ECE 541
Stochastic Signals and Systems
Problem Set 9 Solutions

Problem Solutions : Yates and Goodman, 9.1.4 9.2.2 9.2.6 9.3.2 9.4.2 9.4.6 9.4.7 and 9.5.3

Problem 9.1.4 Solution
The joint PDF of X and Y is

$$f_{X,Y}(x, y) = \begin{cases} 6(y - x) & 0 \le x \le y \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

(a) The conditional PDF of X given Y is found by dividing the joint PDF by the marginal with respect to Y. For $y < 0$ or $y > 1$, $f_Y(y) = 0$. For $0 \le y \le 1$,

$$f_Y(y) = \int_0^y 6(y - x)\, dx = 6xy - 3x^2 \big|_0^y = 3y^2 \tag{2}$$

The complete expression for the marginal PDF of Y is

$$f_Y(y) = \begin{cases} 3y^2 & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Thus for $0 < y \le 1$,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \begin{cases} \frac{6(y-x)}{3y^2} & 0 \le x \le y \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

(b) The minimum mean square estimator of X given $Y = y$ is

$$\hat{X}_M(y) = E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx \tag{5}$$

$$= \int_0^y \frac{6x(y - x)}{3y^2}\, dx = \frac{3x^2 y - 2x^3}{3y^2} \bigg|_{x=0}^{x=y} = y/3 \tag{6}$$

(c) First we must find the marginal PDF for X. For $0 \le x \le 1$,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \int_x^1 6(y - x)\, dy = 3y^2 - 6xy \big|_{y=x}^{y=1} \tag{7}$$

$$= 3(1 - 2x + x^2) = 3(1 - x)^2 \tag{8}$$

The conditional PDF of Y given X is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \begin{cases} \frac{2(y-x)}{1-2x+x^2} & x \le y \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

(d) The minimum mean square estimator of Y given X is

$$\hat{Y}_M(x) = E\left[Y|X = x\right] = \int_{-\infty}^{\infty} y f_{Y|X}\left(y|x\right) \, dy \tag{10}$$

$$= \int_x^1 \frac{2y(y - x)}{1 - 2x + x^2} \, dy \tag{11}$$

$$= \left.\frac{(2/3)y^3 - y^2x}{1 - 2x + x^2}\right|_{y=x}^{y=1} = \frac{2 - 3x + x^3}{3(1 - x)^2}. \tag{12}$$

Perhaps surprisingly, this result can be simplified to

$$\hat{Y}_M(x) = \frac{x}{3} + \frac{2}{3}. \tag{13}$$

Problem 9.2.2 Solution

The problem statement tells us that

$$f_V\left(v\right) = \begin{cases} 1/12 & -6 \le v \le 6, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Furthermore, we are also told that $R = V + X$ where X is a Gaussian $(0, \sqrt{3})$ random variable.

(a) The expected value of R is the expected value V plus the expected value of X. We already know that X has zero expected value, and that V is uniformly distributed between -6 and 6 volts and therefore also has zero expected value. So

$$E\left[R\right] = E\left[V + X\right] = E\left[V\right] + E\left[X\right] = 0. \tag{2}$$

(b) Because X and V are independent random variables, the variance of R is the sum of the variance of V and the variance of X.

$$\text{Var}[R] = \text{Var}[V] + \text{Var}[X] = 12 + 3 = 15. \tag{3}$$

(c) Since $E[R] = E[V] = 0$,

$$\text{Cov}\left[V, R\right] = E\left[VR\right] = E\left[V(V + X)\right] = E\left[V^2\right] = \text{Var}[V]. \tag{4}$$

(d) The correlation coefficient of V and R is

$$\rho_{V,R} = \frac{\text{Cov}\left[V, R\right]}{\sqrt{\text{Var}[V]\,\text{Var}[R]}} = \frac{\text{Var}[V]}{\sqrt{\text{Var}[V]\,\text{Var}[R]}} = \frac{\sigma_V}{\sigma_R}. \tag{5}$$

The LMSE estimate of V given R is

$$\hat{V}(R) = \rho_{V,R}\frac{\sigma_V}{\sigma_R}\left(R - E\left[R\right]\right) + E\left[V\right] = \frac{\sigma_V^2}{\sigma_R^2}R = \frac{12}{15}R. \tag{6}$$

Therefore $a^* = 12/15 = 4/5$ and $b^* = 0$.

