Solutions to Chapter 11

11.1. For minimum variance, we want to minimize diagonal terms of

$$\epsilon^2 \triangleq \overline{(\hat{Y}-Y)(\hat{Y}-Y)^T},$$

where $\hat{Y} = AX$.

$$\epsilon^2 = \overline{(AX-Y)(AX-Y)^T} = AK_1A^T + K_2 - AK_{12} - K_{21}A^T.$$

Now write $A = A_0 + \delta$; is $\delta = 0$ for minimum variance?

$$\begin{aligned}
\epsilon^2 &= (A_0+\delta)K_1(A_0+\delta)^T + K_2 - (A_0+\delta)K_{12} - K_{21}(A_0+\delta)^T \\
&= A_0K_1A_0^T + K_2 - A_0K_{12} - K_{21}A_0^T + \delta K_1\delta^T + A_0K_1\delta^T + \delta K_1A_0^T - \delta K_{12} - K_{21}\delta^T \\
&= \overline{(A_0X-Y)(A_0X-Y)^T} + \delta K_1\delta^T + A_0K_1\delta^T + \delta K_1A_0^T - \delta K_{12} - K_{21}\delta^T \\
&= \overline{(A_0X-Y)(A_0X-Y)^T} + \delta K_1\delta^T + K_{21}\delta^T + \delta K_{12} - \delta K_{12} - K_{21}\delta^T \\
&= \overline{(A_0X-Y)(A_0X-Y)^T} + \delta K_1\delta^T + 0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\text{tra}\epsilon^2 &= \text{tra}\left[(Y-A_0X)(Y-A_0X)^T\right] + \text{tra}\left(\delta K_1\delta^T\right) \\
&= \sum_{i=1}^{n}\left[y_i - \sum_j a_{ij}^{(0)}x_j\right]^2 + \text{tra}\left[\delta CC^T\delta^T\right] \\
&\geq 0.
\end{aligned}$$

$\text{tra}\left[\delta CC^T\delta^T\right] = \text{tra}\left[\delta C(\delta C)^T\right] \geq 0$ where we use factorization $K_1 = CC^T$.
Clearly set $\delta = 0$ for $\text{tra}\epsilon^2$ to be minimum.

11.2. Suppose $Y_1 = Y - \mu_2$ and $X_1 = X - \mu_1$. Then we know that $E[Y_1] = 0 = E[X_1]$, $K_1 = E[X_1X_1^T], K_2 = E[Y_1Y_1^T], K_{12} = E[X_1Y_1^T], K_{21} = E[Y_1X_1^T]$. Hence from problem 11.1, the minimum variance estimator Y_1 of the form $\hat{Y}_1 = AX_1$ is given by

$$\hat{Y}_1 = K_{21}K_1^{-1}(X-\mu_1).$$

But is $\hat{Y}_1$ equal to $\hat{Y} - \mu_2$?

Write $Y = Y_1 + \mu_2$. Let β be such that $\hat{Y} = \hat{Y}_1 + \beta$. Then

$$\begin{aligned}
\epsilon_T^2 &= \overline{(\hat{Y}-Y)(\hat{Y}-Y)^T} \\
&= \overline{[(\hat{Y}_1-Y_1)+(\beta-\mu_2)][(\hat{Y}_1-Y_1)+(\beta-\mu_2)]^T} \\
&= \overline{(\hat{Y}_1-Y_1)(\hat{Y}_1-Y_1)^T} + (\beta-\mu_2)(\beta-\mu_2)^T \\
&\quad \text{(because } \overline{(\hat{Y}_1-Y_1)} = E[K_{21}K_1^{-1}X_1] - E[Y_1] = 0 - 0 = 0) \\
&= \epsilon^2 + (\beta-\mu_2)(\beta-\mu_2)^T
\end{aligned}$$

$\text{tr}\epsilon_T^2 = \text{tr}\epsilon^2 + \text{tr}[(\beta-\mu_2)(\beta-\mu_2)^T] \geq 0$ is minimum when $\beta = \mu_2$. Hence

$$\hat{Y} = \hat{Y}_1 + \mu_2 = \mu_2 + K_{21}K_1^{-1}(X-\mu_1).$$

11.3. Given $\epsilon^2 = E[(X - E[X|Y])^2] = E[X(X - E[X|Y])] - E[E[X|Y][X - E[X|Y]]]$.
However, since the estimate is orthogonal to the error, we have

$$E[E[X|Y][X - E[X|Y]]] = 0. \tag{1}$$

Therefore,

$$\epsilon^2 = E[X[X - E[X|Y]]]. \tag{2}$$

By the same reasoning, from Eq. 1: $E[E[X|Y][X - E[X|Y]]] = 0 \implies E[XE[X|Y]] = E[[E[X|Y]]^2]$. But from Eq. 2,

$$\epsilon^2 = E[X^2] - E[XE[X|Y]] = E[X^2] - E[(E[X|Y])^2].$$

Now let us extend this to random N-vectors.

$$\begin{aligned} \epsilon^2 &= E[[X - E[X|Y]][X - E[X|Y]]^T] & (3) \\ &= E[XX^T] - E[XE[X|Y]^T] - E[E[X|Y]X^T] + E[E[X|Y]E[X|Y]^T]. & (4) \end{aligned}$$

However, using orthogonality again,

$$E[E[X|Y][X - E[X|Y]^T]] = 0$$

or

$$E[E[X|Y]X^T] = E[E[X|Y]E[X|Y]^T]. \tag{5}$$

Using Eq. 4 and Eq. 5 together,

$$\epsilon^2 = E[XX^T] - E[XE[X|Y]^T] = E[X[X - E[X|Y]]^T]. \tag{6}$$

Also:

$$E[XE[X|Y]^T] = E[E[X|Y]E^T[X|Y]]. \tag{7}$$

Using Eq. 6 and 7 together, we have

$$\epsilon^2 = E[XX^T] - E[E[X|Y]E^T[X|Y]].$$

11.4. $G[N] \triangleq E[X[n]|Y_N]$, $Y_N \triangleq [Y[n], \ldots, Y[n-N]]$.
From problem 8.55, we know $G[N]$ is a Martingale sequence in the parameter N.

$$\begin{aligned} E[X^2[n]] &= E[(X[n] - E[X[n]|Y_N] + E[X[n]|Y_N])^2] \\ &= E[(X[n] - E[X[n]|Y_N])^2] + E[E^2[X[n]|Y_N]]. \\ &\quad \text{(by the orthogonality principle)} \end{aligned}$$

$E[E^2[X[n]|Y_N]] \leq E[X^2[n]]$

$$\sigma_G^2[N] = E[E^2[X[n]|Y_N]] - E^2[E[X[n]|Y_N]] \leq E[X^2[n]] - E^2[X[n]].$$

Therefore, $\sigma_G^2[N] \leq \sigma_X^2[n]$. Let $C \triangleq \sigma_X^2[n] \implies \sigma_G^2[N] \leq C$ for all N. By the Martingale convergence theorem 8.8-4, we can conclude that the limit $\lim_{N\to\infty} E[X[n]|Y_N]$ exists with probability 1.

