

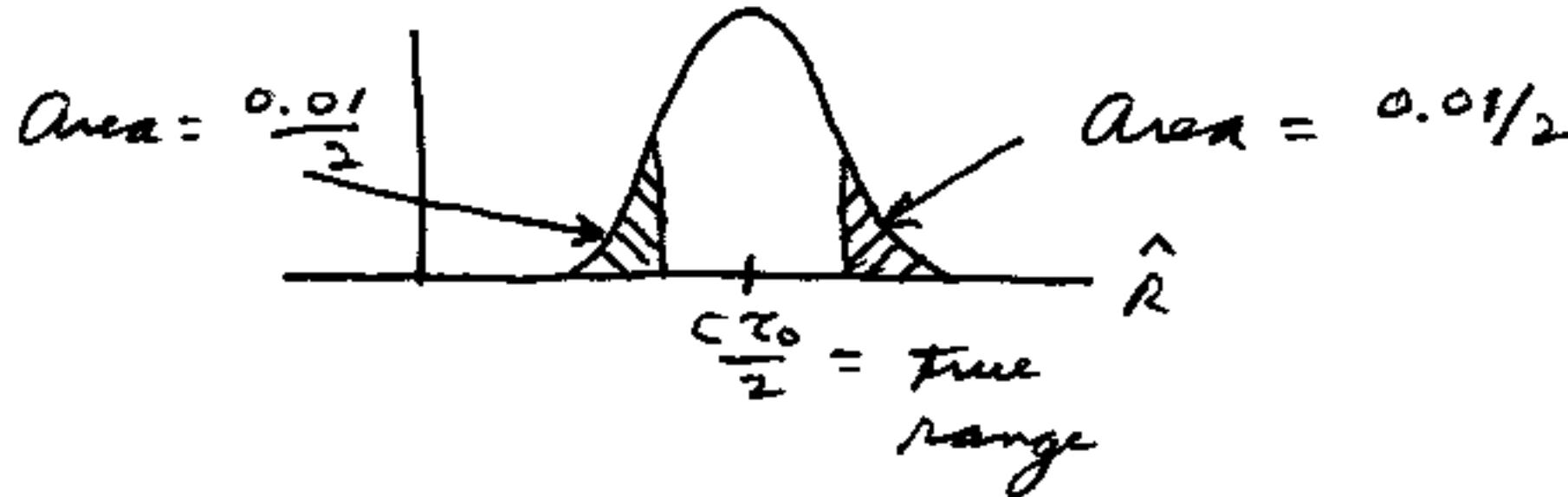
PROBLEM SOLUTIONS

FUNDAMENTALS OF STATISTICAL  
SIGNAL PROCESSING: ESTIMATION  
THEORY

BY STEVEN KAY

## Chapter 1

- 1) Since  $R = Cx_0/2$ , we use  $\hat{R} = \hat{C}\hat{x}_0/2$ .  
 The PDF is from  $\hat{x}_0 \sim N(x_0, \sigma_{\hat{x}_0}^2)$ ,  
 $\hat{R} \sim N(Cx_0/2, \frac{C^2}{4} \sigma_{\hat{x}_0}^2)$



To be within 100 m we must have

$$\Pr \{ |\hat{R} - Cx_0/2| < 100 \} = 0.99 \\ \Rightarrow \Pr \{ \underbrace{\left| \frac{\hat{R} - Cx_0/2}{\frac{C}{2} \sigma_{\hat{x}_0}} \right|}_{N(0,1)} < \frac{100}{\frac{C}{2} \sigma_{\hat{x}_0}} \} = 0.99$$

$$\Rightarrow \frac{100}{\frac{C}{2} \sigma_{\hat{x}_0}} = 2.58 \text{ or } \sigma_{\hat{x}_0} = 2.6 \times 10^{-7} \text{ sec} \\ = 0.26 \text{ usec}$$

- 2) No, in fact  $\sigma$  could have been any value.  
 If  $\sigma$  were indeed 100, then  
 $p(x; \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-100)^2}$

and the probability of  $x$  being in the interval  $(-97, 103)$  or  $\mu \pm 3\sigma$  is 0.999.  
 Hence, his assertion is likely to be correct. However, we cannot be

certain, since if  $\theta = 99$ , then the probability of  $x$  being in the observed interval (for a single experiment) is

$$\int_{-97}^{103} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-99)^2} dx =$$

$$\int_{-2}^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du = 0.977.$$

Thus,  $\theta = 99$  is also highly likely.

3)  $x = \theta + w$

$$p(x; \theta) = p_w(x - \theta)$$

For  $\omega$  a random variable independent of  $w$ ,

$$p(x|\theta) = \frac{p_{x\theta}(x, \theta)}{p(\theta)} = \frac{p_{w\theta}(x-\theta, \theta)}{p(\theta)}$$

$$= \frac{p_w(x-\theta)p(\theta)}{p(\theta)} = p_w(x-\theta)$$

which is the same as before. If  $w$  and  $\theta$  are not independent, then

$$p(x|\theta) = \frac{p_{w|\theta}(x-\theta)p(\theta)}{p(\theta)} = p_{w|\theta}(x-\theta)$$

which will be different than  $p_w(x-\theta)$ .

In general,  $p(x; \theta) \neq p(x|\theta)$ .

- 4) As shown in text  $E(\hat{A}) = A$ . Also,

$$E(\check{A}) = \frac{1}{N+2} (2A + (N-2)A + 2A) = A$$

Also, we know that  $\text{var}(\hat{A}) = \sigma^2/N = 1/N$  and

$$\begin{aligned}\text{var}(\check{A}) &= \frac{1}{(N+2)^2} \left[ 4\sigma^2 + \sum_{n=1}^{N-2} \sigma^2 + 4\sigma^2 \right] \\ &= \frac{N+6}{(N+2)^2} \sigma^2 = \frac{N+6}{(N+2)^2}\end{aligned}$$

$$\begin{aligned}\text{var}(\check{A}) - \text{var}(\hat{A}) &= \frac{N+6}{(N+2)^2} - \frac{1}{N} \\ &= \frac{N(N+6) - (N+2)^2}{N(N+2)^2} \\ &= \frac{2N-4}{N(N+2)^2} > 0 \quad \text{for } N > 2\end{aligned}$$

Hence, both estimators yield the correct value on the average but  $\hat{A}$  has less variance. Conclusion is the same for any value of  $A$ .

- 5)  $\hat{A}$  is not an estimator since to implement it requires knowledge of  $A$  (to determine the SWR).

## Chapter 2

1)  $E(\hat{\sigma}^2) = E\left(\frac{1}{N} \sum_{n=0}^{N-1} x^2[n]\right) = \frac{1}{N} \sum_{n=0}^{N-1} E(x^2[n])$   
 $= \sigma^2$  for all  $\sigma^2 > 0$  (allowable values)  
 $\Rightarrow$  unbiased

$$\begin{aligned}\text{var}(\hat{\sigma}^2) &= \frac{1}{N^2} \text{var}\left(\sum_n x^2[n]\right) \\ &= \frac{1}{N^2} N \text{var}(x^2[n]) \quad (x[n]'s \text{ are IID} \\ &\qquad \Rightarrow x^2[n] \text{ are IID}) \\ &= \frac{1}{N} \text{var}(x^2[n])\end{aligned}$$

$$\begin{aligned}\text{var}(x^2[n]) &= E(x^4[n]) - E(x^2[n])^2 \\ &= 3\sigma^4 - \sigma^4 = 2\sigma^4\end{aligned}$$

$\Rightarrow \text{var}(\hat{\sigma}^2) = 2\sigma^4/N \rightarrow 0$  as  $N \rightarrow \infty$   
Hence, the PDF of  $\hat{\sigma}^2$  collapses about the true value as  $N \rightarrow \infty$ .

