

ECE531 Lecture 12: Linear Estimation and Causal Wiener-Kolmogorov Filtering

D. Richard Brown III

Worcester Polytechnic Institute

16-Apr-2009

Introduction

- ▶ General (static or dynamic) Bayesian MMSE estimation problem: Determine unknown state/parameter $X[\ell]$ from observations $Y[a], Y[a+1] \dots, Y[b]$ for some $b \geq a$.
 - ▶ Prediction: $\ell > b$.
 - ▶ Estimation: $\ell = b$.
 - ▶ Smoothing: $\ell < b$.

- ▶ General solution is just the conditional mean:

$$\hat{X}_{\text{mmse}}[\ell] = \text{E} \{X[\ell] | Y[a], \dots, Y[b]\}$$

- ▶ This problem is difficult to compute in general. Two approaches to getting useful results:
 - ▶ **Restrict the model**, e.g. linear and white Gaussian. This led to a closed-form batch solution and the discrete-time Kalman-Bucy filter.
 - ▶ **Restrict the class of estimators**, e.g. unbiased or **linear** estimators.
- ▶ **Linear estimators** are particularly interesting since they are computationally convenient and require only partial statistics.

Jointly Gaussian MMSE Estimation

Suppose that $X[\ell], Y[a], \dots, Y[b]$ are jointly Gaussian random vectors. We know that

$$\begin{aligned} \hat{X}_{\text{mmse}}[\ell] &= \mathbb{E}\{X[\ell] \mid Y[a], \dots, Y[b]\} \\ &= \mathbb{E}\{X[\ell] \mid Y_a^b\} \\ &= \mathbb{E}\{X[\ell]\} + \text{cov}\{X[\ell], Y_a^b\} [\text{cov}\{Y_a^b, Y_a^b\}]^{-1} (Y_a^b - \mathbb{E}\{Y_a^b\}) \\ &= H^\top[a, b, \ell] Y_a^b + c[a, b, \ell] \end{aligned}$$

where $H[a, b, \ell]$ is a matrix with dimensions $(b - a + 1)k \times m$ and $c[a, b, \ell]$ is a vector of dimension $m \times 1$.

Note that the MMSE estimate in this case is simply a linear (affine) combination of the “super-vector” of observations Y_a^b plus some constant.

When $X[\ell], Y[a], \dots, Y[b]$ are jointly Gaussian random vectors, **the MMSE estimator is a linear estimator**. In this particular case, the restriction to linear estimators results in no loss of optimality.

Basics of Linear Estimators

A linear estimator for the state/parameter $X[\ell]$ given the observations $Y[a], Y[a+1], \dots, Y[b]$ for some $b \geq a$ is simply a collection of linear weights applied to the observations with possibly a constant offset not depending on the observations. When $b - a$ is finite, we can write

$$\hat{X}[\ell] = H^\top[a, b, \ell] Y_a^b + c[a, b, \ell] \in \mathbb{R}^m$$

Remarks:

- ▶ This expression is for general vector-valued parameters/states.
- ▶ Note that the k th element of the estimate is affected only by the k th row of $H^\top[a, b, \ell]$ and the k th element of $c[a, b, \ell]$:

$$\hat{X}_k[\ell] = h_k^\top[a, b, \ell] Y_a^b + c_k[a, b, \ell] \in \mathbb{R}$$

- ▶ Since we can pick the coefficients in each row of $H^\top[a, b, \ell]$ and $c[a, b, \ell]$ without affecting the other estimates, we can focus here without any loss of generality on the case when $\hat{X}[\ell]$ is a scalar.

Linear MMSE Estimation

For notational convenience, we will write the scalar linear estimator for the parameter/state $X_k[\ell]$ from now on as

$$\hat{X}[\ell] = h^\top Y_a^b + c \quad (1)$$

where c is a scalar and h is a vector with $(b - a + 1)k < \infty$ elements.

Let \mathcal{H} denote the set of all linear estimators for (1). The linear MMSE (LMMSE) estimator is pretty easy to compute from (1):

$$\begin{aligned} \hat{X}_{\text{lmmse}}[\ell] &= \arg \min_{\{h,c\} \in \mathcal{H}} \mathbb{E} \left\{ \left(\hat{X}[\ell] - X[\ell] \right)^2 \right\} \\ &= \arg \min_{\{h,c\} \in \mathcal{H}} \mathbb{E} \left\{ \left(h^\top Y_a^b + c - X[\ell] \right)^2 \right\} \\ &= \arg \min_{\{h,c\} \in \mathcal{H}} \mathbb{E} \left\{ (h^\top Y_a^b + c)^2 - 2(h^\top Y_a^b + c)X[\ell] + X^2[\ell] \right\} \\ &= \arg \min_{\{h,c\} \in \mathcal{H}} \mathbb{E} \left\{ (h^\top Y_a^b + c)^2 \right\} - 2\mathbb{E} \left\{ (h^\top Y_a^b + c)X[\ell] \right\} \end{aligned}$$

Linear MMSE Estimation (continued)

Picking up where we left off...

$$\begin{aligned}\hat{X}_{\text{Immse}}[\ell] &= \arg \min_{\{h,c\} \in \mathcal{H}} \mathbb{E} \left\{ (h^\top Y_a^b + c)^2 \right\} - 2\mathbb{E} \left\{ (h^\top Y_a^b + c)X[\ell] \right\} \\ &= \arg \min_{\{h,c\} \in \mathcal{H}} h^\top \mathbb{E} \left\{ Y_a^b (Y_a^b)^\top \right\} h + 2ch^\top \mathbb{E} \left\{ Y_a^b \right\} + c^2 \\ &\quad - 2 \left[h^\top \mathbb{E} \left\{ Y_a^b X[\ell] \right\} + c\mathbb{E} \left\{ X[\ell] \right\} \right]\end{aligned}$$

What should we do now? Let's take the gradient with respect to $[h, c]^\top$ and set it equal to zero...

$$\begin{bmatrix} 2\mathbb{E} \left\{ Y_a^b (Y_a^b)^\top \right\} h + 2c\mathbb{E} \left\{ Y_a^b \right\} - 2\mathbb{E} \left\{ Y_a^b X[\ell] \right\} \\ 2\mathbb{E} \left\{ (Y_a^b)^\top \right\} h + 2c - 2\mathbb{E} \left\{ X[\ell] \right\} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The second equation implies that $c = \mathbb{E}\{X[\ell]\} - \mathbb{E}\{(Y_a^b)^\top\}h$. Plug this into first equation and solve for h ...