(e) The minimum mean square error in the estimate is

$$e^* = \text{Var}[V](1 - \rho_{V,R}^2) = 12(1 - 12/15) = 12/5 \tag{7}$$

2

Problem 9.2.6 Solution

The linear mean square estimator of X given Y is

$$\hat{X}_L(Y) = \left(\frac{E[XY] - \mu_X \mu_Y}{\text{Var}[Y]}\right)(Y - \mu_Y) + \mu_X. \tag{1}$$

To find the parameters of this estimator, we calculate

$$f_Y(y) = \int_0^y 6(y - x)\, dx = 6xy - 3x^2 \big|_0^y = 3y^2 \qquad (0 \le y \le 1) \tag{2}$$

$$f_X(x) = \int_x^1 6(y - x)\, dy = \begin{cases} 3(1 + -2x + x^2) & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The moments of X and Y are

$$E[Y] = \int_0^1 3y^3\, dy = 3/4 \qquad\qquad E[X] = \int_0^1 3x(1 - 2x + x^2)\, dx = 1/4 \tag{4}$$

$$E[Y^2] = \int_0^1 3y^4\, dy = 3/5 \qquad\qquad E[X^2] = \int_0^1 3x^2(1 + -2x + x^2)\, dx = 1/10 \tag{5}$$

The correlation between X and Y is

$$E[XY] = 6\int_0^1 \int_x^1 xy(y - x)\, dy\, dx = 1/5 \tag{6}$$

Putting these pieces together, the optimal linear estimate of X given Y is

$$\hat{X}_L(Y) = \left(\frac{1/5 - 3/16}{3/5 - (3/4)^2}\right)\left(Y - \frac{3}{4}\right) + \frac{1}{4} = \frac{Y}{3} \tag{7}$$

Problem 9.3.2 Solution

From the problem statement we know that R is an exponential random variable with expected value $1/\mu$. Therefore it has the following probability density function.

$$f_R(r) = \begin{cases} \mu e^{-\mu r} & r \ge 0 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

It is also known that, given $R = r$, the number of phone calls arriving at a telephone switch, N, is a Poisson $(\alpha = rT)$ random variable. So we can write the following conditional probability mass function of N given R.

$$P_{N|R}(n|r) = \begin{cases} \frac{(rT)^n e^{-rT}}{n!} & n = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

(a) The minimum mean square error estimate of N given R is the conditional expected value of N given $R = r$. This is given directly in the problem statement as r.

$$\hat{N}_M(r) = E[N|R = r] = rT \tag{3}$$

3

(b) The maximum a posteriori estimate of N given R is simply the value of n that will maximize $P_{N|R}(n|r)$. That is,

$$\hat{n}_{MAP(r)} = \arg\max_{n \geq 0} P_{N|R}(n|r) = \arg\max_{n \geq 0}(rT)^n e^{-rT}/n! \tag{4}$$

Usually, we set a derivative to zero to solve for the maximizing value. In this case, that technique doesn't work because n is discrete. Since e^{-rT} is a common factor in the maximization, we can define $g(n) = (rT)^n/n!$ so that $\hat{n}_{MAP} = \arg\max_n g(n)$. We observe that

$$g(n) = \frac{rT}{n}g(n-1) \tag{5}$$

this implies that for $n \leq rT$, $g(n) \geq g(n-1)$. Hence the maximizing value of n is the largest n such that $n \leq rT$. That is, $\hat{n}_{MAP} = \lfloor rT \rfloor$.