2) Let  $\hat{\theta} = 2 \frac{1}{N} \sum_{n=0}^{N-1} x[n]$  since  $E(x[n]) = \theta/2$

$$E(\hat{\theta}) = \frac{2}{N} \sum_{n=0}^{N-1} \theta/2 = \theta$$

3)  $\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$   $\hat{A}$  is Gaussian since it is a linear function of independent Gaussian random variables. The mean was found to be A and the variance is

$$\begin{aligned}\text{var}(\hat{A}) &= \text{var}\left(\frac{1}{N} \sum_{n=0}^{N-1} x[n]\right) \\ &= \frac{1}{N^2} \text{var}\left(\sum_{n=0}^{N-1} x[n]\right) \\ &= \frac{1}{N^2} N \text{var}(x[n])\end{aligned}$$

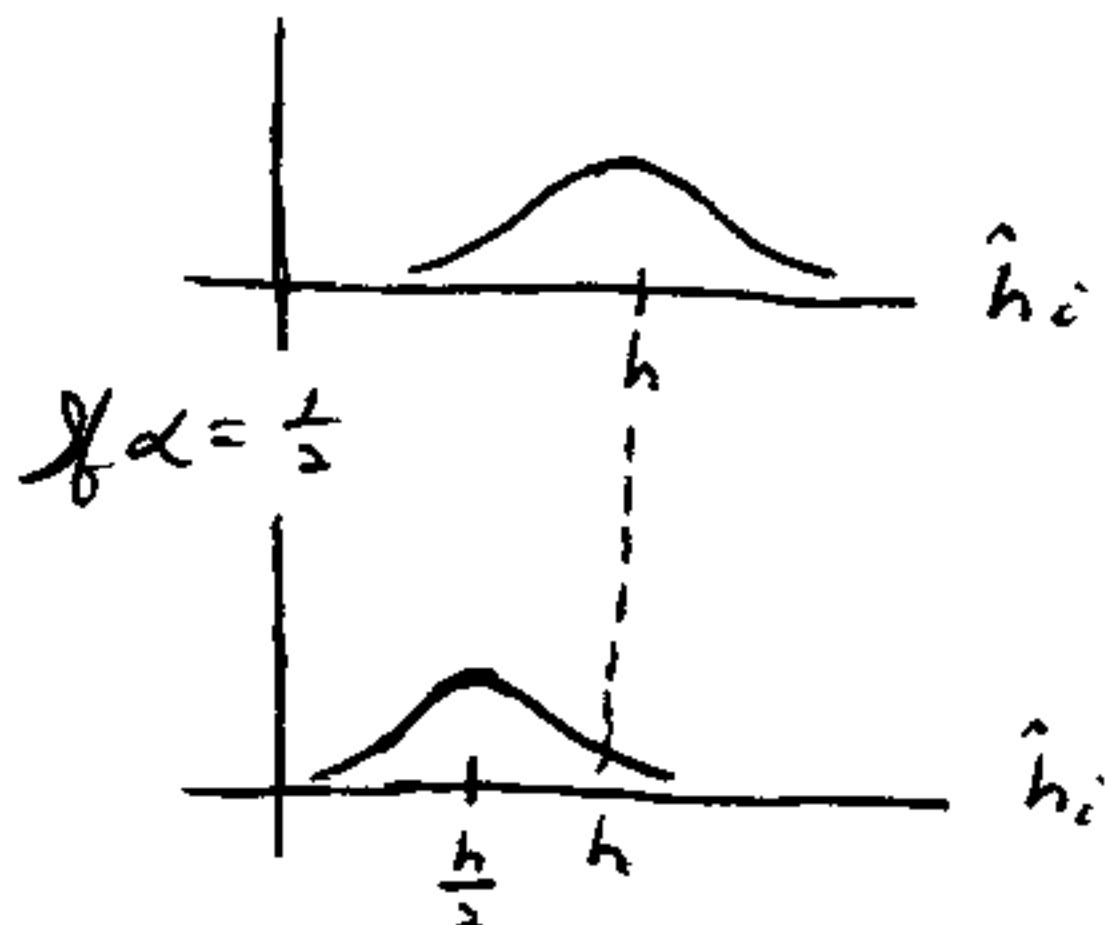
Since the  $x[n]$ 's are IID and thus uncorrelated

$$\Rightarrow \text{var}(\hat{A}) = \sigma^2/N.$$

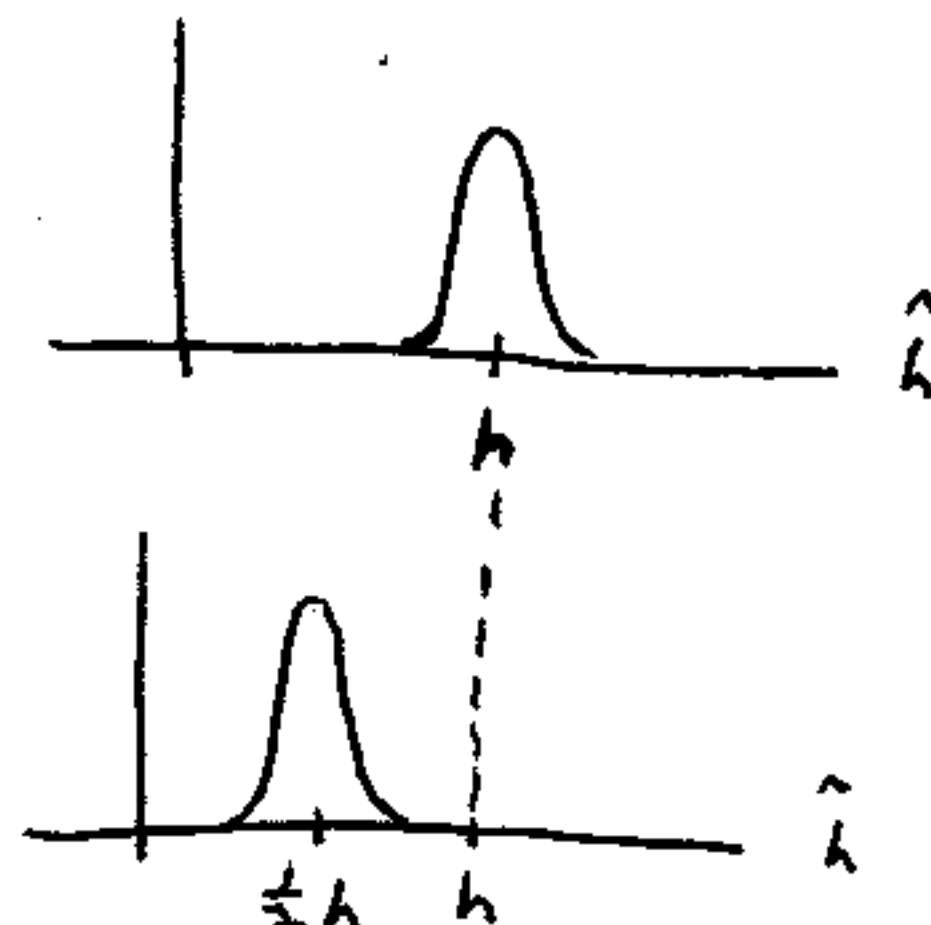
$$4) \quad \hat{h} = \frac{1}{10} \sum_{i=1}^{10} h_i \quad E(\hat{h}) = \frac{1}{10} \sum_{i=1}^{10} E(h_i) \\ = \alpha h$$

$$\text{var}(\hat{h}) = \text{var}(h_i)/10 = 1/10$$

If  $\alpha = 1$ , we have



Before averaging



After averaging

In second case ( $\alpha = \frac{1}{2}$ ), averaging causes the PDF to be more heavily concentrated about the wrong value of  $h$ . The probability

of  $\hat{h}$  being close to  $h$  actually decreases due to averaging. For  $\alpha = 1$ , averaging, of course, is beneficial.

