EL3370 Mathematical Methods in Signals, Systems and Control

Topic 6: Least Squares Estimation

Cristian R. Rojas

Division of Decision and Control Systems
KTH Royal Institute of Technology

Outline

Outline

Hilbert Space of Random Variables

$x_1,\dots,x_n$: finite collection of random variables with $\mathrm{E}\{x_k^2\} < \infty$ for each i. Their *second order statistical information* is given by n *expected values*, $\mathrm{E}\{x_k\}$ ($k = 1,\dots,n$) and the *covariance matrix* $\mathrm{cov}\{x_1,\dots,x_n\} \in \mathbb{R}^{n\times n}$, whose jk-th entry is $\mathrm{E}\{[x_j - \mathrm{E}\{x_j\}][x_k - \mathrm{E}\{x_k\}]\}$.

Define a Hilbert space H of all linear combinations of the x_k's, with inner product $(x, y) := \mathrm{E}\{xy\}$. H has dimension at most n ($< \infty$).

Generalization

$x_1,\dots,x_n$: collection of m-dimensional random vectors with $\mathrm{E}\left\{\|x_k\|^2\right\} < \infty$ for each k.

Let $\mathscr{H}$ be the Hilbert space of all m-dimensional random vectors whose entries are linear combinations of the entries of $x_1,\dots,x_n$, *i.e.*, $x \in \mathscr{H}$ can be expressed as

$$x = K_1 x_1 + \cdots + K_n x_n, \qquad \text{where } K_1,\dots,K_n \in \mathbb{R}^{m\times m}.$$

The inner product of $\mathscr{H}$ is $(x, y) := \mathrm{E}\left\{x^T y\right\} = \mathrm{tr}\,\mathrm{E}\left\{x y^T\right\}$ ($x, y \in \mathscr{H}$).

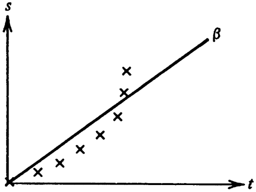
Outline

Least Square Estimate

Suppose that a vector y of measurements $(y_1, \ldots, y_m)$ is available, and we want to find a vector $\beta \in \mathbb{R}^n$ $(n < m)$ s.t. $y \approx W\beta$ in a minimum Euclidean norm sense, *i.e.*, s.t. $\|y - W\beta\|_2$ is minimum, where W is given.

To use the projection theorem, consider the Hilbert space $H = \mathbb{R}^m$, and the closed linear subspace

$$M = \{x \in H : x = W\beta \text{ for some } \beta \in \mathbb{R}^n\} = \mathscr{R}(W).$$



The minimizer β^{opt} should satisfy $(y - W\beta^{\text{opt}}, W\beta) = 0$ for all $\beta \in \mathbb{R}^n$, or

$$\beta^T W^T [y - W\beta^{\text{opt}}] = 0 \quad \text{for all } \beta \in \mathbb{R}^n,$$

i.e., $W^T y = W^T W \beta^{\text{opt}}$. Therefore, if the columns of W are l.i.:

$$\beta^{\text{opt}} = (W^T W)^{-1} W^T y. \qquad (\textit{Least squares solution})$$

Outline

Minimum Variance Estimates

Consider measurements $y = W\beta + \varepsilon$, where both β and ε are random vectors.
We want to minimize $\mathrm{E}\left\{\|\hat{\beta} - \beta\|_2^2\right\}$.

Theorem. Assume that $[\mathrm{E}\{yy^T\}]^{-1}$ exists. Then, the linear estimate $\hat{\beta}$ of β, based on y, minimizing $\mathrm{E}\left\{\|\hat{\beta} - \beta\|_2^2\right\}$ is $\hat{\beta} = \mathrm{E}\{\beta y^T\}[\mathrm{E}\{yy^T\}]^{-1}y$, with error covariance

$$\mathrm{E}\{[\hat{\beta} - \beta][\hat{\beta} - \beta]^T\} = \mathrm{E}\{\beta\beta^T\} - \mathrm{E}\{\beta y^T\}\left[\mathrm{E}\{yy^T\}\right]^{-1}\mathrm{E}\{y\beta^T\}.$$