11.5. The modified Theorem 11.1-3 is given as:

Modified theorem: The LMMSE estimate of the zero-mean sequence $X[n]$ based on the zero-mean random sequence $Y[n]$'s $(p+1)$ most recent terms is

$$\hat{E}\left\{X[n]|Y[n],\ldots,Y[n-p]\right\} = \sum_{i=0}^{p} a_i^{(p)} Y[n-i],$$

where the $a_i^{(p)}$ satisfy the orthogonality condition

$$\left[X[n] - \sum_{i=0}^{p} a_i^{(p)} Y[n-i]\right] \perp Y[n-k], 0 \le k \le p.$$

Further, the LMMSE is given as

$$\epsilon_{\min}^{2(p)} = E\left\{|X[n]|^2\right\} - \sum_{i=0}^{p} a_i^{(p)} E\left\{Y[n-i]X^*[n]\right\}.$$

(a) Eq. (11.1-25) changes to

$$E[X[n]Y^*[n-k]] = \sum_{i=0}^{p} a_i^{(p)} E\left\{Y[n-i]Y^*[n-k]\right\}, 0 \le k \le p$$

via the data modification of $\{Y[n],\ldots,Y[0]\}$ being replaced by $\{Y[n],\ldots,Y[p]\}$. The equation (11.2-4) changes via

$$a^{(p)} \triangleq \left[a_0^{(p)},\ldots,a_p^{(p)}\right]$$

and

$$\underline{Y} \triangleq \left[Y[n], Y[n-1],\ldots,Y[n-p]\right],$$

thereby reversing time from the previous $\underline{Y}$. The cross-covariance vector $\underline{K}_{X\underline{Y}}$ then becomes

$$\underline{K}_{X\underline{Y}} = E\left\{X[n]\underline{Y}^*[n],\ldots,X[n]Y^*[n-p]\right\}$$

and the matrix vector equation (9.1-26) remains essentially the same at

$$\underline{a}^{(p)^T} = \underline{K}_{X\underline{Y}}\underline{K}_{\underline{YY}}^{-1}.$$

The entries in $\underline{K}_{\underline{YY}}^{-1}$ are given as

$$\left(\underline{K}_{\underline{YY}}\right)_{ij} = E\left[Y[n-i+1]Y^*[n-j+1]\right], 1 \le i,j \le p+1.$$

(b) The modified Eq. (11.1-27) is the same as before with the new $\underline{K}_{X\underline{Y}}$ and $\underline{K}_{\underline{YY}}$ inserted into the old version. Nothing else changes

$$\epsilon_{\min}^{2(p)} = \sigma_X^2(n) - \underline{K}_{X\underline{Y}}\underline{K}_{\underline{YY}}^{-1}\underline{K}_{X\underline{Y}}^T.$$

11.6. $X[n]$ is defined as

$$X[n] \triangleq -\sum_{k=1}^{n} \binom{k+2}{2} X[n-k] + W[n],$$

for $n = 1, 2, \ldots$ with $X[0] = W[0], W[n]$ is Gaussian noise with zero mean and unit variance.

(a) $W[n]$ is the innovation sequence for $X[n]$ because

- $W[n]$ is a white (uncorrelated) sequence.
- It is defined as a causal invertible linear transformation on $X[n]$.

(b) A simple substitution will show the result. From the defn. of $X[n]$,

$$W[n] = X[n] + \sum_{k=1}^{n} \binom{k+2}{2} X[n-k] = \sum_{k=0}^{n} \binom{k+2}{2} X[n-k].$$

Then, $W[n] - 2W[n-1] + 3W[n-2] - W[n-3]$
$= \sum_{k=0}^{n} \binom{k+2}{2} X[n-k] - 3\sum_{k=0}^{n-1} \binom{k+2}{2} X[n-1-k] + 3\sum_{k=0}^{n-2} \binom{k+2}{2} X[n-2-k] - \sum_{k=0}^{n-3} \binom{k+2}{2} X[n-3-k]$
$= X[n].$

(c) $X[n]$ is Gaussian since $W[n]$ is Gaussian. $\hat{X}[12|10]$
$\triangleq \hat{E}\left[X[12]|X[0], X[1], \ldots, X[10]\right]$
$= \hat{E}\left[X[12]|W[0], W[1], \ldots, W[10]\right]$
$= \hat{E}\left[W[12] - 3W[11] + 3W[10] - W[9]|W[0], W[1], \ldots, W[10]\right]$
$= \hat{E}\left[W[12]|W[0], W[1], \ldots, W[10]\right] - 3\hat{E}\left[W[11]|W[0], W[1], \ldots, W[10]\right]$
$+ 3\hat{E}\left[W[10]|W[0], W[1], \ldots, W[10]\right] - \hat{E}\left[W[9]|W[0], W[1], \ldots, W[10]\right]$
$= 3W[10] - W[9]$
$=$ (because $W[12] \perp [W[0], W[1], \ldots, W[10]]$ and $W[11] \perp [W[0], W[1], \ldots, W[10]]$)
For M.S. prediction error,

$$\begin{aligned}
E[(X[12] - \hat{X}[12|10])^2] &= E\left[((W[12] - 3W[11] + 3W[10] - W[9]) - (3W[10] - W[9]))\right] \\
&= E[(W[12] - 3W[11])^2] \\
&= E[W^2[12]] + 9E[W^2[11]] - 3E[W[12]W[11]] \\
&= 1 + 9 + 0 = 10.
\end{aligned}$$

$E[W[12]W[11]] = 0$ because $W[12] \perp W[11]$ and $W[n]$ is zero-mean.

11.7. (a) The innovations sequence is clearly $W[n]$, because $X[n] = X[n-1] + W[n]$.

