Homework #8 Solutions

Please submit your assignment as a PDF to the class Gradescope page.

1. Prediction. Consider the wide sense stationary process $Y(n)$ defined by

$$Y(n) = aY(n-1) + X(n)$$

where $X(n)$ is a zero-mean white noise process with variance $\sigma^2$, and $|a| < 1$. We are interested in predicting the future of the process $l$ time steps into the future, i.e. to predict $Y(n+l)$ based on $Y(m)$, $-\infty < m \leq n$.

a. Prove that the optimal linear predictor is

$$\hat{Y}(n+l|n) = a^l Y(n),$$

and explain why this predictor does not need the measurements that proceed $Y(n)$. Hint: use the orthogonality principle.

b. Express, in terms of $a, \sigma^2$ and $l$, the prediction error

$$\epsilon_l^2 = E[(Y(n+l) - \hat{Y}(n+l|n))^2].$$

Check and explain your results for $l = 1$ and $l \to \infty$.

c. Define the $l^{th}$ order innovation process by

$$U_l(n) = Y(n+l) - \hat{Y}(n+l|n).$$

Prove that

$$E[U_l(n)U_l(n-k)] = 0, \quad k \geq l,$$

i.e. the autocorrelation sequence of $U_l(n)$ satisfies $R_{U_l}(k) = 0$, for all $k \geq l$.

Solution (20 points)

a. The linear optimal predictor of $\hat{Y}(n+l)$ based on all $Y(m)$, $-\infty < m \leq n$ is of course a weighted linear combination of the $Y(m)$, i.e.

$$\hat{Y}(n+l|n) = \sum_{k\geq0} \alpha_k Y(n-k), \quad m \geq 0.$$  

Using the orthogonality principle, we can write for all $m \geq 0$

$$E \left[ \left( Y(n+l) - \sum_{k\geq0} \alpha_k Y(n-k) \right) Y(n-m) \right] = 0,$$
which results in the equality
\[ R_Y(m + l) = \sum_{k \geq 0} \alpha_k R_Y(k - m), \quad m \geq 0. \]

To find the optimal \( \alpha_k \), we first find the autocorrelation function. Note that because \( Y(n) \) is recursive, we find the autocorrelation function as follows:

\[ R_Y(m) = E[Y(n + m)Y(n)] = E[(aY(n + m - 1) + X(n))Y(n)] = E[(a(aY(n + m - 2) + X(n - 1)) + X(n))Y(n)] = E[(a^2Y(n + m - 2) + aX(n - 1) + X(n))Y(n)] = \cdots \]

\[ = E \left( a^m Y(n) + \sum_{k=0}^{m-1} a^k X(n - k) \right) Y(n) \]

\[ = a^m R_Y(0). \]

With this expression, it is easy to see that the equality above is satisfied for all \( m \geq 0 \) if \( \alpha_0 = a^l \), and \( \alpha_k = 0 \) for all \( k > 0 \). Thus, \( \hat{Y}(n+l|n) = \sum_{k \geq 0} \alpha_k Y(n-k) = a^l Y(n) \).

b. The prediction error is
\[ \epsilon^2 = E[(Y(n + l) - \hat{Y}(n + l|n))^2] = E[(Y(n + l))^2] - 2a^l E[Y(n + l)Y(n)] + a^{2l} E[Y(n)^2] = (1 + a^{2l}) R_Y(0) - 2a^l R_Y(l) \]

To find \( R_Y(0) \) we write
\[ R_Y(0) = E[Y(n)Y(n)] = E[(aY(n - 1) + X(n))(aY(n - 1) + X(n))] = a^2 R_Y(0) + \sigma^2 \]

\[ \Rightarrow R_Y(0) = \frac{\sigma^2}{1 - a^2}. \]

Plugging this back into the prediction error, we find that
\[ \epsilon^2 = \sigma^2 \frac{1 - a^{2l}}{1 - a^2}. \]

Note that the orthogonality principle could have been invoked after the first equality, resulting in the same prediction error.

