Filtering and Identification

Day 2 - Lecture 1:
Stochastic least squares
Square Root Estimation

Michel Verhaegen

Division of Automatic Control - EE

Recap “goal” of the course

- Estimate linear static models from data:

$$\min_x \|y - Fx\|_W^2$$

  Relevant and to introduce the numerical fundaments of this course!

- Estimate the state of an LTI state space model:

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

  given the model (A, B, C, D) and K and input-output data.

- “Identify” state space model matrices (A, B, C, D, K) from input-output data $\{u(k), y(k)\}_{k=1}^N$.

Overview

- **Main Point Day 1**
- The Stochastic Least Squares Problem
- Updating least squares estimates: The RLS scheme
- A reliable square root solution

Main Point Day 1

The Gauss-Markov Theorem

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

$$\min_x \epsilon^T \epsilon \quad \text{subject to:} \quad y = Fx + L\epsilon \quad \epsilon \sim (0, I)$$

with x deterministic and L invertible, $W = (LL^T)^{-1}$, then,

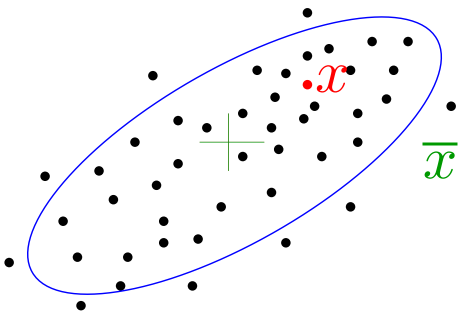
$$\begin{aligned} \text{argmin}_{\widetilde{x}=My} E[(x-\widetilde{x})(x-\widetilde{x})^T] &= (F^T W F)^{-1} F^T W y \\ &= \text{argmin}_x \epsilon^T \epsilon \end{aligned}$$

Overview

- Main Point Day 1
- **The Stochastic Least Squares Problem**
- Updating least squares estimates: The RLS scheme
- A reliable square root solution

Extension: x RV

Distribution Room Temp (T) at 2 time instances:



Legend:

- x | current unknown T
- $\overline{x}$ | mean value
- — | $1-\sigma$ uncertainty
- | ellipsoid

Statistical “Prior” information on x:

$$x \sim (\overline{x}, P) \quad P \geq 0$$

[In Gaussian setting: $f_x \doteq e^{-\frac{1}{2}(x-\overline{x})^T P^{-1}(x-\overline{x})}$]

Statistical Meaning:

$$E[x] = \overline{x} \quad E[(x-\overline{x})(x-\overline{x})^T] = P$$

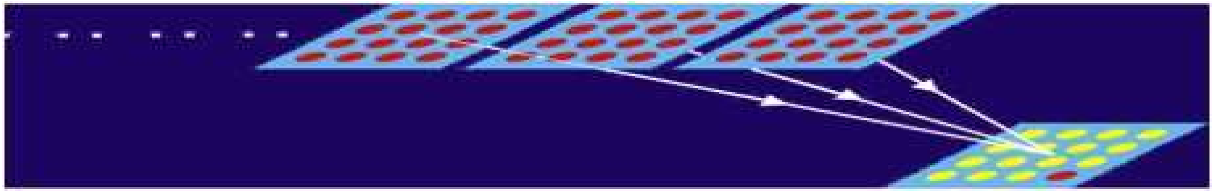
New measurement about x:

$$y = Fx + L\epsilon \quad \epsilon \sim (0, I) \;\; E[(x-\overline{x})\epsilon^T] = 0$$

How to combine these info?

Fusion of Wavefront sensor data

time $k-2$ time $k-1$ time k



Real-life scenario: From the Shack-Hartmann Wavefront (W) sensor data (at times $\leq k$), an estimate of the vectorized the 2D-W (call x) is given as:

$$x \sim (\overline{x}, P) \quad P > 0$$

At time instant $k+1$ we record new sensor data:

$$y(k+1) = Fx + L\epsilon(k+1) \quad \epsilon(k+1) \sim (0, I) \;\; E[(x-\overline{x})\epsilon(k+1)^T] = 0$$

Fuse this information with the estimate of x that you already have! Assumption?

