EL3370 Mathematical Methods in Signals, Systems and Control

Topic 6: Estimation and Optimization in Hilbert Spaces

Cristian R. Rojas

Division of Decision and Control Systems
KTH Royal Institute of Technology

Outline

Outline

Hilbert Space of Random Variables

$x_1,\ldots,x_n$: 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*, $\mathrm{E}\{x_k\}$ ($k = 1,\ldots,n$) and the *covariance matrix* $\mathrm{cov}\{x_1,\ldots,x_n\} \in \mathbb{R}^{n\times n}$, 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

$x_1,\ldots,x_n$: 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 $x_1,\ldots,x_n$, *i.e.*, $x \in \mathscr{H}$ can be expressed as

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

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

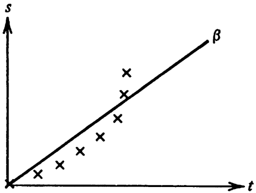
Outline

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 $M = \mathrm{lin}\{y_1, \dots, y_m\}$, which is closed, 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.*, $K = \mathrm{E}\{\beta y^T\}[\mathrm{E}\{yy^T\}]^{-1}$. $\square$

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 $\hat{\beta} = RW^T(WRW^T + Q)^{-1}y = (W^T Q^{-1} W + R^{-1})^{-1} W^T Q^{-1} y$, with error covariance $R - RW^T(WRW^T + Q)^{-1}WR = (W^T Q^{-1} W + R^{-1})^{-1}$ (assuming $Q, R \succ 0$).

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 $\Gamma y = T\,\mathrm{E}\{\beta y^T\}[\mathrm{E}\{yy^T\}]^{-1} y = T\hat{\beta}$. $\square$

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 $P \succ 0$.*

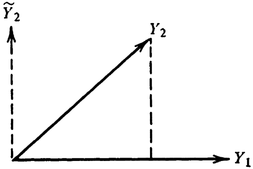
   Proof. From property 1, $P^{1/2}\hat{\beta}$ is the minimum variance estimate of $P^{1/2}\beta$, *i.e.*, $\hat{\beta}$ minimizes $\mathrm{E}\{\|P^{1/2}\hat{\beta} - P^{1/2}\beta\|_2^2\} = \mathrm{E}\{(\hat{\beta}-\beta)^T P(\hat{\beta}-\beta)\}$. $\square$

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. $\square$



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: $x_n = a x_{n-1} + u_{n-1}, \quad |a| < 1.$
   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*: $x_{k+1} = \Phi_k x_k + u_k$ $(k = 0, 1, \dots)$, 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*: $y_k = M_k x_k + w_k$ $(k = 0, 1, \dots)$, 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.

Recursive Estimation (cont.)

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 $Y_{k-1} := \text{lin}\{y_0, \ldots, y_{k-1}\}$.

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

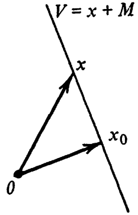
Outline

Minimum Norm Problems

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.)

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

Minimum Norm Problems (cont.)

When M has finite dimension, *e.g.*, $M = \{y_1, \ldots, n\}$, then x_0 is of the form $x_0 = x + \sum_{k=1}^{n} \beta_k y_k$ for some scalars β_k that satisfy the orthogonality conditions

$$\left(x + \sum_{k=1}^{n} \beta_k y_k, y_i \right) = 0, \qquad i = 1, \ldots, n,$$

or

$$\begin{aligned} (y_1, y_1)\beta_1 + \cdots + (y_n, y_1)\beta_n &= -(x, y_1), \\ &\vdots \\ (y_1, y_n)\beta_1 + \cdots + (y_n, y_n)\beta_n &= -(x, y_n), \end{aligned}$$

which is a system of n linear equations in n unknowns.

