$$\widehat{x} = R^{-1} d_1 \quad \|y - F\widehat{x}\|_2 = \|d_2\|_2$$

SensNeq.m

Summary of Lecture 2

- Refreshment of characterization of RPs in Time (- and Frequency) domain
- The linear least squares problem: unknown x deterministic.
- Gauss-Markov theorem (MVUE).
- More on Numerics (proofs for the afternoon).

This afternoon session

Preparation:

Study Chapters 2(2.6-2.7) and 4 (4.1 - 4.5.2)

Get Homework 1

Next lecture:

Addressing your questions on Homework 1.

This afternoon