The Stochastic Least Squares (SLS) Problem

Given the prior on the RV $x \sim (\overline{x}, P)$ with $P \geq 0$ and given the observations:

$$y = Fx + L\epsilon \quad \epsilon \sim (0, I) \quad E[(x - \overline{x})\epsilon^T] = 0$$

with F, L deterministic L full rank. Then seek among the linear estimators:

$$\widetilde{x} = \begin{bmatrix} M & N \end{bmatrix} \begin{bmatrix} y \\ \overline{x} \end{bmatrix}$$

The unbiased minimum variance estimate, i.e.:

$$E[(x - \widetilde{x})(x - \widetilde{x})^T] \quad \text{is minimized} \quad E[\widetilde{x}] = \overline{x}$$

Solution to the SLS Problem

Theorem SLS: Let the conditions of the SLS problem hold, let $W = (LL^T)^{-1}$ then the solution $\hat{x}$ to SLS is

$$\hat{x} = Ky + (I - KF)\overline{x}$$

with $K = PF^T(FPF^T + W^{-1})^{-1}$ and covariance matrix:

$$E[(x - \hat{x})(x - \hat{x})^T] = (P - PF^T(FPF^T + W^{-1})^{-1}FP)$$

If $P > 0$, we can rewrite K as $(P^{-1} + F^TWF)^{-1}F^TW$ and the covariance matrix as:

$$E[(x - \hat{x})(x - \hat{x})^T] = (P^{-1} + F^TWF)^{-1}$$

Sketch of the Proof SLS solution

Derivation of Numerical Procedure

1. The error of a linear estimator:

$$\tilde{x} = MFx + ML\epsilon + N\overline{x} \Rightarrow x - \tilde{x} = (I - MF)x - ML\epsilon - N\overline{x}$$

2. Unbiased Estimator:

$$E[x - \tilde{x}] = (I - MF)\overline{x} - N\overline{x} = 0 \Rightarrow MF + N = I$$

$$\Rightarrow x - \tilde{x} = (I - MF)(x - \overline{x}) - ML\epsilon$$

3. Covariance matrix: $(W^{-1} = LL^T)$ $\quad E[(x - \tilde{x})(x - \tilde{x})^T]$

$$\begin{aligned} &= (I - MF)E[(x - \overline{x})(x - \overline{x})^T](I - MF)^T + MLE[\epsilon\epsilon^T]L^TM \\ &= (I - MF)P(I - MF)^T + MW^{-1}M^T \end{aligned}$$

Minimizing the Covariance Matrix

$$
\begin{aligned}
E[(x-\tilde{x})(x-\tilde{x})^T] &= (I-MF)P(I-MF)^T + MW^{-1}M^T \\
&= \begin{bmatrix} I & -M \end{bmatrix} \underbrace{\begin{bmatrix} P & PF^T \\ FP & FPF^T + W^{-1} \end{bmatrix}}_{Q} \begin{bmatrix} I \\ -M^T \end{bmatrix}
\end{aligned}
$$

Using Lemma 2.3 p. 19, we can factorize Q as:

$$
\begin{bmatrix} I & PF^T(FPF^T+W^{-1})^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} P - PF^T(FPF^T+W^{-1})^{-1}FP & 0 \\ 0 & FPF^T + W^{-1} \end{bmatrix} [\bullet]^T
$$

Therefore,

$$
E[(x-\tilde{x})(x-\tilde{x})^T] = (P - PF^T(FPF^T+W^{-1})^{-1}FP) +
$$

$$
(PF^T(FPF^T+W^{-1})^{-1} - M)(FPF^T + W^{-1})(\bullet)^T
$$

Solution SLS — first part Theorem

The minimizing variance unbiased estimator is given by:

$$\hat{x} = My + \underbrace{(I - MF)}_{N}\overline{x}$$

with the optimal M given as

$PF^T(FPF^T + W^{-1})^{-1} = K$

The minimal covariance matrix is:

$$E[(x-\tilde{x})(x-\tilde{x})^T] = (P-PF^T(FPF^T+W^{-1})^{-1}FP)$$

Solution SLS — second part Theorem

Using the matrix inversion lemma:

$(A+BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1}+DA^{-1}B)^{-1}DA^{-1}$, we can write K as [a]:

$$
\begin{aligned}
K &= PF^T(W^{-1}+FPF^T)^{-1} \\
&= PF^T\Big(W - WF(P^{-1}+F^TWF)^{-1}F^TW\Big) \\
&= P\Big(I_n - F^TWF(P^{-1}+F^TWF)^{-1}\Big)F^TW \\
&= P\Big((P^{-1}+F^TWF) - F^TWF\Big)(P^{-1}+F^TWF)^{-1}F^TW \\
&= (P^{-1}+F^TWF)^{-1}F^TW = K
\end{aligned}
$$

The two gains K are equivalent on paper. What about inside the computer?

[a] Provided P is invertible!