$$5) \quad X[n] \sim N(0, \sigma^2) \quad \text{or} \quad \frac{X[n]}{\sigma} \sim N(0, 1)$$

$$\Rightarrow \left(\frac{X[n]}{\sigma}\right)^2 \sim \chi_1^2 \quad \text{and}$$

$$y = \left(\frac{X[0]}{\sigma}\right)^2 + \left(\frac{X[1]}{\sigma}\right)^2 \sim \chi_2^2$$

$$\text{or } p(y) = \begin{cases} \frac{1}{2} e^{-y/2} & y > 0 \\ 0 & y \leq 0 \end{cases}$$

Transforming, we have  $\hat{\sigma}^2 = \frac{\sigma^2}{2} y$   
so that

$$\begin{aligned} p(\hat{\sigma}^2) &= \frac{p_y(y(\hat{\sigma}^2))}{|d\hat{\sigma}^2/dy|} \\ &= \frac{\frac{1}{2} e^{-\frac{1}{2}(2\hat{\sigma}^2/\sigma^2)}}{\sigma^2/2} \quad \hat{\sigma}^2 > 0 \end{aligned}$$

$$= \begin{cases} \frac{1}{\sigma^2} e^{-\hat{\sigma}^2/\sigma^2} & \hat{\sigma}^2 > 0 \\ 0 & \hat{\sigma}^2 \leq 0 \end{cases}$$



Clearly, not symmetric.

$$\begin{aligned}
 \text{But } E(\hat{\sigma}^2) &= \int_0^\infty \hat{\sigma}^2 \frac{1}{\sigma^2} e^{-\hat{\sigma}^2/\sigma^2} d\hat{\sigma}^2 \\
 &= \sigma^2 \int_0^\infty u e^{-u} du \\
 &= \sigma^2 [-ue^{-u} - e^{-u}] \Big|_0^\infty \\
 &= \sigma^2
 \end{aligned}$$

$\hat{\sigma}^2$  is unbiased but PDF is not symmetric about  $\sigma^2$ .

b)  $E(\hat{A}) = \sum_{n=0}^{N-1} a_n A = A \Rightarrow \sum_{n=0}^{N-1} a_n = 1$

$$\text{var}(\hat{A}) = \sum_{n=0}^{N-1} a_n^2 \text{var}(x(n)) = \sum_{n=0}^{N-1} a_n^2 \sigma^2$$

$$\text{Let } F = \sigma^2 \sum_{n=0}^{N-1} a_n^2 + \lambda \left( \sum_{n=0}^{N-1} a_n - 1 \right)$$

$$\frac{\partial F}{\partial a_i} = 2\sigma^2 a_i + \lambda = 0 \quad i = 0, 1, \dots, N-1$$

$$\Rightarrow a_i = -\lambda/2\sigma^2 \quad \text{for all } i$$

Thus, the  $a_i$ 's must be equal. But  $\sum_{n=0}^{N-1} a_n = 1 \Rightarrow N a_i = 1$  or  $a_i = 1/N$  or

$\hat{A}$  is just the sample mean estimator or as shown in Example 3.3, the MVU estimator.

$$7) \quad \frac{\hat{\theta} - \theta}{\sqrt{\text{var}(\hat{\theta})}} \sim N(0, 1) \quad \frac{\hat{\theta} - \theta}{\sqrt{\text{var}(\theta)}} \sim N(0, 1)$$

$$P_r \left\{ |\hat{\theta} - \theta| > \epsilon \right\} = P_r \left\{ \left| \frac{\hat{\theta} - \theta}{\sqrt{\text{var}(\hat{\theta})}} \right| > \frac{\epsilon}{\sqrt{\text{var}(\hat{\theta})}} \right\}$$

$$\text{Let } \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

= cumulative distribution function for  $N(0, 1)$

$$\Rightarrow P_r \left\{ |\hat{\theta} - \theta| > \epsilon \right\} = 2 \Phi \left( \frac{-\epsilon}{\sqrt{\text{var}(\hat{\theta})}} \right)$$

If  $\text{var}(\hat{\theta}) < \text{var}(\theta)$

$$\Rightarrow \Phi \left( \frac{-\epsilon}{\sqrt{\text{var}(\hat{\theta})}} \right) < \Phi \left( \frac{-\epsilon}{\sqrt{\text{var}(\theta)}} \right)$$

$$\text{or } P_r \left\{ |\hat{\theta} - \theta| > \epsilon \right\} < P_r \left\{ |\theta - \theta| > \epsilon \right\}$$

Q) From Prob. 2.3  $\hat{A} \sim N(A, \sigma^2/N)$

$$P_r \left\{ |A - A| > \epsilon \right\} = P_r \left\{ \left| \frac{\hat{A} - A}{\sqrt{\sigma^2/N}} \right| > \frac{\epsilon}{\sqrt{\sigma^2/N}} \right\}$$

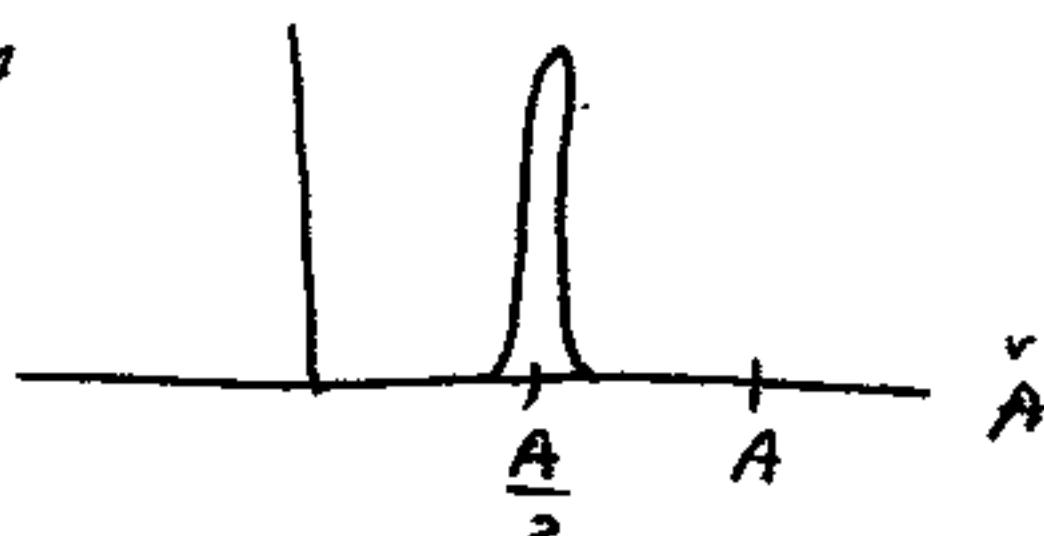
$$= 2 \Phi \left( \frac{-\epsilon}{\sqrt{\sigma^2/N}} \right) \rightarrow 0$$

Since  $\frac{-\epsilon}{\sqrt{\sigma^2/N}} \rightarrow -\infty$  as  $N \rightarrow \infty$

If  $\bar{A} = \frac{1}{2N} \sum_{n=0}^{N-1} x_n$  is used,

$$\hat{A} \sim N(A/2, \sigma^2/4N)$$

As  $N \rightarrow \infty$ , the variance  $\rightarrow 0$  and the PDF approaches



Thus,  $\lim_{N \rightarrow \infty} \Pr\{|\hat{A} - A| > \epsilon\} = 0$  for any  $\epsilon > 0$

$\hat{A}$  is consistent while  $\tilde{A}$  is inconsistent.