(c) The maximum likelihood estimate of N given R selects the value of n that maximizes $f_{R|N=n}(r)$, the conditional PDF of R given N. When dealing with situations in which we mix continuous and discrete random variables, its often helpful to start from first principles. In this case,

$$f_{R|N}(r|n) \, dr = P\left[r < R \leq r + dr | N = n\right] \tag{6}$$

$$= \frac{P\left[r < R \leq r + dr, N = n\right]}{P\left[N = n\right]} \tag{7}$$

$$= \frac{P\left[N = n | R = r\right] P\left[r < R \leq r + dr\right]}{P\left[N = n\right]} \tag{8}$$

In terms of PDFs and PMFs, we have

$$f_{R|N}(r|n) = \frac{P_{N|R}(n|r) f_R(r)}{P_N(n)} \tag{9}$$

To find the value of n that maximizes $f_{R|N}(r|n)$, we need to find the denominator $P_N(n)$.

$$P_N(n) = \int_{-\infty}^{\infty} P_{N|R}(n|r) f_R(r) \, dr \tag{10}$$

$$= \int_0^{\infty} \frac{(rT)^n e^{-rT}}{n!} \mu e^{-\mu r} \, dr \tag{11}$$

$$= \frac{\mu T^n}{n!(\mu + T)} \int_0^{\infty} r^n (\mu + T) e^{-(\mu+T)r} \, dr \tag{12}$$

$$= \frac{\mu T^n}{n!(\mu + T)} E\left[X^n\right] \tag{13}$$

where X is an exponential random variable with expected value $1/(\mu+T)$. There are several ways to derive the nth moment of an exponential random variable including integration by parts. In Example 6.5, the MGF is used to show that $E[X^n] = n!/(\mu + T)^n$. Hence, for $n \geq 0$,

$$P_N(n) = \frac{\mu T^n}{(\mu + T)^{n+1}} \tag{14}$$

4

Finally, the conditional PDF of R given N is

$$f_{R|N}(r|n) = \frac{P_{N|R}(n|r) f_R(r)}{P_N(n)} = \frac{\frac{(rT)^n e^{-rT}}{n!} \mu e^{-\mu r}}{\frac{\mu T^n}{(\mu+T)^{n+1}}} \tag{15}$$

$$= (\mu + T) \frac{[(\mu+T)r]^n e^{-(\mu+T)r}}{n!} \tag{16}$$

The ML estimate of N given R is

$$\hat{n}_{ML}(r) = \arg\max_{n \geq 0} f_{R|N}(r|n) = \arg\max_{n \geq 0} (\mu + T) \frac{[(\mu+T)r]^n e^{-(\mu+T)r}}{n!} \tag{17}$$

This maximization is exactly the same as in the previous part except rT is replaced by $(\mu + T)r$. The maximizing value of n is $\hat{n}_{ML} = \lfloor (\mu + T)r \rfloor$.

Problem 9.4.2 Solution

From the problem statement, we learn for vectors $\mathbf{X} = \begin{bmatrix} X_1 & X_2 & X_3 \end{bmatrix}'$ and $\mathbf{W} = \begin{bmatrix} W_1 & W_2 \end{bmatrix}'$ that

$$E[\mathbf{X}] = \mathbf{0}, \qquad \mathbf{R_X} = \begin{bmatrix} 1 & 3/4 & 1/2 \\ 3/4 & 1 & 3/4 \\ 1/2 & 3/4 & 1 \end{bmatrix}, \qquad E[\mathbf{W}] = \mathbf{0}, \qquad \mathbf{R_W} = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} \tag{1}$$

In addition,

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \mathbf{AX} + \mathbf{W} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \mathbf{X} + \mathbf{W}. \tag{2}$$

(a) Since $E[\mathbf{Y}] = \mathbf{A}E[\mathbf{X}] = \mathbf{0}$, we can apply Theorem 9.7(a) which states that the minimum mean square error estimate of X_1 is $\hat{X}_1(\mathbf{Y}) = \hat{\mathbf{a}}'\mathbf{Y}$ where $\hat{\mathbf{a}} = \mathbf{R_Y}^{-1}\mathbf{R}_{\mathbf{Y}X_1}$. First we find $\mathbf{R_Y}$.

$$\mathbf{R_Y} = E[\mathbf{YY}'] = E[(\mathbf{AX} + \mathbf{W})(\mathbf{AX} + \mathbf{W})'] \tag{3}$$

$$= E[(\mathbf{AX} + \mathbf{W})(\mathbf{X}'\mathbf{A}' + \mathbf{W}')] \tag{4}$$

$$= E[\mathbf{AXX}'\mathbf{A}'] + E[\mathbf{WX}'\mathbf{A}] + E[\mathbf{AXW}'] + E[\mathbf{WW}'] \tag{5}$$