Linear MMSE Estimation (continued)

First equation:

$$2\mathbb{E} \left\{ Y_a^b (Y_a^b)^\top \right\} h + 2c \mathbb{E} \left\{ Y_a^b \right\} - 2\mathbb{E} \left\{ Y_a^b X[\ell] \right\} = 0$$

Plug in $c = \mathbb{E}\{X[\ell]\} - \mathbb{E}\{(Y_a^b)^\top\} h \dots$

$$\mathbb{E} \left\{ Y_a^b (Y_a^b)^\top \right\} h + \mathbb{E} \left\{ Y_a^b \right\} \mathbb{E} \left\{ X[\ell] \right\} - \mathbb{E} \left\{ Y_a^b \right\} \mathbb{E} \left\{ (Y_a^b)^\top \right\} h - \mathbb{E} \left\{ Y_a^b X[\ell] \right\} = 0$$

Collecting terms and recognizing that

$$\text{cov}(Y, Y) = \mathbb{E}\{YY^\top\} - \mathbb{E}\{Y\}\mathbb{E}\{Y^\top\}$$

$$\text{cov}(Y, X) = \mathbb{E}\{YX\} - \mathbb{E}\{Y\}\mathbb{E}\{X\} \quad (\text{when } X \text{ is a scalar})$$

we can write

$$h_{l\text{mmse}} = \left[\text{cov}(Y_a^b, Y_a^b) \right]^{-1} \text{cov}(Y_a^b, X[\ell])$$

Sanity check: What are the dimensions of $h_{l\text{mmse}}$?

Linear MMSE Estimation (continued)

Putting it all together, we have

$$\begin{aligned}
 \hat{X}_{\text{lmmse}}[\ell] &= h_{\text{lmmse}}^{\top} Y_a^b + c \\
 &= h_{\text{lmmse}}^{\top} Y_a^b + \mathbf{E}\{X[\ell]\} - h_{\text{lmmse}}^{\top} \mathbf{E}\{Y_a^b\} \\
 &= \mathbf{E}\{X[\ell]\} + h_{\text{lmmse}}^{\top} (Y_a^b - \mathbf{E}\{Y_a^b\}) \\
 &= \mathbf{E}\{X[\ell]\} + \text{cov}(X[\ell], Y_a^b) [\text{cov}(Y_a^b, Y_a^b)]^{-1} (Y_a^b - \mathbf{E}\{Y_a^b\})
 \end{aligned}$$

This should look familiar.

- ▶ It should be clear that $\hat{X}_{\text{lmmse}}[\ell] = \hat{X}_{\text{mmse}}[\ell]$ when $X[\ell], Y[a], \dots, Y[b]$ are jointly Gaussian.
- ▶ Computation of $\hat{X}_{\text{lmmse}}[\ell]$ only requires knowledge of the observation/state means and covariances (second order statistics). This is much more appealing than requiring full knowledge of the joint distributions.
- ▶ Since the role of c is only to compensate for any non-zero mean of $\hat{X}_{\text{lmmse}}[\ell]$ and any non-zero mean of Y_a^b , we can assume without loss of generality that $\mathbf{E}\{X[\ell]\} \equiv 0$ and $\mathbf{E}\{Y_a^b\} \equiv 0$ (and hence $c = 0$) from now on.

Extension to Vector LMMSE State Estimator

Our result for the scalar LMMSE state estimator ($X[\ell] \in \mathbb{R}$) :

$$h_{\text{lmmse}} = \left[\text{cov}(Y_a^b, Y_a^b) \right]^{-1} \underbrace{\text{cov}(Y_a^b, X[\ell])}_{\text{vector}}$$

The vector LMMSE state estimator ($X[\ell] \in \mathbb{R}^m$) can be written as

$$H_{\text{lmmse}} := [h_{\text{lmmse}}^0, \dots, h_{\text{lmmse}}^{m-1}]$$

and we can write

$$\begin{aligned} H_{\text{lmmse}} &= \left[\text{cov}(Y_a^b, Y_a^b) \right]^{-1} \left[\text{cov}(Y_a^b, X_0[\ell]), \dots, \text{cov}(Y_a^b, X_{m-1}[\ell]) \right] \\ &= \left[\text{cov}(Y_a^b, Y_a^b) \right]^{-1} \underbrace{\text{cov}(Y_a^b, X[\ell])}_{\text{matrix}} \end{aligned}$$

where $X[\ell] = [X_0[\ell], \dots, X_{m-1}[\ell]]^\top$.

Remarks

- ▶ Note that the only assumptions/restrictions that we have made so far in the development of the LMMSE estimator are
 1. We only consider linear (affine) estimators.
 2. We only consider the MSE cost assignment.
 3. The number of observations $b - a + 1$ is finite.
 4. The matrix $\text{cov}(Y_a^b, Y_a^b)$ is invertible.
- ▶ Other than some mild regularity conditions (finite expectations for the observations and the state), we have made no assumptions about the statistical nature of the observations or the unknown state/parameter.
- ▶ Hence, this result is quite general.
- ▶ Practicality of direct implementation is limited somewhat by the requirement to compute a matrix inverse. The Kalman filter (and its extensions) alleviates this problem in some cases.
- ▶ What can we say about the case when $\text{cov}(Y_a^b, Y_a^b)$ is not invertible?

Covariance Matrix and MSE of Linear Estimation

In the general vector parameter case, let

$$\begin{aligned} C_{XX} &:= \text{cov}(X[\ell], X[\ell]) \\ C_{XY} &:= \text{cov}(X[\ell], Y_a^b) = C_{YX}^\top \\ C_{YY} &:= \text{cov}(Y_a^b, Y_a^b) \end{aligned}$$

We can write the covariance matrix of the LMMSE estimator as

$$\begin{aligned} C_{\text{lmmse}} &:= \mathbb{E} \left\{ (X[\ell] - \hat{X}[\ell])(X[\ell] - \hat{X}[\ell])^\top \right\} \\ &= \mathbb{E} \left\{ \left[X[\ell] - \mathbb{E}\{X[\ell]\} - C_{XY}C_{YY}^{-1} \left(Y_a^b - \mathbb{E}\{Y_a^b\} \right) \right] [\text{same}]^\top \right\} \\ &= C_{XX} - 2C_{XY}C_{YY}^{-1}C_{YX} + C_{XY}C_{YY}^{-1}C_{YY}C_{YY}^{-1}C_{YX} \\ &= C_{XX} - C_{XY}C_{YY}^{-1}C_{YX} \end{aligned}$$

The variance of each individual parameter estimate is on the diagonal of this covariance matrix. What can we say about the overall MSE?

Two-Coefficient Scalar LMMSE Estimation

Consider a two-coefficient time-invariant linear estimator $h = [h_0, h_1]^\top$ for the scalar state $X[\ell]$ based on the scalar observations $Y[\ell - 1]$ and $Y[\ell]$. Also assume that $X[\ell]$ and $Y[\ell]$ are wide-sense stationary (and cross-stationary) and that $E\{X[\ell]\} = E\{Y[\ell]\} = 0$ such that $c = 0$.