(b)

$$\hat{X}[n|n] = \hat{X}[n-1|n-1] = G_n \left(Y[n] - \hat{X}[n-1|n-1]\right).$$

(c)

$$G_n = \epsilon^2[n]\left[\epsilon^2[n] + \sigma_v^2\right]^{-1}; n \geq 1.$$
$$\epsilon^2[n] = \epsilon^2[n-1](1 - G_{n-1}) + 1; n \geq 1.$$

$\epsilon^2[0] = E[X^2[0]] = 0.$

11.8. Given $2Y[n+2] + Y[n+1] + Y[n] = 2W[n]$, $W[n] \sim N(0,1)$, $Y[0] = 0$, $Y[1] = 1$, we have for state-space representation

$$\begin{bmatrix} Y[n+2] \\ Y[n+1] \end{bmatrix} = \begin{bmatrix} -0.5 & -0.5 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} Y[n+1] \\ Y[n] \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} W[n].$$

If the state vector $X[n]$ is defined as

$$\underline{X}[n] \triangleq \begin{bmatrix} Y[n+2] \\ Y[n+1] \end{bmatrix},$$

we have

$$\underline{X}[n] = \begin{bmatrix} -0.5 & -0.5 \\ 1 & 0 \end{bmatrix} \underline{X}[n-1] + \begin{bmatrix} 1 \\ 0 \end{bmatrix} W[n].$$

Also, $E[2Y[n+2] + Y[n+1] + Y[n]] = 2E[W[n]]$, and so we have

$$2\mu_Y[n+2] + \mu_Y[n+1] + \mu_Y[n] = 2\mu_W[n]$$

. Obviously, $\mu_Y[0] = 0$ and $\mu_Y[1] = 1$. Then,
$2\mu_Y[2] = -\mu_Y[1] - \mu_Y[0] = -1$ or $\mu_Y[2] = -0.5$.
$2\mu_Y[3] = -\mu_Y[2] - \mu_Y[1] = -0.5$ or $\mu_Y[3] = -0.25$. etc.
We also have $E\left[[2Y[n+2] + Y[n+1] + Y[n]]Y[n]\right] = 2E[W[n]Y[n]] = 0$.
$E\left[[2Y[n+2] + Y[n+1] + Y[n]]Y[n+1]\right] = 2E[W[n]Y[n+1]] = 0$.
Therefore, we can calculate $R[n_1, n_2]$ for all n_1, n_2. Note for initial conditions, we have

(a) $R_Y[0,0] = 0 = R_Y[0,1] = R_Y[1,0]$, and $R_Y[1,1] = 1, R_Y[2,1] = -0.5$ etc.

(b) $R_Y[n_1, n_2] = R_Y[n_2, n_1]$ since Y is a real sequence.

11.9. (a) From the equation, we have $X[n] = AX[n-1] + BW[n]$. We have

$$\mu_X[n] \triangleq E[X[n]] = E\left[AX[n-1] + BW[n]\right] = A\mu_X[n-1] + B\mu_W[n].$$

From the equation, $Y[n] = X[n] + V[n]$. We also have

$$\mu_Y[n] \triangleq E[Y[n]] = E[X[n]] + E[V[n]] = \mu_X[n] + \mu_V[n].$$

(b) Since $\mu_X[n], \mu_Y[n]$ are deterministic variables, we have

$$\begin{aligned}
\hat{X}[n|n] &\triangleq E[X[n]|Y[0], \ldots, Y[n]] \\
&= E[X_C[n] + \mu_X[n]|Y[0], \ldots, Y[n]] \\
&= E[X_C[n]|Y[0], \ldots, Y[n]] + \mu_X[n] \\
&= E[X_C[n]|Y_C[0] + \mu_Y[0], \ldots, Y_C[n] + \mu_Y[n]] + \mu_X[n] \\
&= E[X_C[n]|Y_C[0], \ldots, Y_C[n]] + \mu_X[n] \\
&= \hat{X}_C[n|n] + \mu_X[n].
\end{aligned}$$

(c) Since $X_C[n] = AX_C[n-1] + BW_C[n]$ and $Y_C[n] = X_C[n] + V_C[n]$ are the same as in (11.2-6 and 7), we can use the same estimate equation to estimate $\hat{X}_C[n|n]$, i.e.,

$$\hat{X}_C[n|n] = A\hat{X}_C[n-1|n-1] + G_n\left[Y_C[n] - A\hat{X}[n-1|n-1]\right].$$

So,

$$\hat{X}_C[n|n] = A\hat{X}[n-1|n-1] + G_n\left[Y[n] - A\hat{X}[n-1|n-1]\right] - A\mu_X[n-1] - G_n\mu_Y[n] + G_nA\mu_X[n-1].$$

Therefore,

$$\hat{X}_C[n|n] = A\hat{X}[n-1|n-1] + G_n\left[Y[n] - A\hat{X}[n-1|n-1]\right] - (G_nA - A)\mu_X[n-1] - G_n\mu_Y[n].$$

(d) Since the gain and error covariance equations just depend on $\sigma_V^2[n], \sigma_W^2[n]$ and dynamical model's coefficients, the gain and error covariance equations do not change.

11.10. $Y[n] = C_n X[n] + V[n]$

We arrive at following equation simply by following the procedure developed in the book.

$$\hat{X}[n] = A_n \left[(I - G_{n-1} C_{n-1}) \hat{X}[n-1] + G_{n-1} Y[n-1] \right],$$

with

$$G_n = E\left[X[n] \tilde{Y}^T[n] \right] \left[\sigma^2_{\tilde{Y}}[n] \right]^{-1}$$

and $\tilde{Y}[n] = Y[n] - C_n \hat{X}[n]$ where $\hat{X}[n] \triangleq \hat{X}[n|n-1]$. However,

$$\begin{aligned} E[X[n]\tilde{Y}^T[n]] &= E\left[X[n](Y[n] - C_n \hat{X}[n])^T \right] \\ &= E\left[X[n](C_n X[n] + V[n] - C_n \hat{X}[n])^T \right] \\ &= E\left[X[n](X[n] - \hat{X}[n])^T \right] C_n^T. \end{aligned}$$

By the orthogonality principe, $E\left[\hat{X}[n](X[n] - \hat{X}[n])^T \right] = 0$. So

$$E[X[n]\tilde{Y}^T[n]] = E\left[(X[n] - \hat{X}[n])(X[n] - \hat{X}[n])^T \right] C_n^T = \epsilon^2[n] C_n^T.$$