Since $\mathbf{X}$ and $\mathbf{W}$ are independent, $E[\mathbf{WX}'] = \mathbf{0}$ and $E[\mathbf{XW}'] = \mathbf{0}$. This implies

$$\mathbf{R_Y} = \mathbf{A}E[\mathbf{XX}']\mathbf{A}' + E[\mathbf{WW}'] \tag{6}$$

$$= \mathbf{A}\mathbf{R_X}\mathbf{A}' + \mathbf{R_W} \tag{7}$$

$$= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3/4 & 1/2 \\ 3/4 & 1 & 3/4 \\ 1/2 & 3/4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} = \begin{bmatrix} 3.6 & 3 \\ 3 & 3.6 \end{bmatrix}. \tag{8}$$

Once again, independence of $\mathbf{W}$ and X_1 yields

$$\mathbf{R}_{\mathbf{Y}X_1} = E[\mathbf{Y}X_1] = E[(\mathbf{AX} + \mathbf{W})X_1] = \mathbf{A}E[\mathbf{X}X_1]. \tag{9}$$

This implies

$$\mathbf{R}_{\mathbf{Y}X_1} = \mathbf{A} \begin{bmatrix} E\left[X_1^2\right] \\ E\left[X_2 X_1\right] \\ E\left[X_3 X_1\right] \end{bmatrix} = \mathbf{A} \begin{bmatrix} R_\mathbf{X}(1,1) \\ R_\mathbf{X}(2,1) \\ R_\mathbf{X}(3,1) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3/4 \\ 1/2 \end{bmatrix} = \begin{bmatrix} 7/4 \\ 5/4 \end{bmatrix}. \tag{10}$$

Putting these facts together, we find that

$$\hat{\mathbf{a}} = \mathbf{R}_{\mathbf{Y}}^{-1} \mathbf{R}_{\mathbf{Y}X_1} = \begin{bmatrix} 10/11 & -25/33 \\ -25/33 & 10/11 \end{bmatrix} \begin{bmatrix} 7/4 \\ 5/4 \end{bmatrix} = \frac{1}{132} \begin{bmatrix} 85 \\ -25 \end{bmatrix}. \tag{11}$$

Thus the linear MMSE estimator of X_1 given $\mathbf{Y}$ is

$$\hat{X}_1(\mathbf{Y}) = \hat{\mathbf{a}}'\mathbf{Y} = \frac{85}{132}Y_1 - \frac{25}{132}Y_2 = 0.6439Y_1 - 0.1894Y_2. \tag{12}$$

(b) By Theorem 9.7(c), the mean squared error of the optimal estimator is

$$e_L^* = \mathrm{Var}[X_1] - \hat{\mathbf{a}}'\mathbf{R}_{\mathbf{Y}X_1} \tag{13}$$

$$= R_\mathbf{X}(1,1) - \mathbf{R}_{\mathbf{Y}X_1}' \mathbf{R}_{\mathbf{Y}}^{-1} \mathbf{R}_{\mathbf{Y}X_1} \tag{14}$$

$$= 1 - \begin{bmatrix} 7/4 & 5/4 \end{bmatrix} \begin{bmatrix} 10/11 & -25/33 \\ -25/33 & 10/11 \end{bmatrix} \begin{bmatrix} 7/4 \\ 5/4 \end{bmatrix} = \frac{29}{264} = 0.1098. \tag{15}$$

In Problem 9.4.1, we solved essentialy the same problem but the observations $\mathbf{Y}$ were not subjected to the additive noise $\mathbf{W}$. In comparing the estimators, we see that the additive noise perturbs the estimator somewhat but not dramatically because the correaltion structure of $\mathbf{X}$ and the mapping $\mathbf{A}$ from $\mathbf{X}$ to $\mathbf{Y}$ remains unchanged. On the other hand, in the noiseless case, the resulting mean square error was about half as much, $3/52 = 0.0577$ versus 0.1098.