9)  $\hat{\theta} = (\hat{A})^2$  where  $\hat{A} \sim N(A, \sigma^2/N)$

$$\begin{aligned} E(\hat{\theta}) &= E(\hat{A}^2) = \text{var}(\hat{A}) + E(\hat{A})^2 \\ &= \sigma^2/N + A^2 = \theta + \sigma^2/N \neq \theta \end{aligned}$$

$\hat{\theta}$  is biased but as  $N \rightarrow \infty$ , it is unbiased.  $\hat{\theta}$  is said to be asymptotically unbiased (for large data records).

10) Clearly,  $E(\hat{A}) = A$ . To find  $E(\hat{\theta}^2)$

$$E(\hat{\theta}^2) = \frac{1}{N-1} \sum_{n=0}^{N-1} E[(x[n] - \hat{A})^2]$$

$$\text{But } E[(x[n] - \hat{A})^2] = E\left[\left(x[n] - \frac{1}{N} \sum_{m=0}^{N-1} x[m]\right)^2\right]$$

$$= E\left[\left(x[n]\left(1 - \frac{1}{N}\right) - \frac{1}{N} \sum_{\substack{m=0 \\ m \neq n}}^{N-1} x[m]\right)^2\right]$$

$$= \left(\frac{N-1}{N}\right)^2 E(x^2[n]) - 2 \frac{(N-1)}{N^2} E\left[x[n] \sum_{\substack{m=0 \\ m \neq n}}^{N-1} x[m]\right]$$

$$+ \frac{1}{N^2} E \left[ \left( \sum_{\substack{m=0 \\ m \neq n}}^{N-1} x[m] \right)^2 \right]$$

$$= \left( \frac{N-1}{N} \right)^2 (\sigma^2 + A^2) - \frac{2(N-1)}{N^2} E(x[n]) E \left( \sum_{\substack{m=0 \\ m \neq n}}^{N-1} x[m] \right)$$

$$+ \frac{1}{N^2} \left[ \text{var} \left( \sum_{\substack{m=0 \\ m \neq n}}^{N-1} x[m] \right) + E \left( \sum_{\substack{m=0 \\ m \neq n}}^{N-1} x[m] \right)^2 \right]$$

$$= \left( \frac{N-1}{N} \right)^2 (\sigma^2 + A^2) - 2 \frac{(N-1)}{N^2} A (N-1) A$$

$$+ \frac{1}{N^2} (N-1) \sigma^2 + \frac{1}{N^2} [(N-1) A]^2$$

$$= \sigma^2 \left[ \frac{N^2 - 2N + 1 + N-1}{N^2} \right] = \sigma^2 \frac{N-1}{N}$$

$$\Rightarrow E(\hat{\theta}^2) = \frac{1}{N-1} \sum_{n=0}^{N-1} r^2 \frac{N-1}{N} = \sigma^2$$

$\Rightarrow \hat{\theta}$  is unbiased.

II)  $\hat{\theta} = g(x[0])$

$$E(\hat{\theta}) = \theta \Rightarrow \int g(x[0]) p(x[0]) dx[0] = \theta$$

$$\text{or } \int_0^{\infty} g(u) \theta du = 1$$

$$\int_0^{\infty} g(u) du = 1 \quad \text{for all } \theta > 0$$

Now suppose a  $g$  could be found. Then

for any  $\theta_2 < \theta_1$ , we would have

$$\int_0^{\theta_2} g(u) du = 1$$

$$\int_0^{\theta_1} g(u) du = 1$$

and subtracting the two gives

$$\int_{\theta_1}^{\theta_2} g(u) du = 0 \quad \text{for any } \theta_2 < \theta_1$$

Clearly, we must have  $g(u) = 0$  for all  $u$  , which produces a biased estimator.

### Chapter 3

$$1) \quad p(x(n); \theta) = \frac{1}{\theta} (u(x(n)) - u(x(n) - \theta))$$

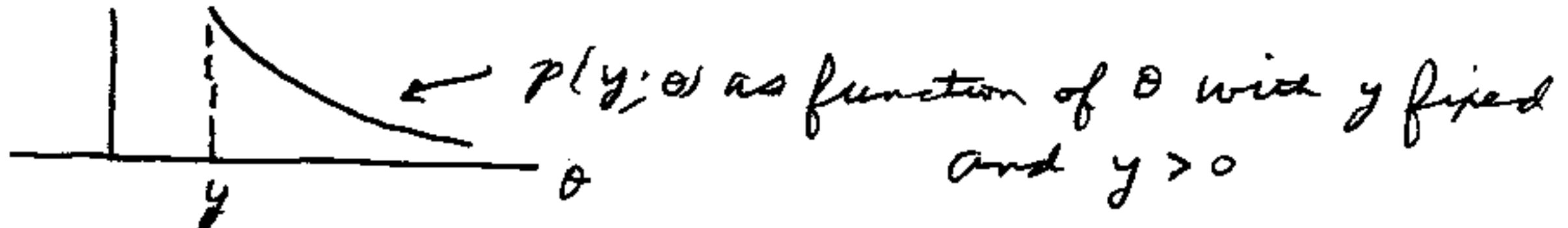
where  $u(x) = 1 \quad x > 0$   
 $\quad \quad \quad \quad \quad 0 \quad x \leq 0$

Since  $p(x; \theta) = \prod_{n=0}^{N-1} p(x(n); \theta)$ , it is enough to show that

$$E\left[\frac{\partial \ln p(x(n); \theta)}{\partial \theta}\right] \neq 0$$

(The expectation will be independent of  $n$ ).  
 Let  $y = x(n)$  so that

$$p(y; \theta) = \frac{1}{\theta} (u(y) - u(y - \theta))$$



For  $\theta > y$

$$\begin{aligned} E\left[\frac{\partial \ln p(y; \theta)}{\partial \theta}\right] &= E\left(\frac{\partial \ln 1/\theta}{\partial \theta}\right) \\ &= -1/\theta \neq 0 \end{aligned}$$

$$2) \quad x(0) = A + w(0)$$

$$p_x(x(0); A) = p_w(x(0) - A) = p(x(0) - A)$$

$$I(A) = E\left[\left(\frac{\partial \ln p(x(0) - A)}{\partial A}\right)^2\right]$$

$$\begin{aligned}
 &= E \left[ \left( \frac{\partial \ln p(x(0)-A)}{\partial (x(0)-A)} (-, ) \right)^2 \right] \\
 &= E \left[ \left( \frac{1}{p(x(0)-A)} \frac{\partial p(x(0)-A)}{\partial (x(0)-A)} \right)^2 \right] \\
 &= \int_{-\infty}^{\infty} \left( \frac{\partial p(x(0)-A)}{\partial (x(0)-A)} \right)^2 \frac{1}{p^2(x(0)-A)} p_x(x(0); A) dx(0) \\
 &= \int_{-\infty}^{\infty} \left( \frac{\partial p(x(0)-A)}{\partial (x(0)-A)} \right)^2 \frac{1}{p^2(x(0)-A)} p(x(0)-A) dx(0)
 \end{aligned}$$

Letting  $u = x(0) - A$

$$I(A) = \int_{-\infty}^{\infty} \frac{\left( \frac{dp(u)}{du} \right)^2}{p(u)} du$$