Also

$$\begin{aligned} \sigma^2_{\tilde{Y}}[n] &= E[\tilde{Y}[n]\tilde{Y}^T[n]] \\ &= E\left[\left(C_n(X[n] - \hat{X}[n]) + V_n \right) \left(C_n(X[n] - \hat{X}[n]) + V_n \right)^T \right] \\ &= C_n \epsilon^2[n] C_n^T + \sigma^2_V[n]. \end{aligned}$$

So, $G[n] = \epsilon^2[n] C_n^T \left[C_n \epsilon^2[n] C_n^T + \sigma^2_V[n] \right]^{-1}$. We know,

$$\epsilon^2[n] = E\left[(X[n] - \hat{X}[n])(X[n] - \hat{X}[n])^T \right] = E\left[X[n] X^T[n] \right] - E\left[\hat{X}[n] \hat{X}^T[n] \right].$$

Now, $X[n] = A_n X[n-1] + B_n W[n]$ and $\hat{X}[n] = A_n \left[\hat{X}[n-1] + G_{n-1} \tilde{Y}[n-1] \right]$. Therefore,

$$E[X[n]X^T[n]] = A_n E\left[X[n-1] X^T[n-1] \right] A_n^T + B_n \sigma^2_W[n] B_n^T,$$

and also

$$E[\hat{X}[n]\hat{X}^T[n]] = A_n E\left[\hat{X}[n-1] \hat{X}^T[n-1] \right] A_n^T + A_n G_{n-1} E\left[\tilde{Y}[n-1] \tilde{Y}^T[n-1] \right] G_{n-1}^T A_n^T.$$

Hence,

$$\epsilon^2[n] = A_n \left(\epsilon^2[n-1] - G_{n-1} \sigma^2_{\tilde{Y}}[n-1] G_{n-1}^T \right) A_n^T + B_n \sigma^2_W[n] B_n^T.$$

From the result above,

$$G_{n-1} \sigma_{\tilde{Y}}[n-1] G_{n-1}^T = E\left[X[n-1] \tilde{Y}^T[n-1] \right] G_{n-1}^T = \epsilon^2[n-1] C_{n-1}^T G_{n-1}^T.$$

Thus, finally

$$\epsilon^2[n] = A_n \epsilon^2[n-1] \left(I - C_{n-1}^T G_{n-1}^T \right) A_n^T + B_n \sigma^2_W[n] B_n^T.$$

11.11. (a) Since $\tilde{Y}[-N] \perp \tilde{Y}[-N+1] \perp \ldots \tilde{Y}[N]$. Using Theorem 11.1-4 property (b), we have

$$\hat{E}\left[X[n]|\tilde{Y}[-N], \tilde{Y}[-N+1], \ldots, \tilde{Y}[N]\right] = \hat{E}[X[n]|\tilde{Y}[-N]] + \hat{E}\left[X[n]|\tilde{Y}[-N+1], \ldots, \tilde{Y}[N]\right].$$

By the same procedure

$$\hat{E}\left[X[n]|\tilde{Y}[-N], \tilde{Y}[-N+1], \ldots, \tilde{Y}[N]\right] = \sum_{k=-N}^{N} \hat{E}[X[n]|\tilde{Y}[k]].$$

Let $\hat{E}\left[X[n]|\tilde{Y}[k]\right] \triangleq g[k]\tilde{Y}[k]$. Then we have

$$\hat{E}\left[X[n]|\tilde{Y}[-N], \tilde{Y}[-N+1], \ldots, \tilde{Y}[N]\right] = \sum_{k=-N}^{N} g[k]\tilde{Y}[k].$$

(b) Let $\hat{X}[N] = \hat{E}\left[X[n]|\tilde{Y}[-N], \ldots, \tilde{Y}[N]\right]$.

$\because E\left[\hat{X}[N]|\hat{X}[0], \hat{X}[1], \ldots, \hat{X}[N-1]\right] = \hat{X}[N-1]$ since $E \triangleq \hat{E}$ in Gaussian case.

$\therefore \hat{E}\left[X[n]|\tilde{Y}[-N], \ldots, \tilde{Y}[N]\right]$ is a Martingale sequence. Using the result of problem 11.4, we conclude

$$\lim_{N\to\infty} \hat{E}\left[X[n]|\tilde{Y}[-N], \ldots, \tilde{Y}[N]\right] = \lim_{N\to\infty} \sum_{k=-N}^{N} g[k]\tilde{Y}[k]$$

exists with probability 1.

11.12. $\hat{R}_N[m] = \frac{1}{N}\sum_{n=0}^{N-1} x[n+m]x^*[n]$

(a)

$$E\left\{\hat{R}_N[m]\right\} = E\left\{\frac{1}{N}\sum_{n=0}^{N-1} x[n+m]x^*[n]\right\} = \frac{1}{N}\sum_{n=0}^{N-1} E\left\{x[n+m]x^*[n]\right\} = R_X[m]$$

(b) Show that $\lim_{N\to\infty} E\left\{|\hat{R}_N[m] - R_X[m]|^2\right\} = 0$.

$\lim_{N\to\infty} E\left\{|\hat{R}_N[m] - R_X[m]|^2\right\}$

$= \lim_{N\to\infty} E\left\{(\hat{R}_N[m] - R_X[m])(\hat{R}^*_N[m] - R^*_N[m])\right\}$

$= \lim_{N\to\infty} E\left\{\hat{R}_N[m]\hat{R}^*_N[m] - \hat{R}_N[m]\hat{R}^*_X[m] - \hat{R}^*_N[m]R_X[m] + R_X[m]R^*_X[m]\right\}$

$= \lim_{N\to\infty} E\left\{\hat{R}_N[m]\hat{R}^*_N[m] - R_X[m]R^*_X[m]\right\}$.