(c) We can estimate random variable X_1 based on the observation of random variable Y_1 using Theorem 9.4. Note that Theorem 9.4 is a special case of Theorem 9.8 in which the observation is a random vector. In any case, from Theorem 9.4, the optimum linear estimate is $\hat{X}_1(Y_1) = a^* Y_1 + b^*$ where

$$a^* = \frac{\mathrm{Cov}\left[X_1, Y_1\right]}{\mathrm{Var}[Y_1]}, \qquad b^* = \mu_{X_1} - a^* \mu_{Y_1}. \tag{16}$$

Since $E[X_i] = \mu_{X_i} = 0$ and $Y_1 = X_1 + X_2 + W_1$, we see that

$$\mu_{Y_1} = E\left[Y_1\right] = E\left[X_1\right] + E\left[X_2\right] + E\left[W_1\right] = 0. \tag{17}$$

These facts, along with independence of X_1 and W_1, imply

$$\mathrm{Cov}\left[X_1, Y_1\right] = E\left[X_1 Y_1\right] = E\left[X_1(X_1 + X_2 + W_1)\right] \tag{18}$$

$$= R_\mathbf{X}(1,1) + R_\mathbf{X}(1,2) = 7/4 \tag{19}$$

In addition, using $\mathbf{R}_\mathbf{Y}$ from part (a), we see that

$$\mathrm{Var}[Y_1] = E\left[Y_1^2\right] = R_\mathbf{Y}(1,1) = 3.6. \tag{20}$$

Thus

$$a^* = \frac{\text{Cov}\left[X_1, Y_1\right]}{\text{Var}[Y_1]} = \frac{7/4}{3.6} = \frac{35}{72}, \qquad b^* = \mu_{X_1} - a^*\mu_{Y_1} = 0. \tag{21}$$

Thus the optimum linear estimate of X_1 given Y_1 is

$$\hat{X}_1(Y_1) = \frac{35}{72}Y_1. \tag{22}$$

From Theorem 9.4(b), the mean square error of this estimator is

$$e_L^* = \sigma_{X_1}^2(1 - \rho_{X_1,Y_1}^2) \tag{23}$$

Since X_1 and Y_1 have zero expected value, $\sigma_{X_1}^2 = R_{\mathbf{X}}(1,1) = 1$ and $\sigma_{Y_1}^2 = R_{\mathbf{Y}}(1,1) = 3.6$. Also, since $\text{Cov}[X_1, Y_1] = 7/4$, we see that

$$\rho_{X_1,Y_1} = \frac{\text{Cov}\left[X_1, Y_1\right]}{\sigma_{X_1}\sigma_{Y_1}} = \frac{7/4}{\sqrt{3.6}} = \frac{\sqrt{490}}{24}. \tag{24}$$

Thus $e_L^* = 1 - (490/24^2) = 0.1493$. As we would expect, the estimate of X_1 based on just Y_1 has larger mean square error than the estimate based on both Y_1 and Y_2.

Problem 9.4.6 Solution

For this problem, let $\mathbf{Y} = \begin{bmatrix} X_1 & X_2 & \cdots & n-1 \end{bmatrix}'$ and let $X = X_n$. Since $E[\mathbf{Y}] = 0$ and $E[X] = 0$, Theorem 9.7(a) tells us that the minimum mean square linear estimate of X given $\mathbf{Y}$ is $\hat{X}_n(\mathbf{Y}) = \hat{\mathbf{a}}'\mathbf{Y}$, where $\hat{\mathbf{a}} = \mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}_{\mathbf{Y}X}$. This implies that $\hat{\mathbf{a}}$ is the solution to

$$\mathbf{R}_{\mathbf{Y}}\hat{\mathbf{a}} = \mathbf{R}_{\mathbf{Y}X}. \tag{1}$$

Note that

$$\mathbf{R}_{\mathbf{Y}} = E\left[\mathbf{YY}'\right] = \begin{bmatrix} 1 & c & \cdots & c^{n-2} \\ c & c^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & c \\ c^{n-2} & \cdots & c & 1 \end{bmatrix}, \quad \mathbf{R}_{\mathbf{Y}X} = E\left[\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{n-1} \end{bmatrix} X_n\right] = \begin{bmatrix} c^{n-1} \\ c^{n-2} \\ \vdots \\ c \end{bmatrix}. \tag{2}$$