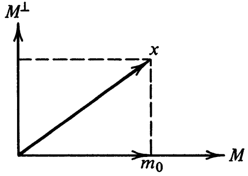
In the cases where V has finite co-dimension, the solution is given by the following result:

Theorem
Let $\{y_1,\dots,y_n\}$ 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 $k = 1,\dots,n$. Then $x_0 = \sum_{k=1}^n \beta_k y_k$, where the coefficients β_k satisfy

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

Proof. Let $M = \mathrm{lin}\{y_1,\dots,y_n\}$, which is closed. The linear variety of vectors $x \in H$ satisfying $(x, y_k) = c_k$ for $k = 1,\dots,n$ 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 $(*)$. $\square$

These two situations can be seen as duals of each other, because they are related, via translation, to the problem of projecting $x \in H$ into a linear subspace M of finite dimension (in the first case), or to its orthogonal complement $M^\perp$ (in the second case):



The situations are completely symmetrical, because $(M^\perp)^\perp = M$!

In both cases, since $H = M \oplus M^\perp$, x needs to be decomposed as $x = m_0 + m_0^\perp$, where $m_0 \in M$ and $m_0^\perp \in M^\perp$. As M has finite dimension, computing m_0 is a finite dimensional problem. Once m_0 is found, we directly obtain $m_0^\perp = x - m_0$!

See end of slides for a wider range of problems reducible to finite dimensions.

Minimum Norm Problems (cont.)

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,

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

Outline

Hilbert Space of Random Variables

Least Square Estimate

Minimum Variance Estimates

Recursive Estimation

Minimum Norm Problems

Optimization in RH_2

Bonus Slides

Motivation

Many problems in control, signal processing and communications can be posed as minimization problems in RH_2. This is because the variance of a discrete-time stationary process $x[k] = H(q)e[k]$, where e is white noise of variance λ and $H(q)$ is a stable rational transfer function (*i.e.*, in RH_2), can be written as

$$\text{var}\{x[k]\} = \text{E}\{x^2[k]\} = \frac{\lambda}{2\pi}\int_{-\pi}^{\pi} |H(e^{i\omega})|^2 d\omega = \lambda \|H\|_2^2.$$

Example. Let x and y be stationary processes with joint spectrum $\begin{bmatrix} \Phi_x(\omega) & \Phi_{xy}(\omega) \\ \Phi_{xy}(-\omega) & \Phi_y(\omega) \end{bmatrix}$.

The problem of finding a predictor $H \in RH_2$ that allows us to estimate $y[k]$ as $H(q)x[k]$ with minimum variance error corresponds to solving

$$\min_{H \in RH_2} \frac{1}{2\pi}\int_{-\pi}^{\pi} [H(e^{i\omega}) \; -1] \begin{bmatrix} \Phi_x(\omega) & \Phi_{xy}(\omega) \\ \Phi_{xy}(-\omega) & \Phi_y(\omega) \end{bmatrix} \begin{bmatrix} H(e^{-i\omega}) \\ -1 \end{bmatrix} d\omega = \|AH - B\|_2^2,$$

where $A \in RH_2$ and $B \in RL_2$ are s.t. $|A(e^{i\omega})|^2 = \Phi_x(e^{i\omega})$ and $A(e^{i\omega})B(e^{-i\omega}) = \Phi_{xy}(e^{i\omega})$ (obtained via spectral factorization).

To solve problems of the form $\min_{H \in RH_2} \|AH - B\|_2^2$, the following lemmas are useful:

Lemma. The orthogonal complement of RH_2 in RL_2 consists exactly of those $f \in RL_2$ which are analytic in $\mathbb{D}$ and s.t. $f(0) = 0$.