For \( l = 1 \), the prediction error variance is equal to the prediction error variance.
of a first-order autoregressive process based on its past, and for \( l = \infty \), it is equal to the variance of the process.

c. We note that \( U_l(n) \) is the error process due to predicting \( Y(n_l) \), and thus it is orthogonal to all the observations used in its prediction. Since \( U_l(n-k) \) for \( k \geq l \) is a linear combination of these observations, \( U_l(n) \) and \( U_l(n-k) \) are orthogonal for \( k \geq l \).

2. **ARMA process prediction.** Let \( \{ Y(n) \} \) be an ARMA(1,1) process described by

\[
Y(n) = -\alpha Y(n-1) + X(n) + \beta X(n-1),
\]

where \( \{ X(n) \} \) is a zero-mean, unit-variance white noise process, and \( |\alpha| < 1 \). Find the optimal linear predictor for \( Y(n) \) in terms of \( \alpha \) and \( \beta \) for

a. \( |\beta| < 1 \)

b. \( |\beta| > 1 \)

**Solution** (20 points)

We first obtain the power spectral density of \( Y(n) \). Taking Fourier transforms on both sides of the system equation, we obtain

\[
Y(f) = -\alpha e^{-j2\pi f} Y(f) + X(f) + \beta e^{-j2\pi f} X(f),
\]

which is equivalent to

\[
Y(f) = \frac{1 + \beta e^{-j2\pi f}}{1 + \alpha e^{-j2\pi f}} X(f).
\]

In other words, \( Y(n) \) can be viewed as the output of an LTI system with input \( X(n) \) and transfer function

\[
H(f) = \frac{1 + \beta e^{-j2\pi f}}{1 + \alpha e^{-j2\pi f}}.
\]

Thus, the power spectral density of \( Y(n) \) is

\[
S_Y(f) = |H(f)|^2 S_X(f) = \frac{(1 + \beta e^{-j2\pi f})(1 + \beta e^{j2\pi f})}{(1 + \alpha e^{-j2\pi f})(1 + \alpha e^{j2\pi f})}.
\]

Recall from the lectures that the transfer function \( P(f) \) for the optimal linear predictor of \( Y(n) \) based on all \( Y(m) \) for \( m < n \) is

\[
P(f) = \frac{1}{S_Y^+(f)} [e^{j2\pi f} S_Y^+(f)]^+.
\]

With these expressions, we compute the spectral factorization of \( S_Y(f) \) and \( P(f) \) for both \( |\beta| < 1 \) and \( |\beta| > 1 \).

a. When \( |\beta| < 1 \), the causal factor of \( S_Y(f) \) is

\[
S_Y^+(f) = \frac{1 + \beta e^{-j2\pi f}}{1 + \alpha e^{-j2\pi f}}.
\]
Plugging it into the general formula for the optimal linear predictor, we have

\[
P(f) = \frac{1 + \alpha e^{-j2\pi f}}{1 + \beta e^{-j2\pi f}} \left[ e^{j2\pi f} \frac{1 + \beta e^{-j2\pi f}}{1 + \alpha e^{-j2\pi f}} \right]_+ + \frac{1 + \alpha e^{-j2\pi f}}{1 + \beta e^{-j2\pi f}} \left[ e^{j2\pi f} + \beta \right]_+
\]

Let’s now deal with the right hand term of \( P(f) \). The goal is to find the causal part of the Fourier Transform of \( e^{j2\pi f} + \beta \). Recall from lecture that the causal Fourier Transform of a system is equivalent to the Fourier Transform of the causal component of the system, i.e. the causal part of the time domain representation of the system. So in the following steps, we write out the system as a recognizable Fourier transform, identify the time domain representation of the signal, and then take the Fourier transform of the causal part of that time domain representation:

\[
\left[ \frac{e^{j2\pi f} + \beta}{1 + \alpha e^{-j2\pi f}} \right]_+ = \left[ \frac{e^{j2\pi f}}{1 + \alpha e^{-j2\pi f}} + \frac{\beta}{1 + \alpha e^{-j2\pi f}} \right]_+ \\
= \left[ \frac{e^{j2\pi f}}{1 + \alpha e^{-j2\pi f}} \sum_{n_1=0}^{\infty} (-\alpha e^{-j2\pi f})^{n_1} + \beta \sum_{n_2=0}^{\infty} (-\alpha e^{-j2\pi f})^{n_2} \right]_+ \\
= \left[ \sum_{n_1=0}^{\infty} (-\alpha)^{n_1} e^{-j2\pi f(n_1-1)} + \beta \sum_{n_2=0}^{\infty} (-\alpha)^{n_2} e^{-j2\pi f n_2} \right]_+ \\
= \left[ \sum_{n'_1=-\infty}^{\infty} (-\alpha)^{n'_1+1} e^{-j2\pi f n'_1} + \beta \sum_{n_2=0}^{\infty} (-\alpha)^{n_2} e^{-j2\pi f n_2} \right]_+ \\
= \left[ \sum_{n'_1=-\infty}^{\infty} (-\alpha)^{n'_1+1} e^{-j2\pi f n'_1} u(n'_1 + 1) + \beta \sum_{n_2=0}^{\infty} (-\alpha)^{n_2} e^{-j2\pi f n_2} u(n_2) \right]_+
\]

where \( u(n) \) is the step function, i.e. \( u(n) = 1, n \geq 0 \) and \( u(n) = 0, n < 0 \). Note that the expression above in \([\cdot]_+\) is now a recognizable Fourier transform of the following time domain signal

\[ (-\alpha)^{n+1} u(n+1) + \beta (-\alpha)^n u(n) \]

The causal part of this signal is only the part that is defined for \( n \geq 0 \), i.e.

\[ [(-\alpha)^{n+1} u(n+1) + \beta (-\alpha)^n u(n)]_+ = (\beta - \alpha) (-\alpha)^n u(n). \]

Now we take the Fourier transform of this causal signal to finally obtain the relationship

\[
\left[ \frac{e^{j2\pi f} + \beta}{1 + \alpha e^{-j2\pi f}} \right]_+ = \frac{\beta - \alpha}{1 + \alpha e^{-j2\pi f}}
\]
which we can plug back into $P(f)$ to obtain

$$P(f) = \frac{\beta - \alpha}{1 + \beta e^{-j2\pi f}}$$

$$= (\beta - \alpha) \sum_{n=0}^{\infty} (-\beta e^{-2\pi f})^n$$

$$= \sum_{n=0}^{\infty} (\beta - \alpha)(-\beta)^n e^{-2\pi fn}.$$  

Again we can easily identify the inverse Fourier transform to be

$$p(n) = (\beta - \alpha)(-\beta)^n u(n).$$

Finally, we convolve the predictor impulse response with the input values $Y(m)$ for $m < n$ to obtain the explicit form of the predictor as

$$\hat{Y}(n|n-1) = \sum_{k=0}^{\infty} (\beta - \alpha)(-\beta)^k Y(n-1-k).$$

b. When $|\beta| > 1$, the causal factor of $S_Y(f)$ is

$$S_Y^+(f) = \frac{1 + \beta e^{j2\pi f}}{1 + \alpha e^{-j2\pi f}}$$

$$= \frac{e^{-j2\pi f} + \beta}{1 + \alpha e^{-j2\pi f}}$$

$$= \beta \frac{1 + \beta e^{-j2\pi f}}{1 + \alpha e^{-j2\pi f}}.$$

Repeating the analysis from part (a), we obtain the optimal linear predictor

$$p(n) = (\beta^{-1} - \alpha)(-\beta^{-1})^n u(n)$$

and the explicit form of the predictor

$$\hat{Y}(n|n-1) = \sum_{k=0}^{\infty} (\beta^{-1} - \alpha)(-\beta^{-1})^k Y(n-1-k).$$

3. Comparing filters. Let

$$Y(n) = \frac{1}{2} Y(n-1) + U(n),$$

where $\{U(n)\}$ is a zero-mean white noise process with variance $\sigma_u^2 = 1$, and let

$$Z(n) = Y(n) + W(n),$$

where $\{W(n)\}$ is a zero-mean WSS random process with power spectral density

$$S_W(f) = \frac{1}{1 + \alpha \cos(2\pi f)}, \quad -1/2 \leq f \leq 1/2,$$
and is independent of \( \{Y(n)\} \).

a. For \( \alpha = -0.8 \), find the optimal linear estimator of \( Y(0) \) given \( Z(k), -\infty < k < \infty \).

b. For \( \alpha = -0.8 \), find the optimal linear estimator of \( Y(0) \) given \( Z(k), -\infty < k \leq 0 \).

c. How do the mean squared error of the estimates in parts (a) and (b) compare?

d. Would your answer to part (c) change if \( \alpha \neq -0.8 \)? Explain.

