Linear estimation of random processes

Mar-9-2016 - Review

Linear estimation

\[ X(t) \xrightarrow{\text{Channel}} Y(t) \xrightarrow{\text{Estimator}} \hat{X}(t) \]

LTI filter

- Let \( X(t) \) and \( Y(t) \) be jointly WSS with zero mean, auto correlation function \( R_X(t) \) and \( R_Y(t) \), and cross correlation function \( R_{XY}(t) = R_Y(t-T) \).
- We observe \( Y(s) \) for \( t-a \leq s \leq t+b \) and wish to estimate \( X(t) \) based on \( \{Y(s)\}_{s=t-a}^{t+b} \) such that the estimate \( \hat{X}(t) \) minimizes
  \[ \text{MSE} = E[(X(t) - \hat{X}(t))^2] \]
- We focus on the linear estimate of the form
  \[ \hat{X}(t) = \int_{t-a}^{t+b} h(u) Y(u) du \]
  \[ = \int_{-b}^{a} h(s) Y(t-s) ds \]

Finding the best \( h(t) \) is difficult for general \( a, b \).
But when \( a, b \to \infty \), then the solution is quite simple, and the problem is called "infinite smoothing." And when \( a \to \infty \), and \( b = 0 \), the solution is "Wiener filter" and the problem is often called "filtering."

By the orthogonality principle, \((X(t) - \hat{X}(t)) \perp Y(s) \) for \( t-a \leq s \leq t+b \).

\[ (X(t) - \hat{X}(t)) \perp Y(t-s) \] for \(-b \leq s \leq a \).

Hence, for \(-b \leq t \leq a\),

\[ E[(X(t) - \int_{-b}^{a} h(s) Y(t-s) ds) Y(t-T)] = 0 \]

or equivalently,

\[ R_{XY}(t) = \int_{-b}^{a} h(s) E[Y(t-s)Y(t-T)]ds \]

\[ = \int_{-b}^{a} h(s) R_{Y}(t-s) ds \] for \(-b \leq t \leq a\).
**Infinite smoothing**

Let $a, b \to \infty$. Then, the orthogonality condition becomes

$$R_{xy}(t) = \int_{-\infty}^{\infty} h(t) R_y(t-s) \, ds$$

$$= h(t) \ast R_y(t)$$

By taking Fourier transforms on both sides, we have

$$S_{xy}(f) = H(f) S_y(f).$$

Hence, the best linear estimator is given by $H(f) = \frac{S_{xy}(f)}{S_y(f)}$ and in time domain

$$h(t) = F^{-1}[H(f)] \quad (\text{cf. For scalar case, } h^* = \frac{\text{cov}(X,Y)}{\text{Var}(Y)})$$

The corresponding MSE is

$$E[(x(t) - \hat{x}(t))^2] = E[(x(t) - \hat{x}(t))x(t)]$$

$$= E[x^2(t)] - E[\hat{x}(t)x(t)].$$

To evaluate the second term, consider

$$R_{xx}(t) = E[x(t+u)x(t)]$$

$$= E[x(t) \int_{-\infty}^{\infty} h(s) y(t-s) \, ds]$$

$$= \int_{-\infty}^{\infty} h(s) E[x(t) y(t-s)] \, ds$$

$$= \int_{-\infty}^{\infty} h(s) R_{xy}(ts) \, ds$$

Back to the MSE calculation,

$$\text{MSE} = E[x^2(t)] - h(-t) \ast R_{xx}(t) \big|_{t=0}$$

Alternatively, let's use the power spectral densities and note

$$S_{xx}(f) = H(-f) S_{xy}(f)$$

$$= \frac{S_{xy}(f)}{S_y(f)} S_{xx}(f)$$

$$= \frac{|S_{xy}(f)|^2}{S_y(f)} \quad \text{and hence, } \text{MSE} = E[x^2(t)] - R_{xx}(0)$$

$$= \int_{-\infty}^{\infty} S_x(f) - \frac{|S_{xy}(f)|^2}{S_y(f)} \, df.$$
Example: Let \( X(t) \) and \( Z(t) \) be zero mean, uncorrelated WSS processes with power spectral densities \( S_x(f) \) and \( S_z(f) \). Let \( Y(t) = X(t) + Z(t) \).

Find the linear MMSE estimate of \( X(t) \) given \( \{Y(t)\}_{t=-\infty}^{\infty} \), that is,

\[
\hat{X}(t) = h(t) \ast X(t).
\]

Now \( R_Y(t) = R_X(t) + R_Z(t) \) and hence \( S_Y(f) = S_X(f) + S_z(f) \).

Similarly, \( R_{XY}(t) = R_X(t) \) and \( S_{XY}(f) = S_X(f) \).