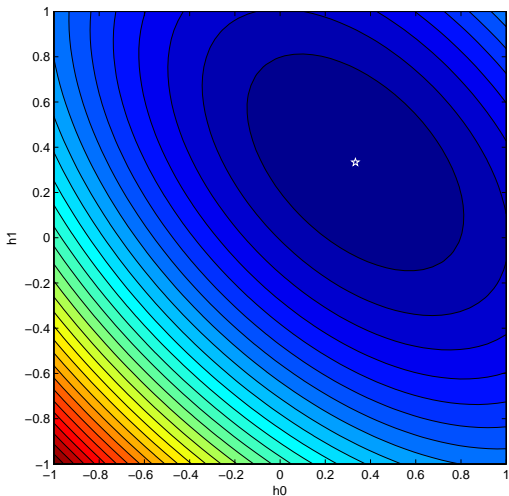
The MSE of the linear estimator h can be written as

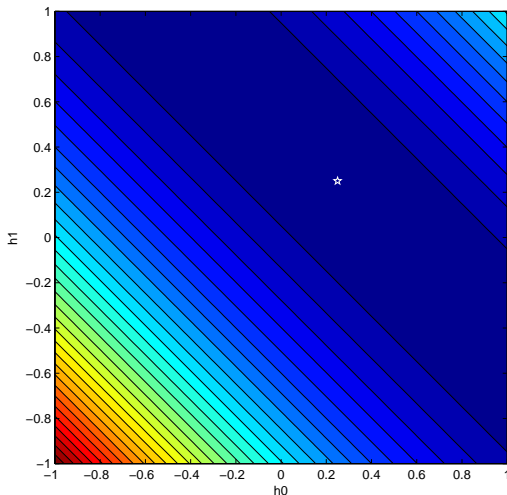
$$\begin{aligned} \text{MSE} &:= E\left\{(\hat{X}[\ell] - X[\ell])^2\right\} \\ &= h^\top E\left\{Y_a^b (Y_a^b)^\top\right\} h - 2h^\top E\left\{Y_a^b X[\ell]\right\} + E\left\{X^2[\ell]\right\} \\ &= h^\top C_{YY} h - 2h^\top C_{YX} + \sigma_X^2 \end{aligned}$$

An LMMSE estimator must satisfy $C_{YY} h_{\text{lmmse}} = C_{YX}$. When C_{YY} is invertible, the unique LMMSE estimator is simply

$$h_{\text{lmmse}} = C_{YY}^{-1} C_{YX}.$$

Since we only have two coefficients in our linear estimator, we can plot the MSE as a function of these coefficients to get some intuition...

Unique LMMSE Solution (C_{YY} is invertible)

Non-Unique LMMSE Solution (C_{YY} is not invertible)

Hint: Use Matlab's `pinv` function to find the minimum norm solution.

Infinite Number of Observations

The case we have yet to consider is when $b - a$ is not finite. For notational simplicity, we will focus here on scalar states $X[\ell] \in \mathbb{R}$ and scalar observations $Y[\ell] \in \mathbb{R}$. The main ideas developed here can be extended to the vector cases without too much difficulty.

In the cases when $a = -\infty$, or $b = \infty$, or both $a = -\infty$ and $b = \infty$, we can no longer use our matrix-vector notation. Instead, a linear estimator for $X[\ell]$ must be written as the sum

$$\hat{X}[\ell] = \sum_{k=a}^b h[\ell, k] Y[k] + c[\ell] \in \mathbb{R}$$

In particular, we are going to have to be careful about the exchange of expectations and summations since the summations are no longer finite.

Exchange of Summation and Expectation

Theorem (Poor V.C.1)

Given $E\{Y^2[k]\} < \infty$ for all k , $h[\ell, k] \in \mathbb{R}$ for all ℓ, k , and

$$\hat{X}[\ell] = \sum_{k=a}^b h[\ell, k]Y[k] + c[\ell]$$

for $b \geq a$, then $E\{(\hat{X}[\ell])^2\} < \infty$. Moreover, if Z is any random variable satisfying $E\{Z^2\} < \infty$ then

$$E\{Z\hat{X}[\ell]\} = \sum_{k=a}^b h[\ell, k]E\{ZY[k]\} + c[\ell]E\{Z\}.$$

This theorem is obviously true when a and b are finite. The second result is a little bit tricky (but important) to show for the cases when $a = -\infty$, or $b = \infty$, or both $a = -\infty$ and $b = \infty$.

Convergence In Mean Square for Infinite Sums

Although not explicit in the theorem, the proof for case when we have an infinite number of observations is going to require that the infinite sums of random variables in the estimates converge in mean square to their limit. Specifically, when $a = -\infty$, and b is finite, the estimate is given as

$$\hat{X}[\ell] = \sum_{k=-\infty}^b h[\ell, k]Y[k] + c[\ell] \in \mathbb{R}.$$

We require the infinite sum to converge in mean square to its limit, i.e.

$$\lim_{m \rightarrow -\infty} \mathbb{E} \left\{ \left(\sum_{k=m}^b h[\ell, k]Y[k] + c[\ell] - \hat{X}[\ell] \right)^2 \right\} = 0.$$

We also require the same sort of mean square convergence for the case when a is finite and $b = \infty$ as well as the case when $a = -\infty$ and $b = \infty$.

Exchange of Summation and Expectation

We focus here on the case when $a = -\infty$ and b is finite since the other cases can be shown similarly.

Proof of the first consequence.

To show the first consequence, we can write for a value of $-\infty < m \leq b$

$$\hat{X}[\ell] = \sum_{k=m}^b h[\ell, k]Y[k] + c[\ell] + \left(\hat{X}[\ell] - \sum_{k=m}^b h[\ell, k]Y[k] - c[\ell] \right)$$

and since $(x + y)^2 \leq 4x^2 + 4y^2$, we can write

$$\mathbb{E} \left\{ (\hat{X}[\ell])^2 \right\} \leq 4\mathbb{E} \left\{ \left(\sum_{k=m}^b h[\ell, k]Y[k] + c[\ell] \right)^2 \right\} + 4\mathbb{E} \left\{ \left(\hat{X}[\ell] - \sum_{k=m}^b h[\ell, k]Y[k] - c[\ell] \right)^2 \right\}$$

The first term is a finite sum with finite coefficients and hence must be finite under our assumption that $\mathbb{E}\{Y^2[\ell]\} < \infty$. Under our mean-squared convergence requirement, the second term goes to zero as $m \rightarrow -\infty$. Hence, there must be a finite value of m that makes this term finite. Hence $\mathbb{E} \left\{ (\hat{X}[\ell])^2 \right\} < \infty$. □

Proof of the second consequence.