For  $p(u) = \frac{1}{\sqrt{2}\sigma} e^{-\frac{u^2}{2\sigma^2}}$  which is even in  $u$

$$\frac{dp}{du} = -\frac{1}{\sigma} \frac{1}{\sqrt{2}\sigma} e^{-\frac{u^2}{2\sigma^2}} u > 0$$

$$I(A) = 2 \int_0^{\infty} \frac{\frac{1}{\sigma^4} e^{-\frac{u^2}{2\sigma^2}}}{\frac{1}{\sqrt{2}\sigma} e^{-\frac{u^2}{2\sigma^2}}} du$$

$$= \frac{2\sqrt{2}\sigma}{\sigma^4} \int_0^{\infty} e^{-\frac{u^2}{2\sigma^2}} du = \frac{2\sqrt{2}\sigma}{\sigma^4} \frac{\sqrt{\pi}}{2}$$

$$= \frac{\pi}{\sigma^2}$$

$\Rightarrow \text{var}(\hat{A}) \geq \sigma^2/2$  The CRLB is half of that for the Gaussian case. In fact,

it can be shown that the Gaussian PDF produces the largest CRLB.

$$3) p(\mathbf{x}; A) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x(n) - Ar^n)^2}$$

$$\begin{aligned} \frac{\partial \ln p}{\partial A} &= -\frac{1}{2\sigma^2} \sum_n (x(n) - Ar^n) (-r^n) \\ &= \frac{1}{\sigma^2} \sum_n (x(n) - Ar^n) r^n \end{aligned}$$

$$\frac{\partial^2 \ln p}{\partial A^2} = -\frac{1}{\sigma^2} \sum_n r^{2n}$$

$$-E\left[\frac{\partial^2 \ln p}{\partial A^2}\right] = \frac{1}{\sigma^2} \sum_n r^{2n}$$

$$\text{or } \text{var}(\hat{A}) \geq \frac{\sigma^2}{\sum_{n=0}^{N-1} r^{2n}} \quad (\text{or use (3.14)})$$

To show that an efficient estimator exists

$$\begin{aligned} \frac{\partial \ln p}{\partial A} &= \frac{1}{\sigma^2} \left( \sum_n x(n) r^n - A \sum_n r^{2n} \right) \\ &= \underbrace{\frac{\sum_n r^{2n}}{\sigma^2}}_{I(A)} \left( \underbrace{\frac{\sum_n x(n) r^n}{\sum_n r^{2n}}}_{\hat{A}} - A \right) \end{aligned}$$

$\hat{A}$  is efficient and  $1/I(A)$  is its variance.

$$\text{var}(\hat{A}) \rightarrow \sigma^2(1-r^2) \quad 0 < r < 1$$

$$\rightarrow 0 \quad r \geq 1$$

as  $N \rightarrow \infty$ .

$$5) p(\underline{x}; r) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x(n) - r^n)^2}$$

$$\frac{\partial \ln p}{\partial r} = \frac{1}{\sigma^2} \sum_n (x(n) - r^n) n r^{n-1}$$

Can't put into form  $I(r)(\hat{r} - r) \Rightarrow$   
no efficient estimator.

$$\frac{\partial^2 \ln p}{\partial r^2} = \frac{1}{\sigma^2} \sum_n [x(n) n(n-1) r^{n-2} - n(2n-1) r^{2n-2}]$$

$$E \left[ \frac{\partial^2 \ln p}{\partial r^2} \right] = \frac{1}{\sigma^2} \sum_n [n(n-1) r^{2n-2} - n(2n-1) r^{2n-2}] \\ = -\frac{1}{\sigma^2} \sum_n n^2 r^{2n-2}$$

$$\text{var}(\hat{r}) \geq \frac{\sigma^2}{\sum_{n=0}^{N-1} n^2 r^{2n-2}} \quad (\text{could also use (3.14)})$$

$$5) p(\underline{x}; A) = \frac{1}{(2\pi)^{N/2} \det(\Sigma)^{1/2}} e^{-\frac{1}{2} (\underline{x} - \underline{A}\underline{1})^T \Sigma^{-1} (\underline{x} - \underline{A}\underline{1})}$$

$$\text{where } \underline{1} = [1 1 \dots 1]^T$$

$$\frac{\partial \ln p}{\partial A} = -\frac{1}{2} \frac{\partial}{\partial A} \left[ \underline{x}^T \Sigma^{-1} \underline{x} - 2 \underline{1}^T \Sigma^{-1} \underline{x} A + \underline{1}^T \Sigma^{-1} \underline{1} A^2 \right]$$

$$\begin{aligned}
 &= \underline{I}^T \underline{C}^{-1} \underline{x} - \underline{I}^T \underline{C}^{-1} \underline{I} A \\
 &= \underbrace{\underline{I}^T \underline{C}^{-1} \underline{I}}_{\mathcal{I}(A)} \left( \underbrace{\frac{\underline{I}^T \underline{C}^{-1} \underline{x}}{\underline{I}^T \underline{C}^{-1} \underline{I}} - A}_{\hat{A}} \right)
 \end{aligned}$$

$\hat{A}$  is efficient and has variance  $1/\mathcal{I}(A)$ .

6)  $x(0) \sim N(\theta, 1)$

$$x(1) \sim N(\theta, 1) \quad \theta \geq 0$$

$$N(\theta, 2) \quad \theta < 0$$

$$p(x; \theta) = \begin{cases} \frac{1}{2\pi} e^{-\frac{1}{2} [(x(0)-\theta)^2 + (x(1)-\theta)^2]} & \theta \geq 0 \\ \frac{1}{2\pi\sqrt{2}} e^{-\frac{1}{2} [(x(0)-\theta)^2 + \frac{1}{2}(x(1)-\theta)^2]} & \theta < 0 \end{cases}$$

For  $\theta \geq 0$

$$\begin{aligned}
 \frac{\partial \ln p}{\partial \theta} &= -\frac{1}{2} [2(x(0)-\theta)(-1) + 2(x(1)-\theta)(-1)] \\
 &= (x(0)-\theta) + (x(1)-\theta)
 \end{aligned}$$

$$\frac{\partial^2 \ln p}{\partial \theta^2} = -2 \Rightarrow E\left(-\frac{\partial^2 \ln p}{\partial \theta^2}\right) = 2$$

$$\text{var}(\hat{\theta}) \geq \frac{1}{2}$$

For  $\theta < 0$

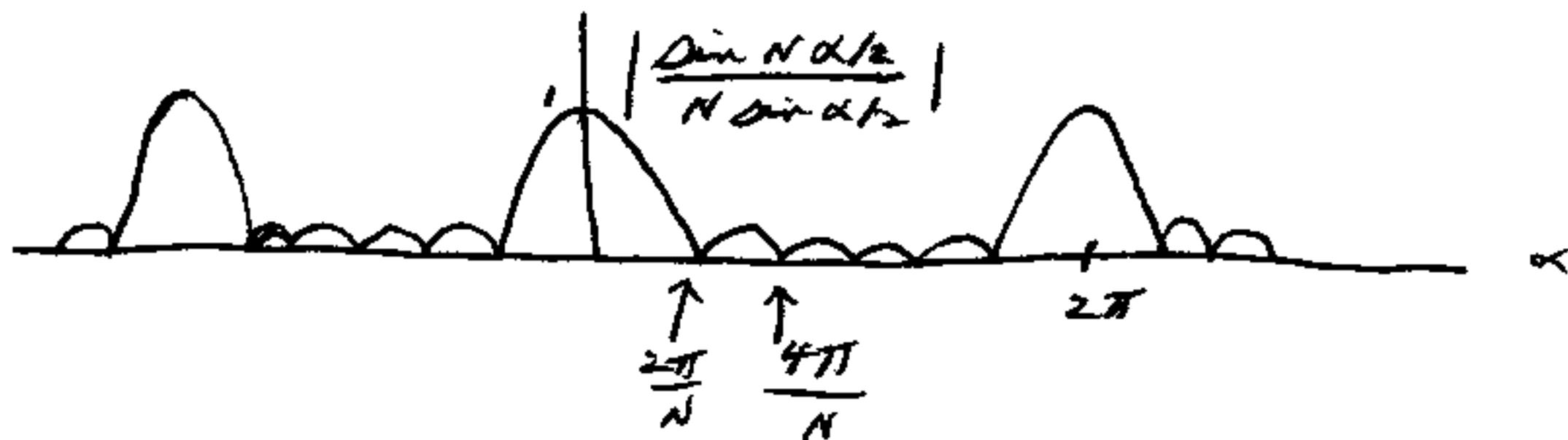
$$\begin{aligned}
 \frac{\partial \ln p}{\partial \theta} &= -\frac{1}{2} [-2(x(0)-\theta) - (x(1)-\theta)] \\
 &= (x(0)-\theta) + \frac{1}{2}(x(1)-\theta)
 \end{aligned}$$