Proof. Let $\hat{\beta} = Ky$, with $K \in \mathbb{R}^{n\times m}$. If we consider the Hilbert space H generated from the entries of y and β, and let $M = \mathrm{clin}\{y_1, \ldots, y_m\}$, the projection theorem gives $(\beta - \hat{\beta}) \perp M$, or $\mathrm{E}\{\beta_k y^T\} = \mathrm{E}\{K_k yy^T\} = K_k \mathrm{E}\{yy^T\}$ (where K_k is the k-th row of K), *i.e.*, $K = \mathrm{E}\{\beta y^T\}[\mathrm{E}\{yy^T\}]^{-1}$. □

Corollary. If $\mathrm{E}\{\varepsilon\varepsilon^T\} = Q \geq 0$, $\mathrm{E}\{\beta\beta^T\} = R \geq 0$, $\mathrm{E}\{\varepsilon\beta^T\} = 0$, with $WRW^T + Q \succ 0$, then $\hat{\beta} = RW^T(WRW^T + Q)^{-1}y = (W^TQ^{-1}W + R^{-1})^{-1}W^TQ^{-1}y$, with error covariance $R - RW^T(WRW^T + Q)^{-1}WR = (W^TQ^{-1}W + R^{-1})^{-1}$ (assuming $Q, R \succ 0$).

Minimum Variance Estimates (cont.)

Properties

1. *The minimum variance linear estimate of a linear function of β, e.g., $T\beta$, is $T\hat{\beta}$.*

   Proof. If Γy is the optimal estimate of $T\beta$, then the projection theorem gives $\mathrm{E}\{y(T\beta - \Gamma y)^T\} = 0$, or $\Gamma y = T\,\mathrm{E}\{\beta y^T\}[\mathrm{E}\{yy^T\}]^{-1} y = T\hat{\beta}$. □

2. *If $\hat{\beta}$ is the linear minimum variance estimate of β, then it is also the linear estimate minimizing $E\{(\hat{\beta}-\beta)^T P(\hat{\beta}-\beta)\}$ for every $P \succ 0$.*

   Proof. From property 1, $P^{1/2}\hat{\beta}$ is the minimum variance estimate of $P^{1/2}\beta$, *i.e.*, $\hat{\beta}$ minimizes $\mathrm{E}\{\|P^{1/2}\hat{\beta} - P^{1/2}\beta\|_2^2\} = \mathrm{E}\{(\hat{\beta}-\beta)^T P(\hat{\beta}-\beta)\}$. □

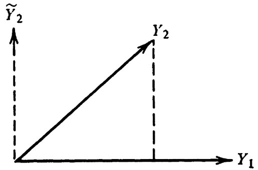
Properties (cont.)

3. *Let $\beta \in H$ (Hilbert space of random variables) and let $\hat{\beta}_1$ denote its orthogonal projection on a closed subspace Y_1 of H. Let y_2 be a vector of m random variables generating $Y_2 \subseteq H$, $\hat{y}_2$ the component-wise projection of y_2 into Y_1, and $\tilde{y}_2 := y_2 - \hat{y}_2$. Then, the projection of β into $Y_1 + Y_2$ is*

$$\hat{\beta} = \hat{\beta}_1 + \mathrm{E}\{\beta \tilde{y}_2^T\}[\mathrm{E}\{\tilde{y}_2 \tilde{y}_2^T\}]^{-1} \tilde{y}_2.$$

Proof

Let $\tilde{Y}_2$ be s.t. $\tilde{Y}_2 \perp Y_1$ and $Y_1 \oplus \tilde{Y}_2 = Y_1 + Y_2$.
Also, if Y_2 is generated by a finite set of vectors, $\tilde{Y}_2$ is generated by those vectors minus their projections into Y_1 (*why?*).
Since the projection into $Y_1 \oplus \tilde{Y}_2$ is equal to the projection into Y_1 plus the projection into $\tilde{Y}_2$, the result follows. □



Example

Assume we have an optimal estimate $\hat{\beta}$ of a random $\beta \in \mathbb{R}^n$, with $\mathrm{E}\{(\hat{\beta}-\beta)(\hat{\beta}-\beta)^T\} = R$. Given new measurements $y = W\beta + \varepsilon$, where ε has zero mean, covariance Q, and is uncorrelated with β and previous measurements, we want to update $\hat{\beta}$ to, say, $\hat{\hat{\beta}}$.

The best estimate of y based on past measurements is $\hat{y} = W\hat{\beta}$ (*why?*), so $\tilde{y} = y - W\hat{\beta} = W(\beta - \hat{\beta}) + \varepsilon$.

