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

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. $|\beta| > 1$

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.

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

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
\[
\frac{e^{j2\pi f}}{1 - \frac{1}{2} e^{-j2\pi f}},
\]
find \( H_d(f) \) for \( d = 1 \).