To show the second consequence, we can write for a value of $-\infty < m \leq b$

$$\mathbb{E}\{Z\hat{X}[\ell]\} - \sum_{k=m}^b h[\ell, k]\mathbb{E}\{ZY[k]\} + c[\ell]\mathbb{E}\{Z\} = \mathbb{E}\left\{Z\left(\hat{X}[\ell] - \sum_{k=m}^b h[\ell, k]Y[k] - c[\ell]\right)\right\}$$

where the exchange of expectation and summation is valid since the summation is finite. The Schwarz inequality implies that the squared value of the RHS is upper bounded by

$$\left|\mathbb{E}\left\{Z\left(\hat{X}[\ell] - \sum_{k=m}^b h[\ell, k]Y[k] - c[\ell]\right)\right\}\right|^2 \leq \mathbb{E}\{Z^2\} \mathbb{E}\left\{\left(\hat{X}[\ell] - \sum_{k=m}^b h[\ell, k]Y[k] - c[\ell]\right)^2\right\}$$

A condition of the theorem is that $\mathbb{E}\{Z^2\} < \infty$. Under our mean-squared convergence requirement, the second term on the RHS goes to zero as $m \rightarrow -\infty$. Hence the whole RHS goes to zero and we can conclude that $\lim_{m \rightarrow -\infty} \text{LHS} = 0$, or that

$$\mathbb{E}\{Z\hat{X}[\ell]\} = \sum_{k=-\infty}^b h[\ell, k]\mathbb{E}\{ZY[k]\} + c[\ell]\mathbb{E}\{Z\}.$$



Remarks

- ▶ The critical consequence of the theorem is that it allows us to exchange the order of expectation and summation, even when the summations are infinite:

$$\mathbb{E}\{Z\hat{X}[\ell]\} = \sum_{k=-\infty}^b h[\ell, k]\mathbb{E}\{ZY[k]\} + c[\ell]\mathbb{E}\{Z\}$$

$$\mathbb{E}\{Z\hat{X}[\ell]\} = \sum_{k=a}^{\infty} h[\ell, k]\mathbb{E}\{ZY[k]\} + c[\ell]\mathbb{E}\{Z\}$$

$$\mathbb{E}\{Z\hat{X}[\ell]\} = \sum_{k=-\infty}^{\infty} h[\ell, k]\mathbb{E}\{ZY[k]\} + c[\ell]\mathbb{E}\{Z\}$$

- ▶ The regularity conditions for this to hold are pretty mild:
 - ▶ $\mathbb{E}\{Z^2\} < \infty$ and $\mathbb{E}\{Y[k]^2\} < \infty$ for all k .
 - ▶ The estimator $\hat{X}[\ell]$ must converge in a mean square sense to its limit.

The Principle of Orthogonality

Theorem

A linear estimator of the scalar state $X[\ell]$ is an LMMSE estimator if and only if

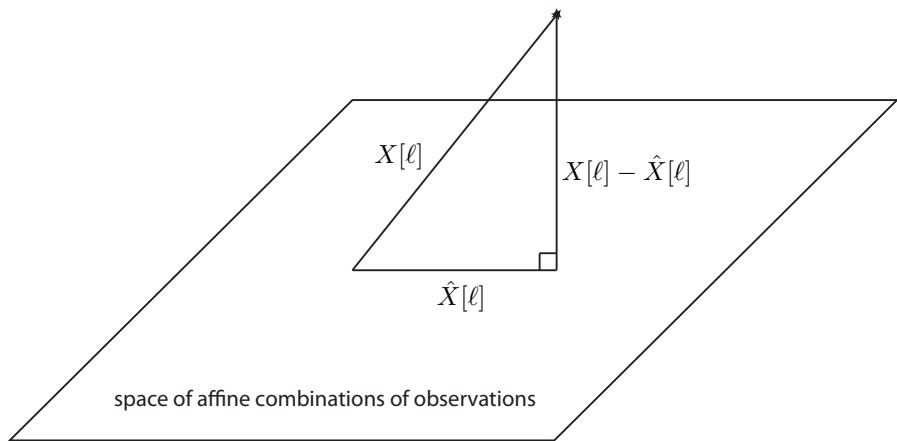
$$\mathbb{E} \left\{ \left(\hat{X}[\ell] - X[\ell] \right) Z \right\} = 0$$

for all Z that are affine functions of the observations $Y[a], \dots, Y[b]$.

Remarks:

- ▶ This is a special case of the “projection theorem” in analysis.
- ▶ The theorem says that $\hat{X}[\ell]$ is an LMMSE estimator (a special affine function of the observations $Y[a], \dots, Y[b]$) if and only if the estimation error $\hat{X}[\ell] - X[\ell]$ is orthogonal (in a statistical sense) to every affine function of those observations.
- ▶ Note that the condition is both necessary and sufficient.

The Principle of Orthogonality: Geometric Intuition



The Principle of Orthogonality

Proof part 1.

We will show that if $\hat{X}[l]$ satisfies the orthogonality condition, it must be an LMMSE estimator. Suppose we have another linear estimator $\tilde{X}[l]$. We can write its MSE as

$$\begin{aligned} \mathbb{E} \left\{ \left(\tilde{X}[l] - X[l] \right)^2 \right\} &= \mathbb{E} \left\{ \left(\tilde{X}[l] - \hat{X}[l] + \hat{X}[l] - X[l] \right)^2 \right\} \\ &= \mathbb{E} \left\{ \left(\tilde{X}[l] - \hat{X}[l] \right)^2 \right\} + 2\mathbb{E} \left\{ \left(\tilde{X}[l] - \hat{X}[l] \right) \left(\hat{X}[l] - X[l] \right) \right\} \\ &\quad + \mathbb{E} \left\{ \left(\hat{X}[l] - X[l] \right)^2 \right\} \end{aligned}$$

Let $Z = \tilde{X}[l] - \hat{X}[l]$ and note that Z is an affine function of the observations. Since $\hat{X}[l]$ satisfies the orthogonality condition, the 2nd term on the RHS is equal to zero and

$$\mathbb{E} \left\{ \left(\tilde{X}[l] - X[l] \right)^2 \right\} = \mathbb{E} \left\{ \left(\tilde{X}[l] - \hat{X}[l] \right)^2 \right\} + \mathbb{E} \left\{ \left(\hat{X}[l] - X[l] \right)^2 \right\} \geq \mathbb{E} \left\{ \left(\hat{X}[l] - X[l] \right)^2 \right\}$$

Since $\tilde{X}[l]$ was arbitrary, this shows that $\hat{X}[l]$ is an LMMSE estimator. \square

Proof part 2.

We will now show that if $\tilde{X}[\ell]$ doesn't satisfy the orthogonality condition, it can't be an LMMSE estimator. Given a linear estimator $\tilde{X}[\ell]$, suppose there is an affine function of the observations Z such that $\mathbb{E}\left\{(\tilde{X}[\ell] - X[\ell])Z\right\} \neq 0$. Let

$$\hat{X}[\ell] = \tilde{X}[\ell] + \frac{\mathbb{E}\{(X[\ell] - \tilde{X}[\ell])Z\}}{\mathbb{E}\{Z^2\}}Z$$

It can be shown via the Shwarz inequality that if $\mathbb{E}\left\{(\tilde{X}[\ell] - X[\ell])Z\right\} \neq 0$ then $\mathbb{E}\{Z^2\} > 0$. It should also be clear that $\hat{X}[\ell]$ is a linear estimator. The MSE of $\hat{X}[\ell]$ is

$$\begin{aligned} \mathbb{E}\left\{\left(X[\ell] - \hat{X}[\ell]\right)^2\right\} &= \mathbb{E}\left\{\left(X[\ell] - \tilde{X}[\ell] - \frac{\mathbb{E}\{(X[\ell] - \tilde{X}[\ell])Z\}}{\mathbb{E}\{Z^2\}}Z\right)^2\right\} \\ &= \mathbb{E}\left\{\left(X[\ell] - \tilde{X}[\ell]\right)^2\right\} - \frac{\mathbb{E}^2\{(X[\ell] - \tilde{X}[\ell])Z\}}{\mathbb{E}\{Z^2\}} \\ &< \mathbb{E}\left\{\left(X[\ell] - \tilde{X}[\ell]\right)^2\right\} \end{aligned}$$