Numerical Example

Consider the stochastic least squares problem with:

$$\overline{x} = 0 \quad P = 10^7 I_3 \quad L = 10^{-6} I_3 \quad x^T = \begin{bmatrix} 1 & -1 & 0.1 \end{bmatrix}$$

The matrices P and L indicate that the prior information is inaccurate and the relationship about x in y (the measurement) is very accurate. The matrix F is generated by the Matlab command:

$$F = \text{gallery('randsvd',3);}$$

Then the sample average of the relative error $\frac{\|\hat{x}-x\|_2}{\|x\|_2}$ for 100 trials with

1. $K = PF^T(FPF^T + W^{-1})^{-1}$ is $\boxed{0.0483}$
2. $K = (P^{-1} + F^T W F)^{-1} F^T W$ is $\boxed{0.1163}$

Recall the second part of the Theorem SLS

Theorem: Let the conditions of the SLS problem hold, let $W = (LL^T)^{-1}$ and $P > 0$ then the solution $\hat{x}$ to SLS is,

$$\hat{x} = Ky + (I - KF)\overline{x}$$

with $K = (P^{-1} + F^T W F)^{-1} F^T W$ and covariance matrix:

$$E[(x - \hat{x})(x - \hat{x})^T] = (P^{-1} + F^T W F)^{-1}$$

Interpretation results of Theorem SLS

1. $E[(x-\hat{x})(x-\hat{x})^T] = (P^{-1} + F^T W F)^{-1}$ is the resulting covariance matrix of fusing the two estimates:

$$x \sim (\overline{x}, P) \quad x \sim ((F^T W F)^{-1} F^T W y, (F^T W F)^{-1})$$

2. $\hat{x} = Ky + (I - KF)\overline{x}$ with $K = (P^{-1} + F^T W F)^{-1} F^T W$ indicates that: For $P^{-1} \to 0$ and $\Rightarrow \hat{x} = \hat{x}_{\mathrm{WLS}}$ no matter what $\overline{x}$ is.

3. It lays the fundaments for Recursive Least Squares (RLS).

Prior $x \sim (\hat{x}_k, P_k)$ with $\hat{x}_k$
UMVE
New Data:

$y_k = F_k x + L_k \epsilon_k$ → SLS → A Posteriori $x \sim (\hat{x}_{k+1}, P_{k+1})$

$\epsilon_k \sim (0, I)$

Overview

- Main Point Day 1
- The Stochastic Least Squares Problem
- **Updating least squares estimates: The RLS scheme**
- A reliable square root solution

The RLS Algorithm

Given initial estimate of the unknown RV x:

$$x \sim (\hat{x}_1, P_1)$$

For k = 1:end,

read data (y_k, F_k, W_k)

$$K_k = (P_k^{-1} + F_k^T W_k F_k)^{-1} F_k^T W_k$$

$$\hat{x}_{k+1} = (I - K_k F_k)\hat{x}_k + K_k y_k$$

$$P_{k+1}^{-1} = (P_k^{-1} + F_k^T W_k F_k)$$

end

Caution: This (information-matrix version) is not a tractable implementation from a numerical point of view.

RLS_demo.m

Overview

- Main Point Day 1
- The Stochastic Least Squares Problem
- Updating least squares estimates: The RLS scheme
- **A reliable square root solution**

A weighted least squares problem

Consider the prior in the SLS given as $x \sim (\overline{x}, P)$,
To define a WLS, we formulate this information into the generalized covariance representation:

$$x = \overline{x} + P^{1/2}\xi \quad \xi \sim (0, I)$$

with $P = P^{1/2}P^{T/2}$. Then we can combine this with the data equation in the WLS:

$$\min_x \nu^T \nu \quad \text{subject to} \quad \begin{bmatrix} \overline{x} \\ y \end{bmatrix} = \begin{bmatrix} I_n \\ F \end{bmatrix} x + \begin{bmatrix} -P^{1/2} & 0 \\ 0 & L \end{bmatrix} \nu$$

with $\nu \sim (0, I_n)$

Instruments to find a square root solution to SLS

What is allowable to define an “Equivalent” WLS problem?

The given WLS problem:

$$\min_x \nu^T \nu \quad \text{subject to} \quad \begin{bmatrix} \overline{x} \\ y \end{bmatrix} = \begin{bmatrix} I_n \\ F \end{bmatrix} x + \begin{bmatrix} -P^{1/2} & 0 \\ 0 & L \end{bmatrix} \nu$$

with $\nu \sim (0, I_n)$, is equivalent to:

$$\min_x (\nu')^T (\nu') \quad \text{subject to} \quad T_\ell \begin{bmatrix} \overline{x} \\ y \end{bmatrix} = T_\ell \begin{bmatrix} I_n \\ F \end{bmatrix} x + T_\ell \begin{bmatrix} -P^{1/2} & 0 \\ 0 & L \end{bmatrix} T_r (\underbrace{T_r^T \nu}_{\nu'})$$

with T_ℓ and T_r resp. invertible and orthogonal. Furthermore,

$\nu' \sim (0, I_n)$

Square Root Solution to SLS (1)