Now apply 4^{th} order moment property for (complex) Gaussian random variables to get

$E\left\{\hat{R}_N[m]\hat{R}^*_N[m]\right\}$

$= \frac{1}{N^2}E\left\{\sum_{n_1=0}^{N-1} X[n_1+m]X^*[n_1]\sum_{n_2=0}^{N-1} X^*[n_2+m]X[n_2]\right\}$

$= \frac{1}{N^2}\sum_{n_1,n_2} E\left\{X[n_1+m]X^*[n_1]X^*[n_2+m]X[n_2]\right\}$

$= \frac{1}{N^2}\sum_{n_1,n_2} E\left\{X[n_1+m]X^*[n_1]\right\}E\left\{X[n_2+m]X^*[n_2]\right\}$

$+\frac{1}{N^2}\sum_{n_1,n_2} E\{X[n_1+m]X^*[n_2+m]\} E\{X^*[n_1]X[n_2]\}$
$+$ extra term which is zero if $X[n]$ has symmetry condition as in (10.6-4) and (10.6-5) which can be restated for a complex random sequence $X[n]=X_r[n]+jX_i[n]$ as
$K_{X_rX_r}[m]=K_{X_iX_i}[m]$ and $K_{X_rX_i}[m]=-K_{X_iX_r}[m]$.
(NOTE: $K[.]=R[.]$ because $X[n]$ is zero mean.)
In this symmetric case, which occurs for the bandpass random process of section 10.6, we see that $E\{X[n_1+m]X[n_2]\}$ is zero as follows:
$E\{X[n_1+m]X[n_2]\}=$
$=E\{(X_r[n_1+m]+jX_i[n_1+m])(X_r[n_2]+jX_r[n_2])\}$
$=E\{X_r[n_1+m]X_r[n_2]-X_i[n_1+m]X_i[n_2]\}+jE\{X_i[n_1+m]X_r[n_2]+X_r[n_1+m]X_i[n_2]\}$
$=K_{X_rX_r}[n_1-n_2+m]-K_{X_iX_i}[n_1-n_2+m]+j\left(K_{X_iX_r}[n_1-n_2+m]+K_{X_rX_i}[n_1-n_2+m]\right)$
$=0+j0=0$
(For more on complex Gaussian random processes and random sequences, see *Discrete Random Signals and Statistical Signal Processing*, C. W. Therrien, Prentice-Hall, 1992.)

So,

$$\begin{aligned}
E\left\{|\hat{R}_N[m]|^2\right\} &= R_X[m]R_X^*[m]+\frac{1}{N^2}\sum_{n_1,n_2=0}^{N-1} R_X[n_1-n_2]R_X^*[n_1-n_2] \\
&= |R_X[m]|^2+\sum_{n=-(N-1)}^{N-1}\frac{N-|n|}{N^2}|R_X[n]|^2.
\end{aligned}$$

Thus $\lim_{N\to\infty} E\left\{|\hat{R}_N[m]-R_X[m]|^2\right\}\le \lim_{N\to\infty}\frac{1}{N}\sum_{m=-\infty}^{\infty}|R_X[m]|^2=0$ for square summable $R_X[m]$.
(NOTE: For real-valued random sequence, get two $O\left(\frac{1}{N}\right)$ terms.)

11.13. (a) The solution to this part is the same as the solution to Problem 8.32 (refer). The power spectral density is given by

$$S_{XX}(w)=10\left(\frac{1-\rho_1^2}{1+\rho_1^2-2\rho_1\cos w}\right)+5\left(\frac{1-\rho_2^2}{1+\rho_2^2-2\rho_2\cos w}\right),$$

where $\rho_1=e^{-\lambda_1},\rho_2=e^{-\lambda_2}$.

(b)

$$\begin{aligned}
E\{I_N(w)\} &= \sum_{m=-(N-1)}^{N-1}\frac{N-|m|}{N}R_X[m]e^{-jwm} \\
&= \sum_{m=0}^{N-1}\frac{N-m}{N}\left(10e^{-\lambda_1 m-jwm}+5e^{-\lambda_2 m-jwm}\right)+\sum\left(10e^{\lambda_1 m-jwm}+5e^{\lambda_2 m-jwm}\right).
\end{aligned}$$

So, $\lim_{N\to\infty}E\{I_N(w)\}=S_X(w)+\lim_{N\to\infty}\left(\sum_{m=0}^{N-1}\frac{-m}{N}\left(10e^{-\lambda_1 m-jwm}+5e^{-\lambda_2 m-jwm}\right)\right)$.
Now, $\left|\sum_{m=0}^{N-1}me^{-\lambda m-jwm}\right|\le\sum_{m=0}^{N-1}me^{-\lambda m}$ for any $\lambda>0$, and $\sum_{m=0}^{\infty}me^{-\lambda m}>\infty$.
Hence $\lim\frac{1}{N}\sum me^{-\lambda m}\to 0$. Thus we can conclude for this case that

$$\lim_{N\to\infty}E\{I_N(w)\}=S_X(w).$$

11.14. $r_0 a_1 = r_1$
$\sigma_X^2 a_1 = \sigma_X^2 \rho \implies a_1 \rho$
$\sigma_e^2 = r_0 - \sum_{m=1}^{p} a_m r_m = \sigma_X^2 - \rho^2 \sigma_X^2 = \sigma_X^2(1-\rho^2)$.

$$\begin{aligned} S_X(w) &= \frac{1}{\frac{1}{\sigma_e^2}\left|1 - \sum_{m=1}^{p} a_m e^{-jwm}\right|^2} \\ &= \frac{\sigma_x^2(1-\rho^2)}{|1-\rho e^{-jw}|^2} \\ &= \frac{\sigma_x^2(1-\rho^2)}{1 - 2\rho\cos w + \rho^2}, |w| \le \pi. \end{aligned}$$

11.15. The Matlab code (below) uses an AR(3) model to generate the N point random sample. Figure 1 is an estimate of the correlation function, computed with $N = 100$. The bottom axis should run from -100 to 100, as the zero shift value for $R[m]$ estimate is in the middle of the plot. Following the first plot, are three AR(3) spectral estimates for $N = 25, 100$, and 512 data points (Fig. 2). Also on each plot is the true psd for our AR(3) model.

```
%This program generates an AR3 estimate of psd of AR3 model.
clear
Pi=3.1415927;
disp('This .m file computes ar(3) parametric psd estimate.');
N=input('choose data length (<=512) = ');
randn('state',0);
w=randn(N,1);
bt=[1.0 0.0 0.0 0.0];
at=[1.0 -1.700 1.530 -0.648];
disp('The true a vector is a = [1.0 -1.7 1.53 -0.648].');
x=filter(bt,at,w);
y=flipud(x);
disp('Now calculating estimate of R[m].');
z=(1./N)*conv(x,y);
figure(1)
plot(z)
title('estimate of correlation function');
pause(5);
R(1,1)=z(N);
R(2,2)=z(N);
R(3,3)=z(N);
R(1,2)=z(N+1);
R(2,3)=z(N+1);
R(1,3)=z(N+2);
R(2,1)=R(1,2);
R(3,2)=R(2,3);
R(3,1)=R(1,3);
R
pause(5);
```