We see that the last column of $c\mathbf{R}_{\mathbf{Y}}$ equals $\mathbf{R}_{\mathbf{Y}X}$. Equivalently, if $\hat{\mathbf{a}} = \begin{bmatrix} 0 & \cdots & 0 & c \end{bmatrix}'$, then $\mathbf{R}_{\mathbf{Y}}\hat{\mathbf{a}} = \mathbf{R}_{\mathbf{Y}X}$. It follows that the optimal linear estimator of X_n given $\mathbf{Y}$ is

$$\hat{X}_n(\mathbf{Y}) = \hat{\mathbf{a}}'\mathbf{Y} = cX_{n-1}, \tag{3}$$

which completes the proof of the claim.

The mean square error of this estimate is

$$e_L^* = E\left[(X_n - cX_{n-1})^2\right] \tag{4}$$
$$= R_{\mathbf{X}}(n,n) - cR_{\mathbf{X}}(n, n-1) - cR_{\mathbf{X}}(n-1, n) + c^2 R_{\mathbf{X}}(n-1, n-1) \tag{5}$$
$$= 1 - 2c^2 + c^2 = 1 - c^2 \tag{6}$$

When c is close to 1, X_{n-1} and X_n are highly correlated and the estimation error will be small.

Comment: We will see in Chapter 11 that correlation matrices with this structure arise frequently in the study of wide sense stationary random sequences. In fact, if you read ahead, you will find that the claim we just proved is the essentially the same as that made in Theorem 11.10.

Problem 9.4.7 Solution

(a) In this case, we use the observation $\mathbf{Y}$ to estimate each X_i. Since $E[X_i] = 0$,

$$E\left[\mathbf{Y}\right] = \sum_{j=1}^{k} E\left[X_j\right] \sqrt{p_j}\mathbf{S}_j + E\left[\mathbf{N}\right] = \mathbf{0}. \tag{1}$$

Thus, Theorem 9.7(a) tells us that the MMSE linear estimate of X_i is $\hat{X}_i(\mathbf{Y}) = \hat{\mathbf{a}}'\mathbf{Y}$ where $\hat{\mathbf{a}} = \mathbf{R_Y^{-1}R}_{\mathbf{Y}X_i}$. First we note that

$$\mathbf{R}_{\mathbf{Y}X_i} = E\left[\mathbf{Y}X_i\right] = E\left[\left(\sum_{j=1}^{k} X_j\sqrt{p_j}\mathbf{S}_j + \mathbf{N}\right)X_i\right] \tag{2}$$

Since $\mathbf{N}$ and X_i are independent, $E[\mathbf{N}X_i] = E[\mathbf{N}]E[X_i] = \mathbf{0}$. Because X_i and X_j are independent for $i \neq j$, $E[X_iX_j] = E[X_i]E[X_j] = 0$ for $i \neq j$. In addition, $E[X_i^2] = 1$, and it follows that

$$\mathbf{R}_{\mathbf{Y}X_i} = \sum_{j=1}^{k} E\left[X_jX_i\right] \sqrt{p_j}\mathbf{S}_j + E\left[\mathbf{N}X_i\right] = \sqrt{p_i}\mathbf{S}_i. \tag{3}$$

For the same reasons,

$$\mathbf{R_Y} = E\left[\mathbf{YY}'\right] = E\left[\left(\sum_{j=1}^{k} \sqrt{p_j}X_j\mathbf{S}_j + \mathbf{N}\right)\left(\sum_{l=1}^{k} \sqrt{p_l}X_l\mathbf{S}'_l + \mathbf{N}'\right)\right] \tag{4}$$

$$= \sum_{j=1}^{k}\sum_{l=1}^{k} \sqrt{p_jp_l}E\left[X_jX_l\right]\mathbf{S}_j\mathbf{S}'_l$$