$$\frac{\partial^2 \ln p}{\partial \theta^2} = -3/2 \Rightarrow E\left(-\frac{\partial^2 \ln p}{\partial \theta^2}\right) = 3/2$$

$$\text{var}(\hat{\theta}) \geq 2/3$$

7) Let  $\alpha = 4\pi f_0$ ,  $\beta = 2\phi$

$$\begin{aligned} \frac{1}{N} \sum_{n=0}^{N-1} \cos(\alpha n + \beta) &= \frac{1}{N} \operatorname{Re} \left( \sum_n e^{j(\alpha n + \beta)} \right) \\ &= \frac{1}{N} \operatorname{Re} \left( e^{j\beta} \frac{1 - e^{j\alpha N}}{1 - e^{j\alpha}} \right) \\ &= \frac{1}{N} \operatorname{Re} \left( e^{j\beta} \frac{e^{j\alpha N/2}}{e^{j\alpha/2}} \frac{e^{-j\alpha N/2} - e^{j\alpha N/2}}{e^{-j\alpha/2} - e^{j\alpha/2}} \right) \\ &= \frac{1}{N} \operatorname{Re} \left[ e^{j\beta} e^{j\alpha \left(\frac{N-1}{2}\right)} \frac{\sin N\alpha/2}{\sin \alpha/2} \right] \\ &= \frac{\sin N\alpha/2}{N \sin \alpha/2} \cos\left(\alpha \left(\frac{N-1}{2}\right) + \beta \right) \end{aligned}$$



As long as  $\alpha$  is not near 0 or  $2\pi$ , this term is approximately zero. But  $\alpha = 4\pi f_0 \Rightarrow f_0$  cannot be near 0 or  $1/2$ .

8)  $p(x; A) = \frac{1}{(2\pi\sigma^2/N)^{1/2}} e^{-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n]-A)^2}$

$$\frac{\partial \ln p}{\partial A} = \frac{1}{\sigma^2} \sum_n (x[n]-A)$$

$$\left(\frac{\partial \ln p}{\partial A}\right)^2 = \frac{1}{\sigma^4} \sum_m \sum_n (x[m]-A)(x[n]-A)$$

$$\begin{aligned} E\left[\left(\frac{\partial \ln p}{\partial A}\right)^2\right] &= \frac{1}{\sigma^4} \sum_m \sum_n \underbrace{E[(x[m]-A)(x[n]-A)]}_{\sigma^2 \delta_{mn}} \\ &= \frac{1}{\sigma^4} \sum_n \sigma^2 = N/\sigma^2 \end{aligned}$$

$$\text{var}(\hat{A}) \geq \sigma^2/N$$

9) Using (3.32)

$$I(A) = \left(\frac{\partial \underline{u}(A)}{\partial A}\right)^T C^{-1} \frac{\partial \underline{u}(A)}{\partial A}$$

$$\underline{u}(A) = [A \ A]^T \Rightarrow \frac{\partial \underline{u}(A)}{\partial A} = 1$$

$$\text{var}(\hat{A}) \geq \frac{1}{1^T C^{-1} 1} \quad (\text{or use approach of Prob 3.5})$$

$$C^{-1} = \frac{1}{\sigma^2} \begin{bmatrix} 1 - \rho & \rho \\ -\rho & 1 \end{bmatrix}$$

$$\Rightarrow \text{var}(\hat{A}) \geq \frac{\sigma^2 (1 - \rho^2)}{2 - 2\rho} = \frac{\sigma^2}{2} (1 + \rho)$$

If  $\rho = 0$ ,  $\text{var}(\hat{A}) \geq \sigma^2/2$ , as expected.  
 If  $\rho \rightarrow 1$ ,  $\text{var}(\hat{A}) \geq \sigma^2$ . This is the same bound as for one sample and occurs because as  $\rho \rightarrow 1$ ,  $w[0]$  and  $w[1]$  will be equal. Hence, we have only one independent data sample. If  $\rho \rightarrow -1$ ,  $\text{var}(\hat{A}) \geq 0$  and in fact, in this case  $w[0] = -w[1]$ . Thus,

$$\begin{aligned}\hat{A} &= \frac{1}{2}(x[0] + x[1]) \\ &= \frac{1}{2}(A + w[0] + A - w[0]) = A\end{aligned}$$

for any realization of the noise samples. Additivity property of Fisher information only holds for independent samples. In this example we could have

$$i(A) \leq I(A) < \infty i(A)$$

where  $i(A) = 1/\sigma^2$ .

$$(10) \quad [I(\underline{\theta})]_{ij} = E \left[ \frac{\partial \ln p}{\partial \theta_i} \frac{\partial \ln p}{\partial \theta_j} \right]$$

$$I(\underline{\theta}) = E \left[ \frac{\partial \ln p}{\partial \underline{\theta}} \frac{\partial \ln p}{\partial \underline{\theta}}^T \right]$$

$$\begin{aligned}\underline{a}^T I(\underline{\theta}) \underline{a} &= E \left[ \underline{a}^T \frac{\partial \ln p}{\partial \underline{\theta}} \frac{\partial \ln p}{\partial \underline{\theta}}^T \underline{a} \right] \\ &= E \left[ (\underline{a}^T \frac{\partial \ln p}{\partial \underline{\theta}})^2 \right] \geq 0\end{aligned}$$

for all  $\underline{a} \Rightarrow I(\underline{\theta})$  is positive semidefinite

for all  $\underline{\theta}$ .

From Prob 3.3  $\underline{\theta} = \begin{pmatrix} A \\ r \end{pmatrix}$  and using  
(3.31)

$$[\underline{\mathcal{I}}(\underline{\theta})]_{ij} = \frac{1}{\sigma^2} \frac{\partial \underline{u}(\underline{\theta})}{\partial \theta_i} \frac{\partial \underline{u}(\underline{\theta})}{\partial \theta_j}$$

where  $\underline{u}(\underline{\theta}) = \begin{bmatrix} A \\ Ar \\ \vdots \\ Ar^{N-1} \end{bmatrix}$

$$\frac{\partial \underline{u}(\underline{\theta})}{\partial A} = [1 \ r \ \dots \ r^{N-1}]^T$$

$$\frac{\partial \underline{u}(\underline{\theta})}{\partial r} = A[0 \ 1 \ \dots \ (N-1)r^{N-2}]^T$$

$$\underline{\mathcal{I}}(\underline{\theta}) = \frac{1}{\sigma^2} \begin{bmatrix} \sum_{n=0}^{N-1} r^n & A \sum_{n=0}^{N-1} n r^{2n-1} \\ A \sum_{n=0}^{N-1} n r^{2n-1} & A^2 \sum_{n=0}^{N-1} n^2 r^{2n-2} \end{bmatrix}$$

If  $A = 0$ ,  $\underline{\mathcal{I}}(\underline{\theta})$  is not positive definite since its determinant is zero. Clearly, in this case there is no information in the data about  $r$ .