Hence, the best estimation filter is given by \( H(f) = \frac{S_{XY}(f)}{S_Y(f)} = \frac{S_X(f)}{S_X(f) + S_z(f)} \) and the corresponding MSE is

\[
\int_{-\infty}^{\infty} \left( S_X(f) - \frac{|S_{XY}(f)|^2}{S_Y(f)} \right) df = \int_{-\infty}^{\infty} \left( S_X(f) - \frac{S_X^2(f)}{S_X(f) + S_z(f)} \right) df
\]

\[
= \int_{-\infty}^{\infty} \frac{S_X(f)S_z(f)}{S_X(f) + S_z(f)} df.
\]

Filtering \((a=\infty, b=0)\)

\( R_{XY}(t) = \int_0^\infty h(s) R_Y(t-s) ds \) or equivalently, find causal \( h(t) \) (i.e., \( h(t) = 0 \ Vt < 0 \)) such that \( R_{XY}(t) = \int_0^\infty h(s) R_Y(t-s) ds \). (Wiener-Hopf equation)

Solution: Wiener filter

\[
H(f) = \frac{1}{S_Y^+(f)} \left[ \frac{S_{XY}(f)}{S_Y(f)} \right]^*_+,
\]

where \( S_Y^+(f) \) and \( S_Y^-(f) \) form the causal-anticausal (cf. Cholesky factorization) decomposition of \( S_Y(f) = S_Y^+(f)S_Y^- (f) \) and \([S(f)]_+\) denotes the Fourier transform of the causal part of \( F^*[S(f)] \).
Basic probability theory

- Law of total probability

* Events: \( P(B) = \sum A_i P(A_i \cap B) \) if \( A_1, A_2, \ldots \) partition \( \Omega \)
* pmf: \( f_X(x) = \sum_y P_{X,Y}(x,y) \)
* pdf: \( f_X(x) = \int f_{X,Y}(x,y) \, dy \)
* mixed: \( f_X(x) = \int P_X(x) f_{Y|X}(y|x) \, dy \)
* \( P_X(x) = \int f_{X,Y}(x,y) \, dy \)

- Bayes rule

* Events: \( P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_i P(B|A_i)P(A_i)} \)
* pmf: \( P_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{\sum_x f_{X,Y}(x,y)} \)
* pdf: \( f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{\int f_{X,Y}(x,y) \, dx} \)

- Application: Signal detection (MAP, ML)

  - optimal: \( \hat{x} \)

- Independence

* Events: \( P(A_{i_1}, A_{i_2}, \ldots, A_{i_k}) = \prod_{j=1}^k P(A_{i_j}) \) for any \( i_1, \ldots, i_k \) and \( k \).
* cdf: \( F_{x_1, \ldots, x_k}(x_1, \ldots, x_k) = F_{x_1}(x_1) \cdots F_{x_k}(x_k) \)
* And similarly, for pmf and pdf.

- Conditional independence: \( X_1 \) and \( X_0 \) are conditionally independent given \( X_2 \) if

\[
  f_{X_1, X_0 | X_2}(x_1, x_0 | x_2) = f_{X_1 | X_2}(x_1 | x_2) f_{X_0 | X_2}(x_0 | x_2)
\]

Equivalently, \( f_{X_0 | X_1, X_2}(x_0 | x_1, x_2) = f_{X_0 | X_2}(x_0 | x_2) \).

X and Y are uncorrelated if \( \text{Cov}(X,Y) = E[(X-EX)(Y-EY)] = 0 \).

Independence \( \Rightarrow \) Uncorrelation, but not vice versa.

Uncorrelation \( \Rightarrow \) independence for Gaussian.
Function of random variables
- Conditional expectation \( E[g(X,Y)|X] \)
- Application: MMSE estimation \( \hat{X}(y) = E[X|Y=y] \) minimizes \( E[(X-\hat{X}(Y))^2] \).
- Probability inequalities: Cauchy-Schwarz, Jensen, Markov, Chebyshev...

Expectation and moments
- For two random variables
  - mean \( E(X) \)
  - second moment \( E(X^2) \)
  - variance \( E(X-E(X))^2 = E(X^2)-(E(X))^2 \)
  - correlation \( E(XY) \)
  - covariance \( E((X-E(X))(Y-E(Y))) = E(XY)-(E(X))(E(Y)) \)
- Similar notions for vectors and processes

- Linear estimation: scalars, vectors, and random processes.
- Convergence
  - Almost sure \( \Rightarrow \) in probability \( \Rightarrow \) in distribution
  - Mean square

Random processes
- IID
- Random walk
- Markov
- Independent increment
- Gaussian
- Gauss-Markov
- Brownian motion
- Poisson