$$+ \sum_{j=1}^{k} \sqrt{p_j}\underbrace{E\left[X_j\mathbf{N}\right]}_{=0}\mathbf{S}_j + \sum_{l=1}^{k} \sqrt{p_l}\underbrace{E\left[X_l\mathbf{N}'\right]}_{=0}\mathbf{S}'_l + E\left[\mathbf{NN}'\right] \tag{5}$$

$$= \sum_{j=1}^{k} p_j\mathbf{S}_j\mathbf{S}'_j + \sigma^2\mathbf{I} \tag{6}$$

Now we use a linear algebra identity. For a matrix $\mathbf{S}$ with columns $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_k$, and a diagonal matrix $\mathbf{P} = \text{diag}[p_1, p_2, \ldots, p_k]$,

$$\sum_{j=1}^{k} p_j\mathbf{S}_j\mathbf{S}'_j = \mathbf{SPS}'. \tag{7}$$

Although this identity may be unfamiliar, it is handy in manipulating correlation matrices. (Also, if this is unfamiliar, you may wish to work out an example with $k = 2$ vectors of length 2 or 3.) Thus,

$$\mathbf{R_Y} = \mathbf{SPS}' + \sigma^2\mathbf{I}, \tag{8}$$

8

and

$$\hat{\mathbf{a}} = \mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}_{\mathbf{Y}X_i} = \left(\mathbf{SPS'} + \sigma^2\mathbf{I}\right)^{-1}\sqrt{p_i}\mathbf{S}_i. \tag{9}$$

Recall that if $\mathbf{C}$ is symmetric, then $\mathbf{C}^{-1}$ is also symmetric. This implies the MMSE estimate of X_i given $\mathbf{Y}$ is

$$\hat{X}_i(\mathbf{Y}) = \hat{\mathbf{a}}'\mathbf{Y} = \sqrt{p_i}\mathbf{S}'_i\left(\mathbf{SPS'} + \sigma^2\mathbf{I}\right)^{-1}\mathbf{Y} \tag{10}$$

(b) We observe that $\mathbf{V} = (\mathbf{SPS'} + \sigma^2\mathbf{I})^{-1}\mathbf{Y}$ is a vector that does not depend on which bit X_i that we want to estimate. Since $\hat{X}_i = \sqrt{p_i}\mathbf{S}'_i\mathbf{V}$, we can form the vector of estimates

$$\hat{\mathbf{X}} = \begin{bmatrix} \hat{X}_1 \\ \vdots \\ \hat{X}_k \end{bmatrix} = \begin{bmatrix} \sqrt{p_1}\mathbf{S}'_1\mathbf{V} \\ \vdots \\ \sqrt{p_k}\mathbf{S}'_k\mathbf{V} \end{bmatrix} = \begin{bmatrix} \sqrt{p_1} & & \\ & \ddots & \\ & & \sqrt{p_k} \end{bmatrix} \begin{bmatrix} \mathbf{S}'_1 \\ \vdots \\ \mathbf{S}'_k \end{bmatrix} \mathbf{V} \tag{11}$$

$$= \mathbf{P}^{1/2}\mathbf{S}'\mathbf{V} \tag{12}$$

$$= \mathbf{P}^{1/2}\mathbf{S}'\left(\mathbf{SPS'} + \sigma^2\mathbf{I}\right)^{-1}\mathbf{Y} \tag{13}$$

Problem 9.5.3 Solution

The solution to this problem is almost the same as the solution to Example 9.10, except perhaps the MATLAB code is somewhat simpler. As in the example, let $\mathbf{W}^{(n)}$, $\mathbf{X}^{(n)}$, and $\mathbf{Y}^{(n)}$ denote the vectors, consisting of the first n components of $\mathbf{W}$, $\mathbf{X}$, and $\mathbf{Y}$. Just as in Examples 9.8 and 9.10, independence of $\mathbf{X}^{(n)}$ and $\mathbf{W}^{(n)}$ implies that the correlation matrix of $\mathbf{Y}^{(n)}$ is

$$\mathbf{R}_{\mathbf{Y}^{(n)}} = E\left[(\mathbf{X}^{(n)} + \mathbf{W}^{(n)})(\mathbf{X}^{(n)} + \mathbf{W}^{(n)})'\right] = \mathbf{R}_{\mathbf{X}^{(n)}} + \mathbf{R}_{\mathbf{W}^{(n)}} \tag{1}$$