Since the MSE of $\tilde{X}[\ell]$ is strictly greater than the MSE of $\hat{X}[\ell]$, it can't be LMMSE. \square

The Principle of Orthogonality Part II

Theorem

A linear estimator of the scalar state $X[\ell]$ is an LMMSE estimator if and only if

$$\mathbf{E}\{\hat{X}[\ell]\} = \mathbf{E}\{X[\ell]\}$$

and

$$\mathbf{E}\left\{\left(\hat{X}[\ell] - X[\ell]\right) Y[\ell]\right\} = 0$$

for all $a \leq \ell \leq b$.

The Principle of Orthogonality Part II

Proof part 1.

We will show that if $\hat{X}[\ell]$ is an LMMSE estimator, then it must satisfy both conditions of PoO part II. From PoO part I, we know

$$\begin{aligned} & \mathbb{E}\{(\hat{X}[\ell] - X[\ell])Z\} = 0 \text{ for any } Z \text{ that is affine in the observations} \\ \Rightarrow & \mathbb{E}\{(\hat{X}[\ell] - X[\ell])1\} = 0 \text{ since } Z=1 \text{ is affine} \\ \Rightarrow & \mathbb{E}\{\hat{X}[\ell]\} = \mathbb{E}\{X[\ell]\} \end{aligned}$$

$\mathbb{E}\left\{(\hat{X}[\ell] - X[\ell])Y[\ell]\right\} = 0$ for all $a \leq \ell \leq b$ is also a direct consequence of the results from PoO part I since $Y[\ell]$ is an affine function of the observations $Y[a], \dots, Y[b]$. \square

The Principle of Orthogonality Part II

Proof part 2.

Now we will show that if $E\{\hat{X}[\ell]\} = E\{X[\ell]\}$ and $E\left\{\left(\hat{X}[\ell] - X[\ell]\right) Y[\ell]\right\} = 0$ for all $a \leq \ell \leq b$, then $\hat{X}[\ell]$ must be an LMMSE estimator. Let $Z = \sum_{k=a}^b g[\ell, k] Y[k] + d[\ell]$ be an affine function of the observations. Then we have

$$\begin{aligned} E\left\{\left(\hat{X}[\ell] - X[\ell]\right) Z\right\} &= E\left\{\left(\hat{X}[\ell] - X[\ell]\right) \left(\sum_{k=a}^b g[\ell, k] Y[k] + d[\ell]\right)\right\} \\ &= \sum_{k=a}^b g[\ell, k] E\left\{\left(\hat{X}[\ell] - X[\ell]\right) Y[k]\right\} + E\left\{\hat{X}[\ell] - X[\ell]\right\} d[\ell] \\ &= 0 + 0 \end{aligned}$$

Since Z is arbitrary, the results of PoO part I imply that $\hat{X}[\ell]$ must be an LMMSE estimator. □

Remark:

- ▶ Note that this result required an exchange of the sum and the expectation.

The Wiener-Hopf Equations

The PoO part II allows us to obtain a series of equations to specify LMMSE estimator coefficients when we have either a finite or an infinite number of observations. We can immediately find the LMMSE affine offset coefficient $c[\ell]$ by using the fact that $E\{\hat{X}[\ell]\} = E\{X[\ell]\}$:

$$E\left\{\sum_{k=a}^b h[\ell, k]Y[k] + c[\ell]\right\} = E\{X[\ell]\}$$

$$\Rightarrow \boxed{c[\ell] = E\{X[\ell]\} - \sum_{k=a}^b h[\ell, k]E\{Y[k]\}}$$

where this result requires the exchange of the order of expectation and summation to be valid.

The Wiener-Hopf Equations

Using the fact that an LMMSE estimator must satisfy

$E \left\{ \left(\hat{X}[\ell] - X[\ell] \right) Y[\ell] \right\} = 0$ for all $\ell = a, \dots, b$, we can write

$$E \left\{ \left(X[\ell] - \sum_{k=a}^b h[\ell, k] Y[k] - c[\ell] \right) Y[\ell] \right\} = 0 \text{ for all } \ell = a, \dots, b$$

Substitute for $c[\ell]$ and do a bit of simplification...

$$E \left\{ \left[X[\ell] - E\{X[\ell]\} - \sum_{k=a}^b h[\ell, k] (Y[k] - E\{Y[k]\}) \right] Y[\ell] \right\} = 0$$

which can be further simplified to

$$\text{cov}(X[\ell], Y[\ell]) = \sum_{k=a}^b h[\ell, k] \text{cov}(Y[k], Y[\ell])$$

for all $\ell = a, \dots, b$. Note that we exchanged the sum and the expectation order to get the final result.

The Wiener-Hopf Equations: Remarks

In the case of finite a and b , it isn't difficult to show that the Wiener-Hopf equations can be written in the same form that we derived earlier:

$$h_{\text{lmmse}} = [\text{cov}(Y_a^b, Y_a^b)]^{-1} \text{cov}(Y_a^b, X[\ell])$$

The cases when $a = -\infty$, or $b = \infty$, or both $a = -\infty$ and $b = \infty$ are a bit messier, however, since we have an infinite number of Wiener-Hopf equations and can't form finite dimensional matrices and vectors. We can only hope to solve these sort of problems with some further simplifying assumptions...

Simplifying assumption number 1: the (scalar) observation sequence is wide-sense stationary, i.e.,

$$\text{cov}(Y[\ell], Y[k]) = C_{YY}[\ell - k] = C_{YY}[k - \ell]$$

Simplifying assumption number 2: the (scalar) observation sequence and (scalar) state sequence are jointly wide-sense stationary, i.e.,

$$\text{cov}(X[\ell], Y[k]) = C_{XY}[\ell - k] = C_{YX}[k - \ell]$$

Non-Causal Wiener-Kolmogorov Filtering

Under our stationarity assumptions, we first consider the case when both $a = -\infty$ and $b = \infty$, i.e.

$$\hat{X}[\ell] = \sum_{k=-\infty}^{\infty} h[\ell, k]Y[k]$$

Note that we are omitting the affine offset coefficient $c[\ell]$ here since the means can all be assumed to be equal to zero without loss of generality. The Wiener-Hopf equations for this problem can then be written as

$$C_{XY}[\ell - j] = \sum_{k=-\infty}^{\infty} h[\ell, k]C_{YY}[k - j] \text{ for } -\infty < j < \infty$$

$$C_{XY}[\tau] = \sum_{\nu=-\infty}^{\infty} h[\ell, \ell - \nu]C_{YY}[\tau - \nu] \text{ for } -\infty < \tau < \infty$$

where we used the substitutions $\tau = \ell - j$ and $\nu = \ell - k$.