The Rational

Recall the given WLS problem:

$$\min_x \nu^T \nu \quad \text{subject to} \quad \begin{bmatrix} \overline{x} \\ y \end{bmatrix} = \begin{bmatrix} I_n \\ F \end{bmatrix} x + \begin{bmatrix} -P^{1/2} & 0 \\ 0 & L \end{bmatrix} \nu$$

with $\nu \sim (0, I_n)$ and $P^{1/2} \in \mathbb{R}^{n \times n}$. Then we use T_ℓ and T_r to "condense" the constraint into:

$$\begin{bmatrix} \star \\ \star \end{bmatrix} = \begin{bmatrix} 0 \\ I_n \end{bmatrix} x + \begin{bmatrix} \star & 0 \\ \star & P_{new}^{1/2} \end{bmatrix} \nu'$$

with $P_{new}^{1/2} \in \mathbb{R}^{n \times n}$.

Square Root Solution to SLS (2)

To achieve (part of) our goal we have to take T_ℓ as:

$$T_\ell \begin{bmatrix} I_n \\ F \end{bmatrix} = \begin{bmatrix} 0 \\ I_n \end{bmatrix} \quad \rightarrow T_\ell ?$$

A left (non-singular) transformation that does the job is:

$$T_\ell = \begin{bmatrix} F & -I \\ I & 0 \end{bmatrix} \quad \text{in} \quad T_\ell \begin{bmatrix} \overline{x} \\ y \end{bmatrix} = T_\ell \begin{bmatrix} I_n \\ F \end{bmatrix} x + T_\ell \begin{bmatrix} -P^{1/2} & 0 \\ 0 & L \end{bmatrix} T_r T_r^T \nu$$

It transforms the constraint set of equations to,

$$\begin{bmatrix} F\overline{x} - y \\ \overline{x} \end{bmatrix} = \begin{bmatrix} \boxed{0} \\ I \end{bmatrix} x + \begin{bmatrix} -FP^{1/2} & -L \\ -P^{1/2} & 0 \end{bmatrix} T_r T_r^T \nu$$

Important is the introduced “zero”!

Square Root Solution to SLS (3)

With the "zero" introduced, we now aim at finding part of ν' by the orthogonal transformation T_r. This is achieved by the following LQ factorization:

$$\begin{bmatrix} -FP^{1/2} & -L \\ -P^{1/2} & 0 \end{bmatrix} T_r = \begin{bmatrix} R & 0 \\ G & S \end{bmatrix}$$

It transforms the constraints into:

$$\begin{bmatrix} F\overline{x} - y \\ \overline{x} \end{bmatrix} = \begin{bmatrix} \boxed{0} \\ I \end{bmatrix} x + \begin{bmatrix} R & 0 \\ G & P_{new}^{1/2} \end{bmatrix} \begin{bmatrix} \kappa \\ \delta \end{bmatrix}$$

with $P_{new}^{1/2} \in \mathbb{R}^{n \times n}$.

Square Root Solution the SLS Problem

Theorem SR-SLS: Let the conditions of the SLS problem hold, let the following LQ factorization by given:

$$\begin{bmatrix} -FP^{1/2} & -L \\ -P^{1/2} & 0 \end{bmatrix} T_r = \begin{bmatrix} R & 0 \\ G & S \end{bmatrix}$$

with T_r orthogonal and the right hand side lower triangular, then the estimate of the SLS equals,

$$\hat{x} = GR^{-1}y + \left(I - GR^{-1}F\right)\overline{x}\left(= \text{argmin}_x \nu^T \nu^T\right)$$

and has covariance matrix,

$$E[(x-\hat{x})(x-\hat{x})^T] = P_{new}^{1/2}(P_{new}^{1/2})^T$$

Numerical Example (Ct'd)

Consider the stochastic least squares problem with:

$$\overline{x} = 0 \quad P = 10^7 I_3 \quad L = 10^{-6} I_3 \quad x^T = \begin{bmatrix} 1 & -1 & 0.1 \end{bmatrix}$$

The matrices P and L indicate that the prior information is inaccurate and the relationship about x in y (the measurement) is very accurate. The matrix F is generated by the Matlab command:

$$F = \text{gallery('randsvd',3);}$$

Then the sample average of the relative error $\frac{\|\hat{x}-x\|_2}{\|x\|_2}$ for 100 trials with

1. $K = GR^{-1}$ (square root solution) is 0.0002
2. $K = (P^{-1} + F^T W F)^{-1} F^T W$ ("worst" normal equation variant) is 0.1163

Summary of Lecture 1 - day 2

- The Stochastic Least squares problem is at the heart of many interesting filtering and estimation schemes $\rightarrow$ demonstrated for the RLS scheme
- Square Root Algorithms are the preferred numerical solutions.