Proof. Let $g \in RH_2$ and $f \in RL_2$ s.t. f is analytic in $\mathbb{D}$ and $f(0) = 0$. Then, (g, f) is the negative of the sum of the residues of $g(z)f(z^{-1})/z$ at its poles in $\mathbb{E}$; since g is analytic in $\mathbb{E}$, as well as $z \mapsto f(z^{-1})/z$ (and at $z = \infty$, $\mathrm{Res}_{z=\infty} f(z^{-1})/z = -\lim_{z\to\infty} zf(z^{-1})/z = f(0) = 0$), we have that $(g, f) = 0$, so $f \in RH_2^\perp$. Conversely, if $f \in RL_2$ is not analytic in $\mathbb{D}$ nor s.t. $f(0) = 0$, it can be decomposed via partial fraction expansion as a sum of a constant plus positive powers and simple fractions. Let $\tilde{f}_1$ consist of the constant and fractions with poles in $\mathbb{E}$, and $\tilde{f}_2 := f - \tilde{f}_1$. Then, let $f_1 := \tilde{f}_1 + \tilde{f}_2(0) \neq 0$ and $f_2 := \tilde{f}_2 - \tilde{f}_2(0)$, so $f_1 \in RH_2$ and f_2 is analytic in $\mathbb{D}$ and s.t. $f_2(0) = 0$. From the previous argument, we have that $(f_1, f) = (f_1, f_1) > 0$, since $f_1 \neq 0$, thus $f \notin RH_2^\perp$ because it is not orthogonal to $f_1 \in RH_2$. □

Corollary. Every function $f \in RL_2$ can be decomposed as $f = f_1 + f_2$, where $f_1 \in RH_2$, $f_2 \in RH_2^\perp$, and $\|f\|_2^2 = \|f_1\|_2^2 + \|f_2\|_2^2$.

Proof. From the proof of the theorem, every $f \in RL_2$ can be decomposed as $f = f_1 + f_2$, where $f_1 \in RH_2$ and $f_2 \in RH_2^\perp$. The identity $\|f\|_2^2 = \|f_1\|_2^2 + \|f_2\|_2^2$ follows from Pythagoras' theorem. □

RL_∞ / RH_∞ spaces

RL_∞: normed space of real-rational functions, analytic in $\mathbb{T}$, with usual addition and scalar multiplication, and norm

$$\|f\|_\infty := \max_{\omega \in [-\pi,\pi]} |f(e^{i\omega})|, \qquad f \in RL_\infty.$$

RH_∞: subspace of RL_∞, of functions analytic in $\overline{\mathbb{E}}$, with same norm as RL_∞.

Note. If $f \in RL_2$ and $g \in RL_\infty$, then $fg \in RL_2$ and $\|fg\|_2 \leqslant \|g\|_\infty \|f\|_2$; similarly for $f \in RH_2$ and $g \in RH_\infty$.

Inner-Outer Factorization

Assume that $f \in RH_2$ has no zeros in $\mathbb{T}$. Then, f can be described as

$$f(z) = K \frac{\prod_{k=1}^{m_1} (z - z_k) \prod_{k=m_1+1}^{m} (z - z_k)}{\prod_{k=1}^{n} (z - p_k)}, \qquad K, z_1, \dots, z_m, p_1, \dots, p_n \in \mathbb{C}, \quad n, m \in \mathbb{N}_0,$$

where $|z_1|, \dots, |z_{m_1}| > 1$ and $|z_{m_1+1}|, \dots, |z_m| < 1$, so it can be decomposed as

$$f(z) = \underbrace{\frac{\prod_{k=1}^{m_1} (z - z_k)}{\prod_{k=1}^{m} (1 - z_k z)}}_{=: f_I(z)} \cdot \underbrace{K \frac{\prod_{k=1}^{m_1} (1 - z_k z) \prod_{k=m_1+1}^{m} (z - z_k)}{\prod_{k=1}^{n} (z - p_k)}}_{=: f_O(z)}.$$

This is the *inner-outer factorization* of f: the *inner function* $f_I \in RH_\infty$ has constant, non-zero modulus in $\mathbb{T}$, *i.e.*, $|f_I(e^{i\omega})| =$ constant for $\omega \in [-\pi, \pi]$, while the *outer function* f_O satisfies $f_O, 1/f_O \in RH_2$.

Note. If $f, g \in RL_2$, and $|g(e^{i\omega})| \equiv \alpha$ (constant), then $\|gf\|_2 = \alpha \|f\|_2$.