Non-Causal Wiener-Kolmogorov Filtering

Inspection of the Wiener-Hopf equations for the estimator $\hat{X}[\ell]$

$$C_{XY}[\tau] = \sum_{\nu=-\infty}^{\infty} h[\ell, \ell - \nu] C_{YY}[\tau - \nu] \text{ for } -\infty < \tau < \infty$$

reveals that the time index ℓ only appears in the coefficient sequence. This is a direct consequence of our stationarity assumption and implies that the filter $h[\ell, k]$ doesn't depend on ℓ (h is shift-invariant). Since ℓ is irrelevant, we can let $h[\nu] := h[\ell, \ell - \nu]$ and rewrite the Wiener-Hopf equations as

$$C_{XY}[\tau] = \sum_{\nu=-\infty}^{\infty} h[\nu] C_{YY}[\tau - \nu] \text{ for } -\infty < \tau < \infty$$

What is this? It is the discrete-time convolution of two infinite length sequences. We can get more insight by moving to the frequency domain...

Non-Causal Wiener-Kolmogorov Filtering

Assuming that the discrete-time Fourier transforms exist, we can define

$$H(\omega) = \sum_{k=-\infty}^{\infty} h[k]e^{-i\omega k} \text{ (transfer function of } h\text{)}$$

$$\phi_{XY}(\omega) = \sum_{k=-\infty}^{\infty} C_{XY}[k]e^{-i\omega k} \text{ (cross spectral density)}$$

$$\phi_{YY}(\omega) = \sum_{k=-\infty}^{\infty} C_{YY}[k]e^{-i\omega k} \text{ (power spectral density)}$$

and the Wiener-Hopf equations become

$$\phi_{XY}(\omega) = H(\omega)\phi_{YY}(\omega) \text{ for all } -\pi \leq \omega \leq \pi$$

Non-Causal Wiener-Kolmogorov Filtering

As long as $\phi_{YY}(\omega) > 0$ for all $-\pi \leq \omega \leq \pi$, we can solve for the required frequency response of the non-causal Wiener-Kolmogorov filter as

$$H(\omega) = \frac{\phi_{XY}(\omega)}{\phi_{YY}(\omega)}$$

and the shift-invariant filter coefficients can be computed via the inverse discrete Fourier transform as

$$h[\nu] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\phi_{XY}(\omega)}{\phi_{YY}(\omega)} e^{i\omega\nu} d\omega$$

for all integer ν .

Non-Causal Wiener-Kolmogorov Filtering Performance

The MSE of the non-causal Wiener-Kolmogorov filter can be derived as

$$\text{MMSE} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[1 - \frac{|\phi_{XY}(\omega)|^2}{\phi_{XX}(\omega)\phi_{YY}(\omega)} \right] \phi_{XX}(\omega) d\omega$$

(see Poor pp. 236-237 for the details).

Interpretation:

- ▶ What can we say about the term $\gamma = \frac{|\phi_{XY}(\omega)|^2}{\phi_{XX}(\omega)\phi_{YY}(\omega)}$?
- ▶ When the sequences X and Y are completely uncorrelated, $\gamma = 0$ and

$$\text{MMSE} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi_{XX}(\omega) d\omega.$$

What is this quantity? The observations don't help us here and the best we can do is estimate with the unconditional mean, i.e. $\hat{X}[\ell] = 0$. The resulting MMSE is then just the power of the random states.

- ▶ What happens when X and Y are perfectly correlated?

Non-Causal Wiener-Kolmogorov Filtering: Main Results

Given observations $\{Y[k]\}_{-\infty}^{\infty}$ (stationary and cross-stationary with the states), and assuming all the discrete-time Fourier transforms exist, the non-causal Wiener-Kolmogorov LMMSE estimator for the state $X[\ell]$ can be expressed as

$$H(\omega) = \frac{\phi_{XY}(\omega)}{\phi_{YY}(\omega)}$$

for all $-\pi \leq \omega \leq \pi$, or, equivalently,

$$h[\nu] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\phi_{XY}(\omega)}{\phi_{YY}(\omega)} e^{i\omega\nu} d\omega$$

for all integer ν . The MSE of the non-causal Wiener-Kolmogorov filter is

$$\text{MMSE} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[1 - \frac{|\phi_{XY}(\omega)|^2}{\phi_{XX}(\omega)\phi_{YY}(\omega)} \right] \phi_{XX}(\omega) d\omega$$

Causal Wiener-Kolmogorov Filtering: Problem Setup

- ▶ In many applications, e.g. real-time, we need a causal estimator.
- ▶ We now consider the problem of LMMSE estimation of $X[\ell]$ given observations $\{Y[k]\}_{-\infty}^{\ell}$, i.e.

$$\hat{X}[\ell] = \sum_{k=-\infty}^{\ell} h[\ell, k]Y[k]$$

- ▶ Notation: Let the space of affine functions of $\{Y[k]\}_{-\infty}^{\infty}$ be denoted as \mathcal{H}^{nc} and let the space of affine functions of $\{Y[k]\}_{-\infty}^{\ell}$ be denoted as \mathcal{H}^c .
- ▶ Notation: Let $\hat{X}^{nc}[\ell]$ be a non-causal Wiener-Kolmogorov filter and let $\hat{X}^c[\ell]$ be a causal Wiener-Kolmogorov filter. Note that

$$\begin{aligned}\hat{X}^{nc}[\ell] &\in \mathcal{H}^{nc} \\ \hat{X}^c[\ell] &\in \mathcal{H}^c \subset \mathcal{H}^{nc}\end{aligned}$$

- ▶ Idea: Can we just **project** our non-causal W-K filter into \mathcal{H}^c ?

Causal Wiener-Kolmogorov Filtering

Let $Z \in \mathcal{H}^c$. The principle of orthogonality implies that, if $\hat{X}^c[\ell]$ is a causal W-K LMMSE filter, then

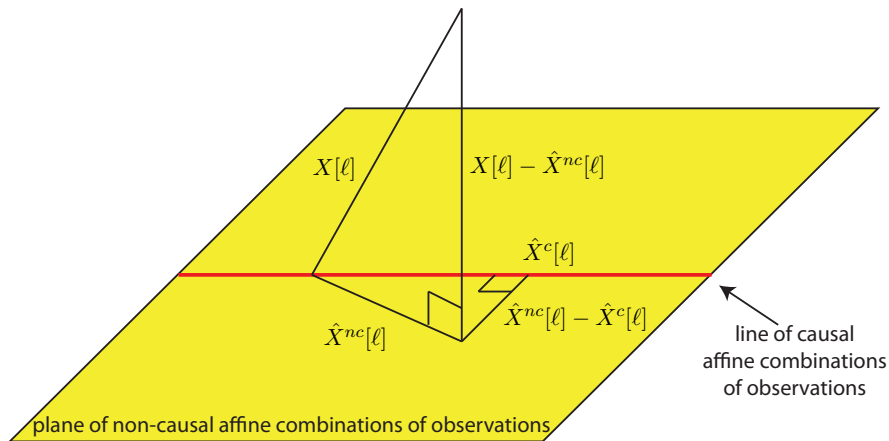
$$\begin{aligned} \mathbb{E} \left\{ (X[\ell] - \hat{X}^c[\ell])Z \right\} &= 0 \\ \mathbb{E} \left\{ (X[\ell] - \hat{X}^{nc}[\ell] + \hat{X}^{nc}[\ell] - \hat{X}^c[\ell])Z \right\} &= 0 \\ \mathbb{E} \left\{ (X[\ell] - \hat{X}^{nc}[\ell])Z \right\} + \mathbb{E} \left\{ (\hat{X}^{nc}[\ell] - \hat{X}^c[\ell])Z \right\} &= 0 \end{aligned}$$

What can we say about the first term here?