- 11) Since  $\underline{\mathcal{I}}(\underline{\theta})$  is positive definite,  $a > 0, c > 0$  and  $\det(\underline{\mathcal{I}}(\underline{\theta})) > 0$  or  $ac - b^2 > 0$ . But

$$(\underline{I}^{-1}(\underline{\theta}))_{ii} = \frac{c}{ac - b^2} = \frac{1}{a - b^2/c} \geq \frac{1}{a}$$

Thus, the CRLB is almost always increased when we estimate additional parameters.

Equality holds if and only if  $b = 0$  or the Fisher information matrix is "decoupled", i.e., it is diagonal. In this case the additional parameter does not affect the CRLB.

$$(12) \quad I^2 = (\underline{e}_i^T \sqrt{\underline{I}(\underline{\theta})} \sqrt{\underline{I}^{-1}(\underline{\theta})} \underline{e}_i)^2$$

$$\text{since } \sqrt{\underline{I}^{-1}(\underline{\theta})} = (\sqrt{\underline{I}(\underline{\theta})})^{-1}$$

$$I^2 \leq \underline{e}_i^T \underline{I}(\underline{\theta}) \underline{e}_i \quad \underline{e}_i^T \underline{I}^{-1}(\underline{\theta}) \underline{e}_i$$

$$I \leq [\underline{I}(\underline{\theta})]_{ii} [\underline{I}^{-1}(\underline{\theta})]_{ii}$$

$$\Rightarrow [\underline{I}^{-1}(\underline{\theta})]_{ii} \geq \frac{1}{[\underline{I}(\underline{\theta})]_{ii}}$$

New bound achieved when an efficient estimator exists and  $\underline{I}(\underline{\theta})$  is diagonal.

$$(13) \quad \text{From (3.33) with } s(n; \underline{\theta}) = \sum_{k=0}^{p-1} A_k n^k$$

$$\begin{aligned}
 [\mathbb{E}(\theta)]_{ij} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left( \frac{\partial}{\partial \theta_i} \sum_{k=0}^{p-1} A_k n^k \right) \\
 &\quad \left( \frac{\partial}{\partial \theta_j} \sum_{k=0}^{p-1} A_k n^k \right) \\
 &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} n^{i-1} n^{j-1} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} n^{i+j-2}
 \end{aligned}$$

where  $\theta_i = A_{i-1}$ ,  $i = 1, 2, \dots, p$

$$(14) \quad \hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x(n)$$

If we condition the mean and variance on  $A$ , then we can regard  $\hat{A}$  as the observed value  $A_0$ . Hence, from Example 3.3

$$E(\hat{A} | A = A_0) = A_0$$

$$\text{var}(\hat{A} | A = A_0) = \sigma^2/N$$

and since  $\sigma^2/N \rightarrow 0$  as  $N \rightarrow \infty$ ,  $\hat{A} \rightarrow A_0$ .

Now consider  $A$  as a random variable  
as  $N \rightarrow \infty$

$$\begin{aligned}
 \text{var}(\hat{\sigma}_A^2) &= \text{var}(\hat{A}^2) \\
 &\rightarrow \text{var}(A^2)
 \end{aligned}$$

$$\begin{aligned}
 \text{But } \text{var}(A^2) &= E(A^4) - E(A^2)^2 \\
 &= 3\sigma_A^4 - \sigma_A^4 = 2\sigma_A^4
 \end{aligned}$$

so that  $\text{var}(\hat{\sigma}_A^2) \rightarrow 2\sigma_A^4$ , which is just the LRLB as  $N \rightarrow \infty$ .  $\hat{\sigma}_A^2$  cannot be estimated without error since even as

$N \rightarrow \infty$ , although we can nullify the noise effects by averaging ( $\hat{A} \rightarrow A_0$ ), we cannot reduce the random nature of  $A$ . This is because we have only one realization of  $A$ . Since  $\hat{\sigma}_A^2$  is the square of  $\hat{A}$ , it will also exhibit the same variability.

15) Because the  $x(w)$ 's are independent

$$I(\rho) = N i(\rho)$$

where  $i(\rho)$  is the Fisher information for a single vector sample. Using (3.32)

$$i(\rho) = \frac{1}{2} \pi \left[ (\underline{C}^{-1}(\rho) \frac{\partial \underline{C}(\rho)}{\partial \rho})^2 \right]$$

$$\underline{C}^{-1}(\rho) = \frac{\begin{bmatrix} 1 & \rho \\ -\rho & 1 \end{bmatrix}}{1-\rho^2} \quad \frac{\partial \underline{C}(\rho)}{\partial \rho} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$D = \underline{C}^{-1}(\rho) \frac{\partial \underline{C}(\rho)}{\partial \rho} = \frac{\begin{bmatrix} -\rho & 1 \\ 1 & -\rho \end{bmatrix}}{1-\rho^2}$$

$$D^2 = \frac{\begin{bmatrix} -\rho & 1 \\ 1 & -\rho \end{bmatrix} \begin{bmatrix} -\rho & 1 \\ 1 & -\rho \end{bmatrix}}{(1-\rho^2)^2} = \frac{\begin{bmatrix} 1+\rho^2 & - \\ - & 1+\rho^2 \end{bmatrix}}{(1-\rho^2)^2}$$

$$i(\rho) = \frac{1}{2} \frac{2+2\rho^2}{(1-\rho^2)^2} = \frac{1+\rho^2}{(1-\rho^2)^2}$$

$$\text{var}(\hat{p}) \geq \frac{(1-p^2)^2}{N(1+p^2)}$$

$$(16) \quad I(P_0) = \frac{1}{2} \text{tr} \left[ (\underline{\Sigma}'(P_0) \frac{\partial \underline{\Sigma}(P_0)}{\partial P_0})^2 \right]$$

$$\begin{aligned} \text{Let } R_{xx}[k] &= \mathcal{F}^{-1}\{P_{xx}(f)\} \\ &= \mathcal{F}^{-1}\{P_0 Q(f)\} \\ &= P_0 \mathcal{F}^{-1}\{Q(f)\} \end{aligned}$$

Let  $\mathcal{F}^{-1}\{Q(f)\} = g[k]$  and construct the Toeplitz autocorrelation matrix  $\underline{\Sigma}_g$  (of dimension  $N \times N$ ), where

$$(\underline{\Sigma}_g)_{ij} = g[i-j]$$

$$\text{Then, } \underline{\Sigma}(P_0) = P_0 \underline{\Sigma}_g \text{ and}$$

$$\underline{\Sigma}'(P_0) \frac{\partial \underline{\Sigma}(P_0)}{\partial P_0} = \frac{1}{P_0} \underline{\Sigma}_g' \underline{\Sigma}_g = \frac{1}{P_0} \underline{\Sigma}$$

$$I(P_0) = \frac{1}{2} \text{tr} \left( \frac{1}{P_0^2} \underline{\Sigma}^2 \right) = \frac{N}{2P_0^2}$$

$$\text{var}(\hat{P}_0) \geq 2P_0^2/N$$

Using the asymptotic form

$$\begin{aligned} I(P_0) &= \sum_{f=-\frac{N}{2}}^{\frac{N}{2}} \left( \frac{\partial \ln P_{xx}(f; P_0)}{\partial P_0} \right)^2 df \\ &= \sum_{f=-\frac{N}{2}}^{\frac{N}{2}} \left( \frac{\partial \ln P_0 Q(f)}{\partial P_0} \right)^2 df \end{aligned}$$

$$= \frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{P_0^2} dt = \frac{N}{2P_0^2}$$

The two CRLB's are identical but in general this will not be true.