**Solution** (20 points)

a. The infinite smoothing filter is

\[
H_{opt}(f) = \frac{S_{YZ}(f)}{S_Z(f)}.
\]

where \( S_{YZ}(f) \) and \( S_Z(f) \) are given by

\[
R_{YZ}(k) = E[Y(n+k)Z(n)] = E[Y(n+k)(Y(n)+W(n))] = R_Y(k)
\]

since \( \{Y(n)\} \) and \( \{W(n)\} \) are independent. Hence

\[
S_{YZ}(f) = S_Y(f).
\]

To find \( S_Y(f) \), we first find the autocorrelation function \( R_Y(f) \). In problem 1 we found that \( R_Y(m) = a^m R_Y(0) \) where \( a \) is the previous step scale factor and \( R_Y(0) = \sigma^2/(1-a^2) \), where \( \sigma^2 \) is the variance of the white noise function. Plugging in \( \sigma = 1 \) and \( a = 1/2 \) as specified for this problem, we find that \( R_Y(m) = \frac{4}{3} \left(\frac{1}{2}\right)^{|m|} \).
Now we find the power spectral density of the autocorrelation:

\[ S_Y(f) = \mathcal{F}\{R_Y(m)\} = \sum_{m=-\infty}^{\infty} \frac{4}{3} \left( \frac{1}{2} \right)^{|m|} e^{-i2\pi mf} \]

\[ = \frac{4}{3} \left[ \sum_{m_1=-\infty}^{-1} \left( \frac{1}{2} \right)^{-m_1} e^{-i2\pi m_1 f} + \sum_{m_2=0}^{\infty} \left( \frac{1}{2} \right)^{m_2} e^{-i2\pi m_2 f} \right] \]

\[ = \frac{4}{3} \left[ \sum_{m_1=0}^{\infty} \left( \frac{1}{2} \right)^{m_1} e^{i2\pi m_1 f} - 1 + \sum_{m_2=0}^{\infty} \left( \frac{1}{2} \right)^{m_2} e^{-i2\pi m_2 f} \right] \]

\[ = \frac{4}{3} \left[ \frac{1}{1 - \frac{1}{2} e^{i2\pi f}} - 1 + \frac{1}{1 - \frac{1}{2} e^{-i2\pi f}} \right] \]

\[ = \frac{1}{\left(1 - \frac{1}{2} e^{-i2\pi f}\right) \left(1 - \frac{1}{2} e^{i2\pi f}\right)}, \]

where the second to last equality is via the convergence of infinite geometric sums.

So

\[ S_Z(f) = S_Y(f) + S_W(f) \]

\[ = \frac{1}{\left(1 - \frac{1}{2} e^{-i2\pi f}\right) \left(1 - \frac{1}{2} e^{i2\pi f}\right)} + \frac{1}{1 + \frac{1}{2} \alpha (e^{-i2\pi f} + e^{i2\pi f})} \]

\[ = \frac{9}{4 \left(1 - \frac{1}{2} e^{-j2\pi f}\right) \left(1 - \frac{1}{2} e^{j2\pi f}\right)}, \]

where we have used the fact that \( \alpha = -0.8 \).