Hence $\mathbb{E} \left\{ (\hat{X}^{nc}[\ell] - \hat{X}^c[\ell])Z \right\} = 0$. What does the principle of orthogonality imply about this result? **Among all of the causal estimators in \mathcal{H}^c , $\hat{X}^c[\ell]$ is the LMMSE estimator of the non-causal $\hat{X}^{nc}[\ell]$.**

Implication: The causal estimator $\hat{X}^c[\ell]$ can be obtained by first projecting $X[\ell]$ onto \mathcal{H}^{nc} to obtain $\hat{X}^{nc}[\ell]$ and then projecting $\hat{X}^{nc}[\ell]$ onto \mathcal{H}^c .

Causal Wiener-Kolmogorov Filtering: Geometric Intuition



Causal Wiener-Kolmogorov Filtering

Suppose we have our non-causal LMMSE Wiener-Kolmogorov filter coefficients already computed as

$$\hat{X}^{nc}[\ell] = \sum_{k=-\infty}^{\infty} h^{nc}[\ell, k]Y[k].$$

How should we perform the projection to obtain a causal LMMSE Wiener-Kolmogorov filter? Let's try this:

$$\hat{X}^c[\ell] = \sum_{k=-\infty}^{\ell} h^{nc}[\ell, k]Y[k].$$

We are just truncating the non-causal filter to get a causal filter. How can we check to see if this is indeed the causal LMMSE Wiener-Kolmogorov filter? Principle of orthogonality...

Causal Wiener-Kolmogorov Filtering

By the principle of orthogonality (part II), we know that $\hat{X}^c[\ell]$ is LMMSE if

$$\mathbf{E} \left\{ (\hat{X}^{nc}[\ell] - \hat{X}^c[\ell])Y[m] \right\} = 0 \text{ for all } m = \dots, \ell - 1, \ell \quad (2)$$

Our causal filter is just a truncation of the non-causal W-K filter, hence

$$\hat{X}^{nc}[\ell] - \hat{X}^c[\ell] = \sum_{k=\ell+1}^{\infty} h^{nc}[\ell, k]Y[k]$$

One condition under which (2) holds is when the $\{Y[k]\}_{-\infty}^{\infty}$ is an uncorrelated sequence. In this case, for any $m \leq \ell$, we can write

$$\begin{aligned} \mathbf{E} \left\{ (\hat{X}^{nc}[\ell] - \hat{X}^c[\ell])Y[m] \right\} &= \mathbf{E} \left\{ \left(\sum_{k=\ell+1}^{\infty} h^{nc}[\ell, k]Y[k] \right) Y[m] \right\} \\ &= \sum_{k=\ell+1}^{\infty} h^{nc}[\ell, k] \mathbf{E} \{ Y[k]Y[m] \} = 0 \end{aligned}$$

Causal Wiener-Kolmogorov Filtering

Requiring $\{Y[k]\}_{-\infty}^{\infty}$ to be uncorrelated is a pretty big restriction on the types of problems we can solve. What can we do if $\{Y[k]\}_{-\infty}^{\infty}$ is correlated?

If we can convert $\{Y[k]\}_{-\infty}^{\infty}$ into an equivalent uncorrelated stationary sequence $\{Z[k]\}_{-\infty}^{\infty}$ by a **causal linear operation**, then we can determine the causal W-K LMMSE filter by

- ▶ First computing the non-causal W-K LMMSE filter coefficients based on the uncorrelated sequence $\{Z[k]\}_{-\infty}^{\infty}$ and
- ▶ then truncating the coefficients so that the resulting filter is causal.

Hence we need to see if we can causally and linearly whiten $\{Y[k]\}_{-\infty}^{\infty}$.

Theorem (Spectral Factorization Theorem)

Suppose $\{Y[k]\}_{-\infty}^{\infty}$ is wide-sense stationary and has a power spectrum $\phi_{YY}(\omega)$ satisfying the "Paley-Wiener condition", given by

$$\int_{-\pi}^{\pi} \ln \phi_{YY}(\omega) d\omega > -\infty.$$

Then $\phi_{YY}(\omega)$ can be factored as $\phi_{YY}(\omega) = \phi_{YY}^+(\omega)\phi_{YY}^-(\omega)$ for $-\pi \leq \omega \leq \pi$ where $\phi_{YY}^+(\omega)$ and $\phi_{YY}^-(\omega)$ are two functions satisfying $|\phi_{YY}^+(\omega)|^2 = |\phi_{YY}^-(\omega)|^2 = \phi_{YY}(\omega)$,

$$\int_{-\pi}^{\pi} \phi_{YY}^+(\omega) e^{in\omega} d\omega = 0 \text{ for all } n < 0, \quad (3)$$

and

$$\int_{-\pi}^{\pi} \phi_{YY}^-(\omega) e^{in\omega} d\omega = 0 \text{ for all } n > 0. \quad (4)$$

Moreover (3) also holds when $\phi_{YY}^+(\omega)$ is replaced with $1/\phi_{YY}^+(\omega)$ and (4) also holds when $\phi_{YY}^-(\omega)$ is replaced with $1/\phi_{YY}^-(\omega)$.

Remarks on the Spectral Factorization Theorem

- ▶ The Paley-Wiener condition implies that $\phi_{YY}(\omega) > 0$ for all $-\pi \leq \omega \leq \pi$, hence $\phi_{YY}^+(\omega) > 0$ and $\phi_{YY}^-(\omega) > 0$ for all $-\pi \leq \omega \leq \pi$ and $1/\phi_{YY}^+(\omega)$ and $1/\phi_{YY}^-(\omega)$ are finite for all $-\pi \leq \omega \leq \pi$.
- ▶ The inverse discrete-time Fourier transforms show that $\phi_{YY}^+(\omega)$ has a **causal** discrete-time impulse response and $\phi_{YY}^-(\omega)$ has an **anti-causal** discrete-time impulse response.
- ▶ The final part of the theorem implies also that $1/\phi_{YY}^+(\omega)$ has a **causal** discrete-time impulse response and $1/\phi_{YY}^-(\omega)$ has an **anti-causal** discrete-time impulse response.
- ▶ Since $\phi_{YY}(\omega)$ is a power spectral density, $\phi_{YY}(\omega) = \phi_{YY}(-\omega)$ and it follows that the spectral factors also have the property that

$$\phi_{YY}^-(\omega) = [\phi_{YY}^+(\omega)]^* .$$

One Way to Whiten $\{Y[k]\}_{-\infty}^{\infty}$

- ▶ Recall from our discussion of the discrete-time Kalman-Bucy filter that the innovation sequence

$$\tilde{Y}[\ell | \ell - 1] := Y[\ell] - H[\ell]\hat{X}[\ell | \ell - 1] = Y[\ell] - \hat{Y}[\ell | \ell - 1]$$

was an uncorrelated sequence, even if the sequence $\{Y[\ell]\}_{-\infty}^{\infty}$ is correlated.