By property 3: $\hat{\hat{\beta}} = \hat{\beta} + \mathrm{E}\{\beta\tilde{y}^T\}[\mathrm{E}\{\tilde{y}\tilde{y}^T\}]^{-1}\tilde{y} = \hat{\beta} + RW^T[WRW^T + Q]^{-1}(y - W\hat{\beta})$.

The error covariance is: $\mathrm{E}\left\{\left(\hat{\hat{\beta}}-\beta\right)\left(\hat{\hat{\beta}}-\beta\right)^T\right\} = R - RW^T[WRW^T + Q]^{-1}WR.$ (*Exercise!*)

Outline

A *discrete random process* is a sequence (x_n) of random variables. (x_n) is *orthogonal* or *white* if $\mathrm{E}\{x_j x_k\} = \alpha_j \delta_{j-k}$, and *orthonormal* if, in addition, $\alpha_j = 1$ ($j \in \mathbb{N}$).

We assume that underlying an observed random process there is an orthonormal process.

Examples ($(u_k)_{k\in\mathbb{Z}}$: orthonormal process)

1. *Moving average*: $x_n = \sum_{k=1}^{\infty} a_k u_{n-k}$, where $\sum_{k=1}^{\infty} |a_k|^2 < \infty$.
2. *Autorregresive of order* 1: $x_n = a x_{n-1} + u_{n-1}$, $|a| < 1$.
   Notice that this process is equivalent to a moving average: $x_n = \sum_{k=1}^{\infty} a^{k-1} u_{n-k}$.
3. *Autorregresive of order* N: $x_n + a_1 x_{n-1} + \cdots + a_N x_{n-N} = u_{n-1}$,
   where the polynomial $s^N + a_1 s^{N-1} + \cdots + a_N$ has all its roots in the open unit disk.

Definition

An *n-dimensional state-space model* of a random process consists of:

1. *State equation*: $x_{k+1} = \Phi_k x_k + u_k$ $(k = 0, 1, \ldots)$, where x_k is an n-dimensional state (random) vector, $\Phi_k \in \mathbb{R}^{n \times n}$ is known, and u_k is an n-dimensional random vector of zero mean and $\mathrm{E}\{u_k u_l^T\} = Q_k \delta_{k-l}$.

2. *Initial random vector*: x_0 with an estimate $\hat{x}_0$ s.t. $\mathrm{E}\{(\hat{x}_0 - x_0)(\hat{x}_0 - x_0)^T\} = P_0$.

3. *Measurements*: $y_k = M_k x_k + w_k$ $(k = 0, 1, \ldots)$, where $M_k \in \mathbb{R}^{m \times n}$ is known, and w_k is an m-dimensional random measurement vector of zero mean and $\mathrm{E}\{w_k w_l^T\} = R_k \delta_{k-l}$, with $R_k \succ 0$.

In addition, assume that x_0, u_j and w_k are uncorrelated for all $j, k \geqslant 0$.

Estimation problem

Find the minimum variance estimate, $\hat{x}_{k|n}$, of x_k given measurements $y_0, \ldots, y_n$.

We will focus only on the *prediction* problem: to find $\hat{x}_{k+1|k}$.

Theorem (Kalman)

$\hat{x}_{k+1|k}$ can be computed recursively from:

$$\hat{x}_{k+1|k} = \Phi_k P_k M_k^T (M_k P_k M_k^T + R_k)^{-1} (y_k - M_k \hat{x}_{k|k-1}) + \Phi_k \hat{x}_{k|k-1},$$

where P_k is the covariance of $\hat{x}_{k|k-1}$, which can also be computed recursively from

$$P_{k+1} = \Phi_k P_k [I - M_k^T (M_k P_k M_k^T + R_k)^{-1} M_k P_k] \Phi_k^T Q_k.$$

The initial conditions for these equations are $\hat{x}_{0|-1} = \hat{x}_0$ and P_0.

Proof

Suppose that measurements $y_0, \ldots, y_{k-1}$ are available, as well as $\hat{x}_{k|k-1}$ and P_k, *i.e.*, we have the projection of x_k onto $Y_{k-1} := \text{clin}\{y_0, \ldots, y_{k-1}\}$.