So, we have

\[ H_{opt}(f) = \frac{S_{YZ}(f)}{S_Z(f)} = \frac{4}{9} \]

and

\[ \hat{Y}(0) = \frac{4}{9} Z(0). \]

b. The answer to part (a) is already causal, so the answer remains the same.

c. Since the estimator remains the same, the error also remains the same.

d. Yes, the answer to (c) would change if \( \alpha \neq -0.8 \). In general, there is no pole-zero cancellation in the non-causal estimator, resulting in a non-causal estimator which uses \( \{Z(k), k \geq 1\} \). This estimator would have a reduced MSE compared with the causal one, since it has access to more information.
4. Estimation with lookahead vs. non-causal Wiener filter. In class we showed that the frequency response of the optimal linear estimator with lookahead of $X(n)$ given $Y(k), -\infty < k \leq n + d$ is given by

$$H_d(f) = \frac{e^{j2\pi fd}}{S_Y^+(f)} \left[ \frac{S_{XY}e^{-j2\pi fd}}{S_Y(f)} \right]_+.$$ 

a. What does the filter $H_d(f)$ converge to as $d \to \infty$?

b. If the frequency response of the non-causal Wiener filter for estimating $X(n)$ from $Y(k), -\infty < k < \infty$ is

$$e^{j2\pi f} \frac{1}{1 - \frac{1}{2}e^{-j2\pi f}},$$

find $H_d(f)$ for $d = 1$.

Solution (20 points)

a. To understand what happens as $d \to \infty$, we define the following quantities:

$$a(n) \triangleq \mathcal{F}^{-1}\left\{ \frac{S_{XY}(f)}{S_Y(f)} \right\} \Rightarrow \frac{S_{XY}(f)}{S_Y(f)} = \sum_{n=-\infty}^{\infty} a(n)e^{-j2\pi fn}$$

$$B(f) \triangleq \frac{S_{XY}(f)e^{-j2\pi fd}}{S_Y^-(f)} = e^{-j2\pi fd} \sum_{n=-\infty}^{\infty} a(n)e^{-j2\pi fn}$$

$$= \sum_{n=-\infty}^{\infty} a(n)e^{-j2\pi f(n+d)}.$$

Now, the causal part of $\left[ \frac{S_{XY}e^{-j2\pi fd}}{S_Y(f)} \right]_+$ is $[B(f)]_+$ for all summation values $n \geq -d$,

$$[B(f)]_+ = \sum_{n=-d}^{\infty} a(n)e^{-j2\pi f(n+d)}.$$
With this new expression we can rewrite the optimal linear estimator

\[
H_d(f) = \frac{e^{j2\pi fd}}{S_Y^+(f)} \sum_{n=-d}^{\infty} a(n)e^{-j2\pi f(n+d)}
\]

\[
= \sum_{n=-\infty}^{\infty} a(n)e^{-j2\pi fn} \frac{1}{S_Y^+(f)}
\]

\[
\Rightarrow \lim_{d \to \infty} H_d(f) = \frac{e^{j2\pi fd}}{S_Y^+(f)} \sum_{n=-\infty}^{\infty} a(n)e^{-j2\pi f(n+d)}
\]

\[
= S_{XY}(f) \frac{1}{S_Y^+(f)} S_Y^+(f)
\]

\[
= \frac{S_{XY}(f)}{S_Y(f)}
\]

which is the non-causal Wiener filter.

b. The non-causal Wiener filter gives us

\[
\frac{S_{XY}(f)}{S_Y(f)} = \frac{e^{j2\pi f}}{1 - \frac{1}{2}e^{-j2\pi f}}
\]

which we can use to find \( H_1(f) \):

\[
H_1(f) = \frac{e^{j2\pi f}}{S_Y^+(f)} \left[ \frac{e^{j2\pi f}}{1 - \frac{1}{2}e^{-j2\pi f}} S_Y^+(f)e^{-j2\pi f} \right]_+
\]

\[
= \frac{e^{j2\pi f}}{S_Y^+(f)} \frac{S_Y^+(f)}{1 - \frac{1}{2}e^{-j2\pi f}}
\]

\[
= \frac{S_Y^+(f)}{1 - \frac{1}{2}e^{-j2\pi f}}
\]

which is the same as the non-causal filter.

Alternatively, it is easy to see that the non-causal Wiener filter does not depend on \( Y(k) \) for \( k \geq n + 2 \), so the non-causal Wiener filter is indeed \( H_1(f) \).