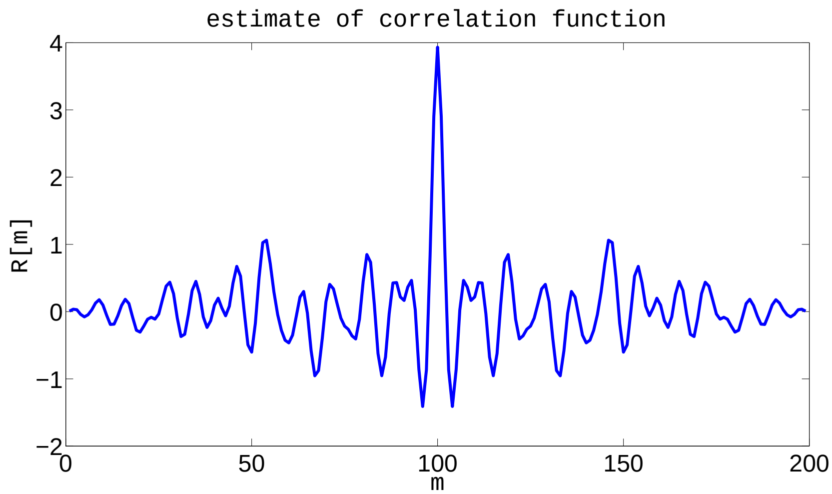
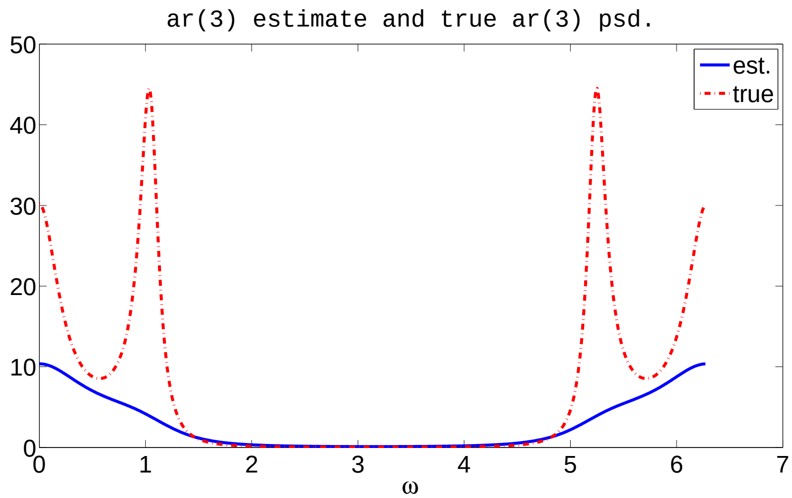


Figure 1: Estimate of correlation function for $N = 100$.

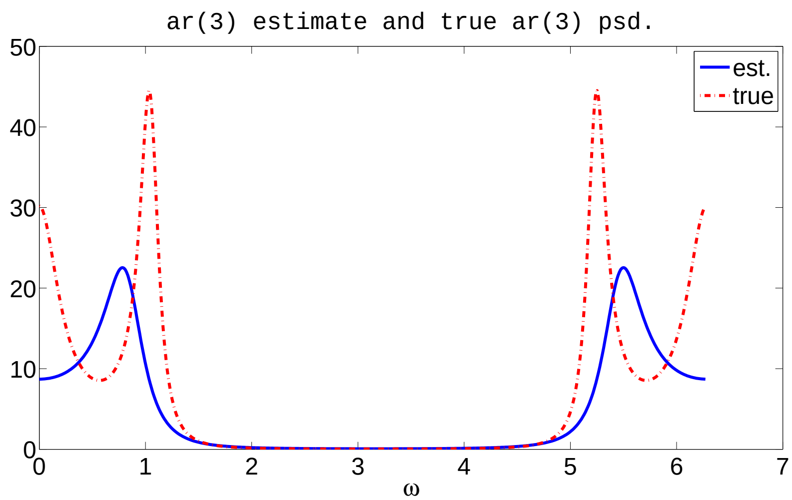
```
r(1,1)=z(N+1);
r(2,1)=z(N+2);
r(3,1)=z(N+3);
r
pause(5);
a=inv(R)*r
disp('May compare a vector values to ar(3) coefficients.');
b1=[1.0 0.0 0.0 0.0];
a1=[1.0 -a(1) -a(2) -a(3)];
[h,w]=freqz(b1,a1,512,'whole');
s=h.*conj(h);
figure(2)
plot(w,s);
title('ar(3) power spectral density estimate');
pause(3);
[ht,w]=freqz(bt,at,512,'whole');
st=ht.*conj(ht);
figure(3)
plot(w,st);
title('true ar(3) power spectral density');
pause(3);
figure(4)
plot(w,s,w,st)
title('ar(3) estimate and true ar(3) psd.');
text(0,-5,'Please press any key to exit.');
```

11.16. The Matlab function VITERBIPATH given below computes the most likely state sequences for the observations.

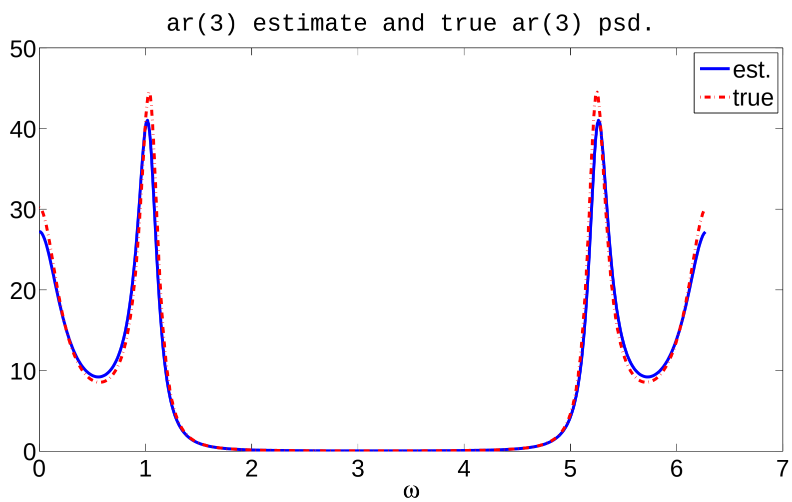
```
% The following function computes the most likely state seqeuence
```