The new measurement is $y_k = M_k x_k + w_k$. From the previous example, we have

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + P_k M_k^T (M_k P_k M_k^T + R_k)^{-1} (y_k - M_k \hat{x}_{k|k-1})$$

and covariance matrix $P_{k|k} = P_k - P_k M_k^T (M_k P_k M_k^T + R_k)^{-1} M_k P_k$.

Since $x_{k+1} = \Phi_k x_k + u_k$, and u_k is uncorrelated to v_k and x_k, Property 1 gives

$$\hat{x}_{k+1|k} = \Phi_k \hat{x}_{k|k},$$

with error covariance $P_{k+1} = \Phi_k P_{k|k} \Phi_k^T + Q_k$. □

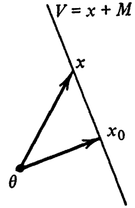
Outline

Dual Approximation Problem

The projection theorem can also be used to explicitly solve some infinite dimensional problems. To this end, we can restate it as:

Theorem (minimum norm problem)
Let M be a closed subspace of a Hilbert space H. Let $x \in H$, and the *linear variety* $V = x + M := \{x + m : m \in M\}$. Then there is a unique $x_0 \in V$ of minimum norm. Furthermore, $x_0 \perp M$.

Proof. Translate V by $-x$, so that V turns into M, and $\|x_0\|$ becomes $\|x_0 - x\|$, so that the projection theorem can be applied. □



Two types of varieties V are of interest: those with finite dimensional M, and those consisting of all $x \in H$ satisfying (for $y_1, \dots, y_n$ l.i.)

$$\begin{aligned} (x, y_1) &= c_1, \\ &\vdots \\ (x, y_n) &= c_n. \end{aligned} \qquad (V \text{ has } \textit{co-dimension } n.)$$

Theorem

Let $\{y_1, \ldots, y_n\}$ be l.i. vectors in a Hilbert space H, and $x_0 \in H$ the vector of minimum norm s.t. $(x, y_k) = c_k$ for $k = 1, \ldots, n$. Then $x_0 = \sum_{k=1}^{n} \beta_k y_k$, where the coefficients β_k satisfy

$$\begin{aligned} (y_1, y_1)\beta_1 + \cdots + (y_n, y_1)\beta_n &= c_1, \\ &\vdots \\ (y_1, y_n)\beta_1 + \cdots + (y_n, y_n)\beta_n &= c_n. \end{aligned} \qquad (*)$$

Proof. Let $M = \text{clin}\{y_1, \ldots, y_n\}$. The linear variety of vectors $x \in H$ satisfying $(x, y_k) = c_k$ for $k = 1, \ldots, n$ is a translation of $M^\perp$. Since $M^\perp$ is closed, existence and uniqueness of x_0 follow from the modified projection theorem (if $M^\perp \neq \{0\}$). Furthermore, $x_0 \perp M^\perp$, *i.e.*, $x_0 \in (M^\perp)^\perp$. Since M is closed, $(M^\perp)^\perp = M$, so $x_0 \in M$, and $x_0 = \sum_{k=1}^{n} \beta_k y_k$ for some coefficients β_k, which must satisfy the constraints $(x_0, y_k) = c_k$; this gives the system of equations $(*)$. □

Example

The shaft angular velocity ω of a DC motor driven by a current u satisfies
$\dot{\omega}(t) + \omega(t) = u(t)$.
The shaft angular position is θ (*i.e.*, $\dot{\theta} = \omega$). The motor is initially at rest: $\theta(0) = \omega(0) = 0$.
We want to find the current of minimum energy, $\int_0^1 u^2(t)dt$, that drives the motor to $\theta(1) = 1$, $\omega(1) = 0$.

This problem can be treated as a minimum norm problem in $L_2[0,1]$: By integration,

$$\omega(1) = \int_0^1 e^{t-1}u(t)dt = (u, y_1) \stackrel{!}{=} 0, \qquad y_1(t) = e^{t-1},$$

$$\theta(1) = \int_0^1 (1 - e^{t-1})u(t)dt = (u, y_2) \stackrel{!}{=} 1, \qquad y_2(t) = 1 - e^{t-1}.$$

According to the previous theorem, $u(t) = \beta_1 e^{t-1} + \beta_2(1 - e^{t-1})$, and by forcing the constraints,

$$u(t) = \frac{1}{3-e}(1 + e - 2e^t), \quad t \in [0,1].$$

Next Topic

Dual Spaces