Theorem. Let $A \in RH_2$ and $B \in RL_2$, where A has no zeros in $\mathbb{T}$. Then,

$$\arg \min_{H \in RH_2} \|AH - B\|_2^2 = \frac{1}{A_O} P_{RH_2}\left(\frac{B}{A_I}\right),$$

where $A = A_I A_O$ is the inner-outer factorization of A, and $P_{RH_2}: RL_2 \to RH_2$ is the *projection operator onto* RH_2, $P_{RH_2}(f) := f_1$ for $f \in RL_2$ (where f_1 is defined as in the Corollary in Slide 24).

Proof. Note that $\|AH - B\|_2^2 = \|A_I A_O H - B\|_2^2 = \alpha \|A_O H - B/A_I\|_2^2$, where $\alpha = \|A_I\|_\infty^2 > 0$. Let $B/A_I = f_1 + f_2$, as in the Corollary of Slide 24, where $f_1 = P_{RH_2}(B/A_I)$, and noting that $A_O H \in RH_2$, we obtain by Pythagoras' theorem

$$\|AH - B\|_2^2 = \alpha \|A_O H - f_1 - f_2\|_2^2 = \alpha \|A_O H - f_1\|_2^2 + \alpha \|f_2\|_2^2 \geqslant \alpha \|f_2\|_2^2,$$

where equality in last step is attained iff $H = f_1/A_O = P_{RH_2}(B/A_I)/A_O$. This concludes the proof. $\square$

Remark. This result can be extended in several ways. *E.g.*, if $H \in RH_2$ is required to fulfil the interpolation constraint $H(x) = y$, for $x, y \in \mathbb{C}$, then H can be written as $H(z) = \frac{z-x}{z} \tilde{H}(z) + y$, where $\tilde{H} \in RH_2$ is arbitrary, so the cost function can be re-written as $\|AH - B\|_2^2 = \|\left(\frac{z-x}{z} A\right) \tilde{H} - (B - yA)\|_2^2$.
Also, the function $\|A_1 H - B_1\|_2^2 + \|A_2 H - B_2\|_2^2$ can be written as $\|AH - B\|_2^2 + \text{constant}$ by completion of squares.

Example

Let $y[k]$ be a signal described by the recursive equation $y[k] = e[k] + \alpha e[k-1]$, where $(e[k])$ is white noise of variance 1 and $|\alpha| > 1$. We want to predict $y[k]$, with error of minimum variance, based on past values of itself, as $\hat{y}[k] = H(q)y[k]$, where $H \in RH_2$ is s.t. $H(z) = h_1 z^{-1} + h_2 z^{-2} + \cdots$, *i.e.*, $H(\infty) = 0$.

Note that $H(\infty) = 0$ means that $H(z) = z^{-1}\tilde{H}(z)$, where $\tilde{H} \in RH_2$. Then,

$$\begin{aligned}
\mathrm{E}\{(y[k] - \hat{y}[k])^2\} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} |1 - H(e^{i\omega})|^2 \Phi_y(\omega) d\omega \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} |1 - H(e^{i\omega})|^2 |1 + \alpha e^{-i\omega}|^2 d\omega \\
&= \|(1 + \alpha q^{-1})[1 - q^{-1}\tilde{H}(q)]\|_2^2 \\
&= \|(1 + \alpha q^{-1})[q - \tilde{H}(q)]\|_2^2 \qquad \text{(multiplying cost by } q\text{, which has} \\
&= \left\| q + \alpha - \frac{q + \alpha}{q}\tilde{H}(q) \right\|_2^2 \qquad \text{constant modulus in } \mathbb{T}) \\
&= \left\| \frac{1 + \alpha q}{q + \alpha}\left(q + \alpha - \frac{q + \alpha}{q}\tilde{H}(q) \right) \right\|_2^2 \\
&= \left\| 1 + \alpha q - \frac{1 + \alpha q}{q}\tilde{H}(q) \right\|_2^2 = \alpha^2 \|q\|_2^2 + \left\| 1 - \frac{1 + \alpha q}{q}\tilde{H}(q) \right\|_2^2 \geqslant \alpha^2 \|q\|_2^2.
\end{aligned}$$

Optimization in RH_2 (cont.)