(a) $N = 25$



(b) $N = 100$



(c) $N = 512$

Figure 2: Comparison of the true spectrum and AR(3) spectral estimate for different values of N.

```
% accounting for the observations.
% N is the number of states and must be an integer.
% L is either the number of observations or the maximum value f
%   the discrete time index
% O is the observation vector and in this program must correspond to
%   the columns of B. For example if you observe
%   {heads, heads, tails, heads, tails} the observation vector would be
%   {1,1,2,1,2} where a 1 would correspond to a head and a 2 would
%   correspond to a tail.
% A is the sate transitionn probability matrix and is 2x2 in this case
%   since there are only two states 'heads' and 'tails'.
% B is the state-conditional output probability matrix; the first column
%   of B is [P[head|state1],P[head|state2]]
% PINT is the inital state probability row vector

function [PSTAR, Q] = VITERBIPATH(N,L,O,A,B,PINT)

% Initialize
Q  = zeros(1,L);
psi = zeros(N,L);
phi = zeros(N,L);
for i = 1:N
phi(i,1) = PINT(i)*B(i,O(1));
end

% Iterate forward
for j = 2:L,
for i = 1:N,
for id = 1:N,
phiid(id) = phi(id,j-1) * A(id,i) * B(i,O(j));
psiid(id) = phi(id,j-1) * A(id,i);
end
phi(i,j) = max(phiid);
[junk, psi(i,j)] = max(psiid);
end
end

% Iterate backwards to find path
[PSTAR, Q(L)] = max(phi(:,L));
for n = L-1:-1:1
Q(n) = psi(Q(n+1), n+1);
end

% Graphical output
figure(1);
clf % clears current window
% The next instruction creates a graphical rectangular space of x-dimension
% going from 0 to L+1 and y-dimension from 0 to N+1.
```

```
axis([0 L+1 0 N+1]);
hold on; % keeps the current graphics
for j = 1:L
plot(j*ones(1,N),1:N,'o');
end
plot(1:L, Q,'-');  % Q is a vector whose components are the states
                       % at time l = 1, ..., L
% Hence a solid line segment drawn from the previous node to the new node
% at tile l.

set(gca, 'XTickLabel', [1:L]); % puts the numbers 1,2,...,L at the ticks
set(gca, 'YTick'     , [1:N]); % puts tick marks on y-label
set(gca, 'YTickLabel', [1:N]); % puts the numbers 1,2,...,N at the ticks

xlabel('time index');
ylabel('state'     );
title('most likely state path determined by Viterbi algorithm');
% grid
hold off; % liberates you from restrictions of
                % the graphical environment above
```

We run the matlab file to check the answer in Example 11.5-5.

```
>> VITERBIPATH(2, 3, [1,2,2], [0.6,0.4,0.3,0.7], [0.3,0.7,0.6,0.4], [0.7,0.3])
ans = 0.0370
```

is the highest probability of observing this sequence probability Now that we are confident that the program works we can try it on the observation vector head, head, head = 1,1,1. The output figure obtained is given in Fig. 3.

```
>> VITERBIPATH(2, 3, [1,1,1], [0.6,0.4,0.3,0.7], [0.3,0.7,0.6,0.4], [0.7,0.3])
ans = 0.0318
```

11.17. The most likely state path is given in Fig. 4.

```
>> VITERBIPATH(2, 3, [1,2,2,1,1], [0.6,0.4,0.3,0.7], [0.3,0.7,0.6,0.4], [0.7,0.3])
ans = 0.0318
```

11.18. We are given

$$Y_1 = \frac{2}{3}X_1 + \frac{1}{3}X_2, Y_2 = \frac{1}{3}X_1 + \frac{2}{3}X_2,$$

or

$$X_1 = 2Y_1 - Y_2, X_2 = -Y_1 + 2Y_2,$$

where X_1: Poisson with λ_1, X_2: Poisson with λ_2, and $P_{X_1X_2}[m,k] = P_{X_1}[m]P_{X_2}[k]$.

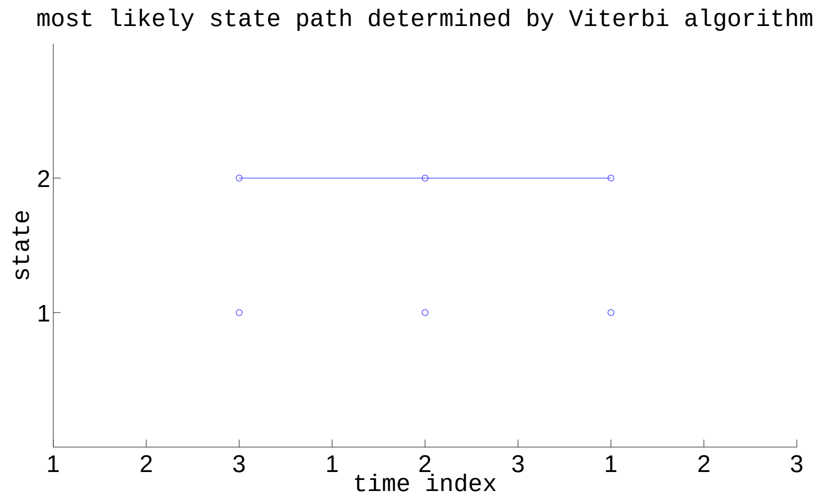


Figure 3:

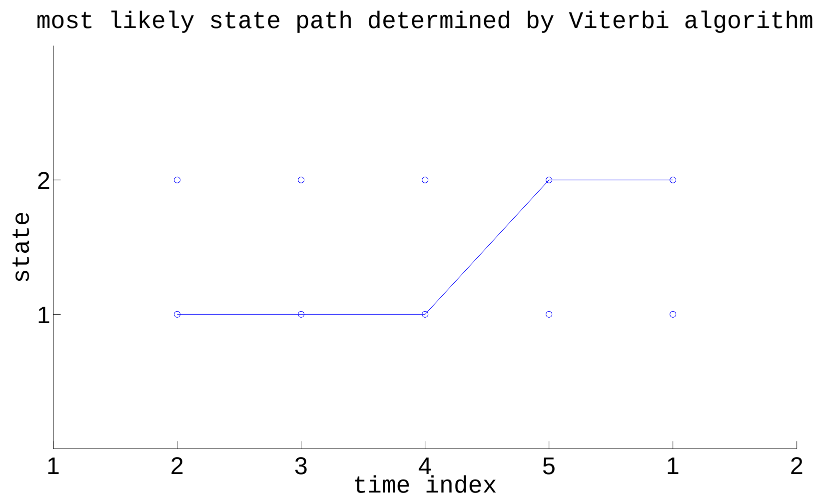


Figure 4:

We know that the maximum likelihood estimators for λ_1, λ_2 are given by

$$\hat{\lambda}_1^{(ml)} = X_1 = 2Y_1 - Y_2,$$

$$\hat{\lambda}_2^{(ml)} = X_2 = -Y_1 + 2Y_2.$$

Now let us see how the EM algorithm yields this result:
E-step: $\tilde{X}^{(k+1)} = E[\tilde{X}^k|\tilde{Y};\tilde{\lambda}^{(k)}]$,
M-step: $\tilde{\lambda}^{(k+1)} = \arg\max_{\tilde{\theta}} \left\{ -\sum_{i=1}^n \theta_i + \tilde{X}^{(k+1)}\tilde{\Gamma} \right\}$,
where $\tilde{\Gamma} \triangleq (\log\theta_1, \log\theta_2, \ldots, \log\theta_n)$. To compute $E[\tilde{X}^{(h)}|\tilde{Y};\tilde{\lambda}^{(h)}]$, we need
$P\left[X_1^{(h)} = x_1, X_2^{(h)} = x_2 | Y_1 = y_1, Y_2 = y_2; \lambda_1^{(h)}, \lambda_2^{(h)}\right]$
$= P\left[X_1^{(h)} = x_1 | Y_1 = y_1; \lambda_1^{(h)}]P[X_2^{(h)} = x_2 | Y_2 = y_2; \lambda_2^{(h)}\right]$;
since X_1, X_2 are independent, $X_1^{(h)}, X_2^{(h)}$ are also independent. However, in this case

$$P[X_1^{(h)} = x_1|\tilde{Y} = \tilde{y}] = \begin{cases} 1, & x_1 = 2y_1 - y_2 \\ 0, & \text{else.} \end{cases}.$$

Likewise

$$P[X_2^{(h)} = x_2|\tilde{Y} = \tilde{y}] = \begin{cases} 1, & x_2 = -y_1 + 2y_2 \\ 0, & \text{else.} \end{cases},$$

independent of the index h. Hence $X_1^{(h+1)} = X_1^{(h)}, X_2^{(h+1)} = X_2^{(h)}$ or for all k

$$X_1^{(k)} = 2Y_1 - Y_2, X_2^{(k)} = -Y_1 + 2Y_2.$$

Thus the E-step becomes

$$\tilde{X}^{(h+1)} = E[\tilde{X}^{(h)}|\tilde{Y}, \tilde{\theta}^{(h)}] = (2Y_1 - Y_2, -Y_1 + 2Y_2)^T.$$

And the M-step becomes

$$\tilde{\lambda}^{(h+1)} = \arg\max_{\tilde{\theta}} \left\{ -\sum_{i=1}^2 \theta_i + X_1^{(h)}\log\theta_1 + X_2^{(h)}\log\theta_2 \right\} \triangleq \arg\max_{\tilde{\theta}} \Lambda(\tilde{\theta}).$$

$\frac{\partial\Lambda}{\partial\theta_1} = 0 \implies 1 = \frac{X_1^{(h+1)}}{\theta_1}$ or $\theta_1 = \lambda_1^{(ml)} = 2Y_1 - Y_2$. $\frac{\partial\Lambda}{\partial\theta_2} = 0 \implies 1 = \frac{X_2^{(h+1)}}{\theta_2}$ or $\theta_2 = \lambda_2^{(ml)} = -Y_1 + 2Y_2$,
independent of k. Hence the E-M algorithm will converge after one iteration.

11.19. From the cited equation, we have

$$\sigma_U^2 \left| 1 - \sum_k a_k e^{-jwk} \right|^2 = \sigma_W^2 \left(1 - \sum_{k \neq 0} c_k e^{-jwk} \right)^2,$$

where σ_U^2 is the minimum M.S. interpolation error and σ_W^2 is the minimum M.S. prediction error, with the respective predictor coefficients $a_k, k = 1, \ldots, p$ and interpolator coefficients $c_k = c_{-k}, k = 1, \ldots, p$ $(c_0 = a_0 = 0)$.

This equation is between two polynomials in the variable e^{-jw}. Setting the constant terms equal, we get $\sigma_k^2 \left(1 + \sum_{k=1}^{p} |a_k|^2\right) = \sigma_W^2$ or

$$\sigma_U^2 = \frac{\sigma_W^2}{1 + \sum |a_k|^2}$$

where $\sigma_U^2 < \sigma_W^2$ if any $a_k \neq 0, k = 1, \dots, p$.

11.20. Set $D^{(k)}[n] \triangleq X[n] - X^{(k+1)}[n]$. Then from the two given equations,

$$D^{(k)}[n] = C * D^{(k-1)}[n] \frac{\sigma_v^2}{\sigma_u^2 + \sigma_v^2} = \sum_l c_l D^{(k-1)}[n-l] \frac{\sigma_v^2}{\sigma_u^2 + \sigma_v^2}.$$

Consider an interval $[-N, N]$ where the solution will be simulated for some large $N < \infty$. Then define the norm

$$||D^{(k)}|| \triangleq \max_{|n| \leq N} |D^{(k)}[n]|.$$

We have

$$\begin{aligned} ||D^{(k)}|| &= \max_{|n| \leq N} \left| \sum_l c_l D^{(k-1)}[n-l] \right| \frac{\sigma_v^2}{\sigma_u^2 + \sigma_v^2} \\ &\leq \left(\sum |c_l| \right) \max_{|n| \leq N} \left| D^{(k-1)}[n-l] \right| \frac{\sigma_v^2}{\sigma_u^2 + \sigma_v^2} \\ &\leq \rho ||D^{(k-1)}||, \end{aligned}$$

where $\rho \triangleq \frac{\sigma_v^2}{\sigma_u^2 + \sigma_v^2} \sum_l |c_l| < 1$. Note that in the above equation, the maximum of $|n|$ should be less than or equal to $N - p$ where p is the order of the Markov model, but $N - p \approx N$.

So, $||D^{(k)}|| \leq \rho ||D^{(k-1)}||, k \geq 1$. As $k \to \infty$, clearly $||D^{(k)}|| \to 0$ and so

$$X^{(k)}[n] \to X[n] \text{ for } -N < n < N.$$

NOTE: We cannot let $N = +\infty$ because then the norm $||D||$ defined as $\max |D[n]|$ might be infinite. Still, N is large compared to the Markov order p should be sufficient. If we know the $Y[n]$ are finite valued, then we could let $N = +\infty$, because the equations are stable (BIBO stable).