Note that $\mathbf{R}_{\mathbf{X}^{(n)}}$ and $\mathbf{R}_{\mathbf{W}^{(n)}}$ are the $n \times n$ upper-left submatrices of $\mathbf{R}_{\mathbf{X}}$ and $\mathbf{R}_{\mathbf{W}}$. In addition,

$$\mathbf{R}_{\mathbf{Y}^{(n)}X} = E\left[\begin{bmatrix} X_1 + W_1 \\ \vdots \\ X_n + W_n \end{bmatrix} X_1\right] = \begin{bmatrix} r_0 \\ \vdots \\ r_{n-1} \end{bmatrix}. \tag{2}$$

Compared to the solution of Example 9.10, the only difference in the solution is in the reversal of the vector $\mathbf{R}_{\mathbf{Y}^{(n)}X}$. The optimal filter based on the first n observations is $\hat{\mathbf{a}}^{(n)} = \mathbf{R}_{\mathbf{Y}^{(n)}}^{-1}\mathbf{R}_{\mathbf{Y}^{(n)}X}$, and the mean square error is

$$e_L^* = \mathrm{Var}[X_1] - (\hat{\mathbf{a}}^{(n)})'\mathbf{R}_{\mathbf{Y}^{(n)}X}. \tag{3}$$

```
function e=mse953(r)
N=length(r);
e=[];
for n=1:N,
    RYX=r(1:n)';
    RY=toeplitz(r(1:n))+0.1*eye(n);
    a=RY\RYX;
    en=r(1)-(a')*RYX;
    e=[e;en];
end
plot(1:N,e);
```

The program `mse953.m` simply calculates the mean square error e_L^*. The input is the vector `r` corresponding to the vector $\begin{bmatrix} r_0 & \cdots & r_{20} \end{bmatrix}$, which holds the first row of the Toeplitz correlation matrix $\mathbf{R_X}$. Note that $\mathbf{R}_{\mathbf{X}^{(n)}}$ is the Toeplitz matrix whose first row is the first n elements of `r`.

To plot the mean square error as a function of the number of observations, n, we generate the vector `r` and then run `mse953(r)`. For the requested cases (a) and (b), the necessary MATLAB commands and corresponding mean square estimation error output as a function of n are shown here:
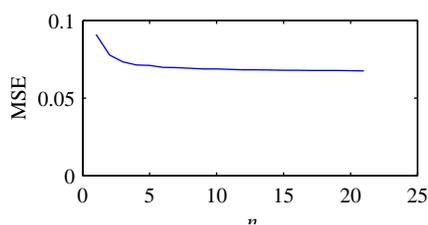

```
ra=sinc(0.1*pi*(0:20));
mse953(ra)
```
(a)
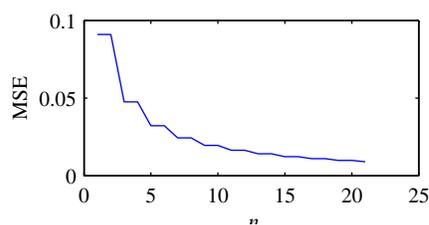
```
rb=cos(0.5*pi*(0:20));
mse953(rb)
```
(b)

In comparing the results of cases (a) and (b), we see that the mean square estimation error depends strongly on the correlation structure given by $r_{|i-j|}$. For case (a), Y_1 is a noisy observation of X_1 and is highly correlated with X_1. The MSE at $n = 1$ is just the variance of W_1. Additional samples of Y_n mostly help to average the additive noise. Also, samples X_n for $n \geq 10$ have very little correlation with X_1. Thus for $n \geq 10$, the samples of Y_n result in almost no improvement in the estimate of X_1.

In case (b), $Y_1 = X_1 + W_1$, just as in case (a), is simply a noisy copy of X_1 and the estimation error is due to the variance of W_1. On the other hand, for case (b), X_5, X_9, X_{13} and X_{17} and X_{21} are completely correlated with X_1. Other samples also have significant correlation with X_1. As a result, the MSE continues to go down with increasing n.