- ▶ Note that the $\tilde{Y}[\ell | \ell - 1]$ is not wide-sense stationary (variance is time-varying).
- ▶ In any case, this suggests that one practical way to build a whitening filter is to do **one-step prediction** of the observations.
- ▶ The only minor detail is that we need to compute a scaled innovation so that resulting sequence is stationary.

One Way to Whiten $\{Y[k]\}_{-\infty}^{\infty}$

- Denote the one-step causal LMMSE predictor of $Y[\ell]$ given the previous observations $\{Y[k]\}_{-\infty}^{\ell-1}$ as

$$\hat{Y}[\ell | \ell - 1] = \sum_{k=-\infty}^{\ell-1} g[\ell - k]Y[k]$$

and denote the MSE of this predictor as

$$\sigma^2[\ell] = \mathbb{E} \left\{ \left(Y[\ell] - \hat{Y}[\ell | \ell - 1] \right)^2 \right\} \quad -\infty < \ell < \infty$$

- If we define a new sequence

$$Z[\ell] = \frac{Y[\ell] - \hat{Y}[\ell | \ell - 1]}{\sigma[\ell]} \quad -\infty < \ell < \infty$$

then it isn't difficult to show that the elements of this sequence have zero mean, unit variance, and are temporally uncorrelated (using the principle of orthogonality, since the predictor is LMMSE).

- You can use the spectral factorization theorem to derive the causal filter coefficients $g[\nu]$ required to implement the one-step predictor (the details are in your textbook pp. 242-246).

Whitening

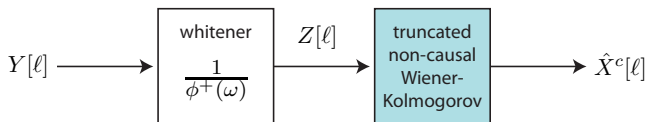
We use the fact that the spectrum at the output of an LTI filter with frequency response $G(\omega)$ is equal to the product of the spectrum at the input of the filter and $|G(\omega)|^2$.

At the input of the filter, we have the correlated observations $Y[\ell]$ with spectrum $\phi_{YY}(\omega)$. At the output of the filter, we want the equivalent sequence $Z[\ell]$ with constant spectrum. What should we use as our filter?

Since our whitening filter must be causal, we should use a filter with frequency response $\frac{1}{\phi_{YY}^+(\omega)}$. The spectral factorization theorem ensures that impulse response of the filter with this frequency response is causal if the Paley-Wiener condition is satisfied.

What is the spectrum of $Z[\ell]$ in this case?

Causal Wiener-Kolmogorov Filtering



Note that the non-causal Wiener-Kolmogorov filter here is based on the whitened observations $Z[\ell]$. That is,

$$H^{nc}(\omega) = \frac{\phi_{XZ}(\omega)}{\phi_{ZZ}(\omega)} \text{ for all } -\pi \leq \omega \leq \pi$$

We know that $\phi_{ZZ}(\omega) \equiv 1$ and it can also be shown without too much difficulty that $\phi_{XZ}(\omega) = G^*(\omega)\phi_{XY}(\omega)$ where $G(\omega) = 1/\phi^+(\omega)$. Hence, the non-causal Wiener-Kolmogorov filter can be expressed as

$$H^{nc}(\omega) = \frac{\phi_{XY}(\omega)}{[\phi^+(\omega)]^*} = \frac{\phi_{XY}(\omega)}{\phi^-(\omega)} \text{ for all } -\pi \leq \omega \leq \pi$$

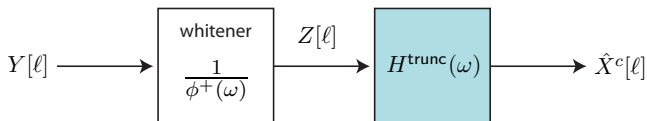
with impulse response $h^{nc}[\nu]$ for $-\infty < \nu < \infty$.

Causal Wiener-Kolmogorov Filtering

Denoting the frequency response of the causally-truncated Wiener-Kolmogorov filter (based on white observations $Z[\ell]$) as

$$H^{\text{trunc}}(\omega) := \sum_{\nu=0}^{\infty} h^{nc}[\nu] e^{-i\omega\nu}$$

we now have



Note that both filters are causal here, hence their series connection is also causal. The frequency response of the causal Wiener-Kolmogorov filter (based on temporally correlated observations $Y[\ell]$) can then be written as

$$H^c(\omega) = \frac{H^{\text{trunc}}(\omega)}{\phi^+(\omega)} \text{ for all } -\pi \leq \omega \leq \pi.$$

Final Remarks

- ▶ Linear estimation is motivated by
 - ▶ Computational convenience.
 - ▶ Analytical tractability.
 - ▶ No loss of optimality under MSE criterion and observations jointly Gaussian with unknown state/parameter.
- ▶ We looked at (not in this order)
 - ▶ Linear MMSE with a finite number of observations (matrix-vector formulation)
 - ▶ Kalman Filter (both static and dynamic states)
 - ▶ Linear MMSE with an infinite number of observations
 - ★ Principle of orthogonality
 - ★ Wiener-Hopf equations
 - ★ Stationarity assumptions
 - ★ Non-causal Wiener-Kolmogorov filtering
 - ★ Causal Wiener-Kolmogorov filtering
 - ★ Spectral factorization
- ▶ Whitening (aka pre-whitening) is a useful tool in many applications.

Comprehensive Final Exam: What You Need to Know

- ▶ Everything from the first half of the course (see end of Lec. 6).
- ▶ Different types of parameter estimation problems (Bayesian and non-random, static and dynamic)
- ▶ Mathematical model of parameter estimation problems.
- ▶ Bayesian parameter estimation (MMSE, MMAE, MAP, ...)
- ▶ Non-random parameter estimation
 - ▶ Unbiased estimators and MVU estimators.
 - ▶ RBLs theorem.
 - ▶ Information inequality and the Cramer-Rao lower bound.
 - ▶ Maximum likelihood estimators.
- ▶ Dynamic parameter/state estimation and Kalman filtering.
- ▶ Linear MMSE estimation with a finite number of observations.

Linear MMSE estimation with an infinite number of observations will not be on the final exam.