Example (cont.)

Also, $\|q\|_2^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega}e^{-i\omega}d\omega = 1$, so the lower bound becomes α^2, which is attained iff $1 - \frac{1+\alpha z}{z}\tilde{H}(z) \equiv 0$. This shows that $\tilde{H}^{\text{opt}}(z) = \frac{z}{1+\alpha z}$, so

$$H^{\text{opt}}(z) = \frac{1}{1+\alpha z},$$

so the optimal predictor satisfies the recursive equation $\alpha\hat{y}[k+1] + \hat{y}[k] = y[k]$.

Note that the minimum attainable variance for the prediction error, α^2, grows with $|\alpha|$!

Exercise: Repeat this example assuming that $|\alpha| < 1$, and show that the minimum variance does not depend on the value of α.

Next Topic

Dual Spaces

Outline

Bonus: Regularization and Representer Theorem

In data science, given a data set $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, where $x_i \in \mathscr{X}$ ($\mathscr{X}$ arbitrary) and $y_i \in \mathbb{R}$ ($i = 1, \ldots, n$), a standard problem is to find a function $f : \mathscr{X} \to \mathbb{R}$ s.t. $f(x_i) \approx y_i$ for all i. To address it, an approach consists in fixing a *feature map* $\psi : \mathscr{X} \to H$, where H is a Hilbert space, and solve (where $\lambda > 0$ is a fixed *regularization parameter*)

$$\min_{g \in H} \sum_{i=1}^{n} \|y_i - g(\psi(x_i))\|^2 + \lambda \|g\|^2, \tag{$\star$}$$

Then, $f = g \circ \psi$ provides the sought function. This is an infinite-dimensional problem, but in many instances it can be solved exactly, thanks to the following result.

Representer Theorem. Let $\Phi : H \to \mathbb{R}$ be a function on a Hilbert space H defined as $\Phi(f) := F(l_1(f), \ldots, l_n(f), \|f\|^2)$, where $l_1, \ldots, l_n \in H^*$, and $F : \mathbb{R}^{n+1} \to \mathbb{R}$ is monotonically increasing w.r.t. its last argument ($\|f\|^2$). Assume that there exists at least one $f^* \in H$ which attains the minimum of Φ over H. Then, there exists a minimizer of Φ of the form $\hat{f} = \sum_{i=1}^{n} \alpha_i y_i$, where y_i is s.t. $l_i(f) = (f, y_i)$ for $i = 1, \ldots, n$, and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$.

This result is most useful when H is a *reproducing kernel Hilbert space*, *i.e.*, a Hilbert space whose elements are real-valued functions on an arbitrary set $\mathscr{X}$, s.t. all the *evaluation functionals* $l_x : H \to \mathbb{R}$, defined as $l_x(f) = f(x)$ ($x \in \mathscr{X}$, $f \in H$), are bounded. For such a Hilbert space, the representer theorem can be directly applied to ($\star$) (*why?*).

Bonus: Regularization and Representer Theorem (cont.)

Proof (of Representer Theorem)
Let $M = \text{lin}\{y_1, \ldots, y_n\}$, and assume that the minimizer f^* does not belong to M. Then, $f^* = f_1 + f_2$, where $f_1 \in M$ and $f_2 \in M^\perp$, but

$$\begin{aligned} F(l_1(f^*), \ldots, l_n(f^*), \|f^*\|^2) &= F(l_1(f_1 + f_2), \ldots, l_n(f_1 + f_2), \|f_1\|^2 + \|f_2\|^2) \\ &= F(l_1(f_1), \ldots, l_n(f_1), \|f_1\|^2 + \|f_2\|^2) \\ &\leqslant F(l_1(f_1), \ldots, l_n(f_1), \|f_1\|^2), \end{aligned}$$

due to the monotonicity of F w.r.t. its last argument, and that $l_i(f_2) = (f_2, y_i) = 0$ (because $y_i \in M$ and $f_2 \in M^\perp$). Therefore, $f_1 \in M$ is a minimizer of Φ too. $\square$