- 17) All elements of  $\underline{\underline{I}}(\underline{\Omega})$  are the same except the sums run from  $n = -M$  to  $n = M$ . Thus, now  $(\underline{\underline{I}}(\underline{\Omega}))_{23} = 0$  since

$$\sum_{n=-M}^M n = 0$$

This makes  $\underline{\underline{I}}(\underline{\Omega})$  diagonal. Letting  $N = 2M + 1$ ,

$$\text{var}(\hat{A}) \geq \frac{2\sigma^2}{N} \text{ same as before}$$

$$\text{var}(\hat{\phi}) \geq \frac{2\sigma^2}{NA^2} = \frac{1}{N\eta} \text{ less than before}$$

$$\text{var}(\hat{f}_0) \geq \frac{\sigma^2}{2A^2\pi^2 \sum_{n=-M}^M n^2}$$

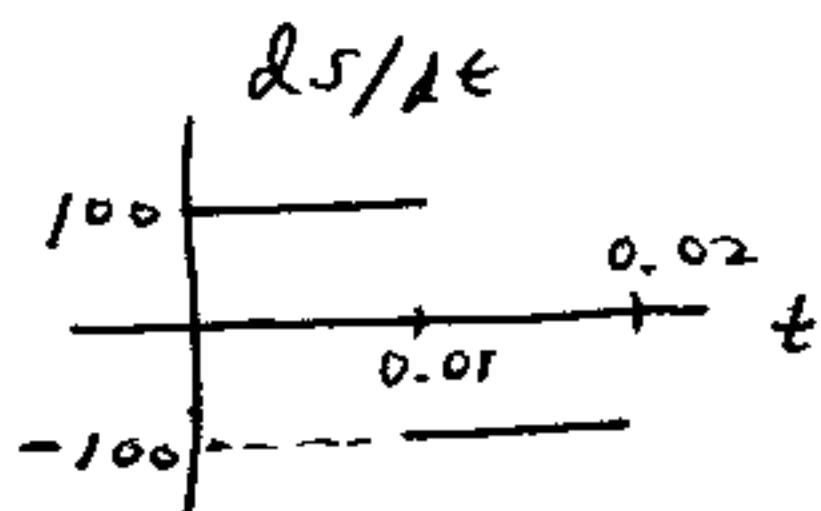
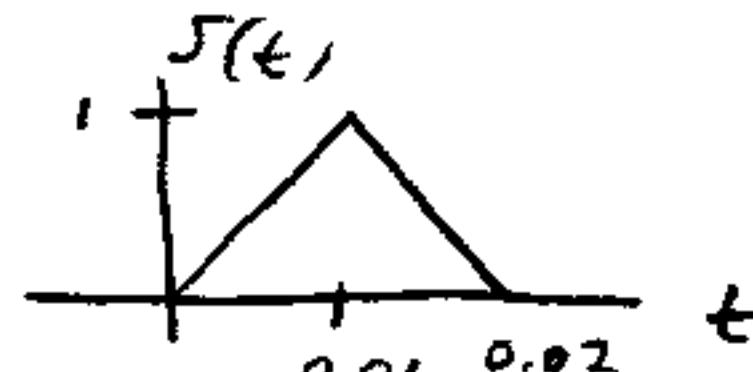
$$\begin{aligned} \text{But } \sum_{n=-M}^M n^2 &= 2 \sum_{n=1}^M n^2 = \frac{2M(M+1)(2M+1)}{6} \\ &= \frac{1}{3} \left( \frac{N-1}{2} \right) \left( \frac{N+1}{2} \right) N = \frac{N(N^2-1)}{12} \end{aligned}$$

$$\text{var}(\hat{f}_0) \geq \frac{6\sigma^2}{A^2\pi^2 N(N^2-1)} = \frac{3}{\eta\pi^2 N(N^2-1)}$$

$$= \frac{12}{(2\pi)^2 \eta N(N^2-1)} \quad \text{Same as before}$$

$$18) \quad \text{var}(\hat{R}) \geq \frac{c^2/4}{\frac{\epsilon}{N_0 t_2} \bar{F}^2}$$

$$\text{where } \bar{F}^2 = \frac{\int_0^{T_s} \left( \frac{ds}{dt} \right)^2 dt}{\int_0^{T_s} s^2(t) dt}$$



$$\bar{F}^2 = \frac{\int_0^{0.02} (100)^2 dt}{\epsilon} = \frac{200}{\epsilon}$$

$$\text{var}(\hat{R}) \geq \frac{c^2/4}{\frac{1}{N_0 t_2} 200} = \frac{(1500)^2/4}{10^6 \cdot 200}$$

$$= 0.00281$$

$$\text{or } \sqrt{\text{var}(\hat{R})} \geq 0.05 \text{ m}$$

19) From Example 3.15

$$\text{var}(\hat{\beta}) \geq \frac{12}{(2\pi)^2 M \cdot \frac{M+1}{M-1} \left(\frac{L}{\lambda}\right)^2 \sin^2 \beta}$$

$$\begin{aligned} \text{For } \beta = 90^\circ, \gamma = 1, F_0 = 10, L &= (M-1)d \\ &= (M-1)\lambda/2 \end{aligned}$$

$$\text{var}(\hat{\beta}) \geq \frac{12}{(2\pi)^2 M \cdot \frac{M+1}{M-1} \left(\frac{M-1}{2}\right)^2}$$

$$= \frac{12}{(2\pi)^2 \frac{M}{4} (M^2-1)}$$

$$M(M^2-1) \geq \frac{48}{(2\pi)^2 (5\pi/180)^2} = 159.7$$

or  $M \geq 6$ . But then

$$\begin{aligned} L &= (M-1)N/2 = (M-1) \frac{L}{2F_0} = \frac{5(3 \times 10^8)}{2 \times 10^6} \\ &= 750 \text{ m} \end{aligned}$$

This is clearly impossible.

$$\begin{aligned} 20) \quad \text{var}(\hat{P}_{xx}(f_1)) &\geq \frac{\left(\frac{\partial P_{xx}(f_1)}{\partial a(1)}\right)^2}{I(a(1))} \\ &= \frac{\left(\frac{\partial P_{xx}(f_1)}{\partial a(1)}\right)^2}{N/(1-a(1))^2} \end{aligned}$$

using results from Example 3.16.

$$P_{xx}(f) = \frac{\sigma_u^2}{|A(f)|^2} \quad \text{where } A(f) = 1 + a_{(1)} e^{-j2\pi f}$$

$$\frac{\partial P_{xx}(f)}{\partial a_{(1)}} = \sigma_u^2 \frac{\partial}{\partial a_{(1)}} \left( \frac{1}{A(f) A^*(f)} \right)$$

$$= -\frac{\sigma_u^2}{|A(f)|^4} \frac{\partial}{\partial a_{(1)}} A(f) A^*(f)$$

$$= -\frac{\sigma_u^2}{|A(f)|^4} (A(f) e^{j2\pi f} + A^*(f) e^{-j2\pi f})$$

$$= -\frac{\sigma_u^2}{|A(f)|^4} \underbrace{2 \operatorname{Re}(A(f) e^{j2\pi f})}_{a_{(1)} + \cos 2\pi f}$$

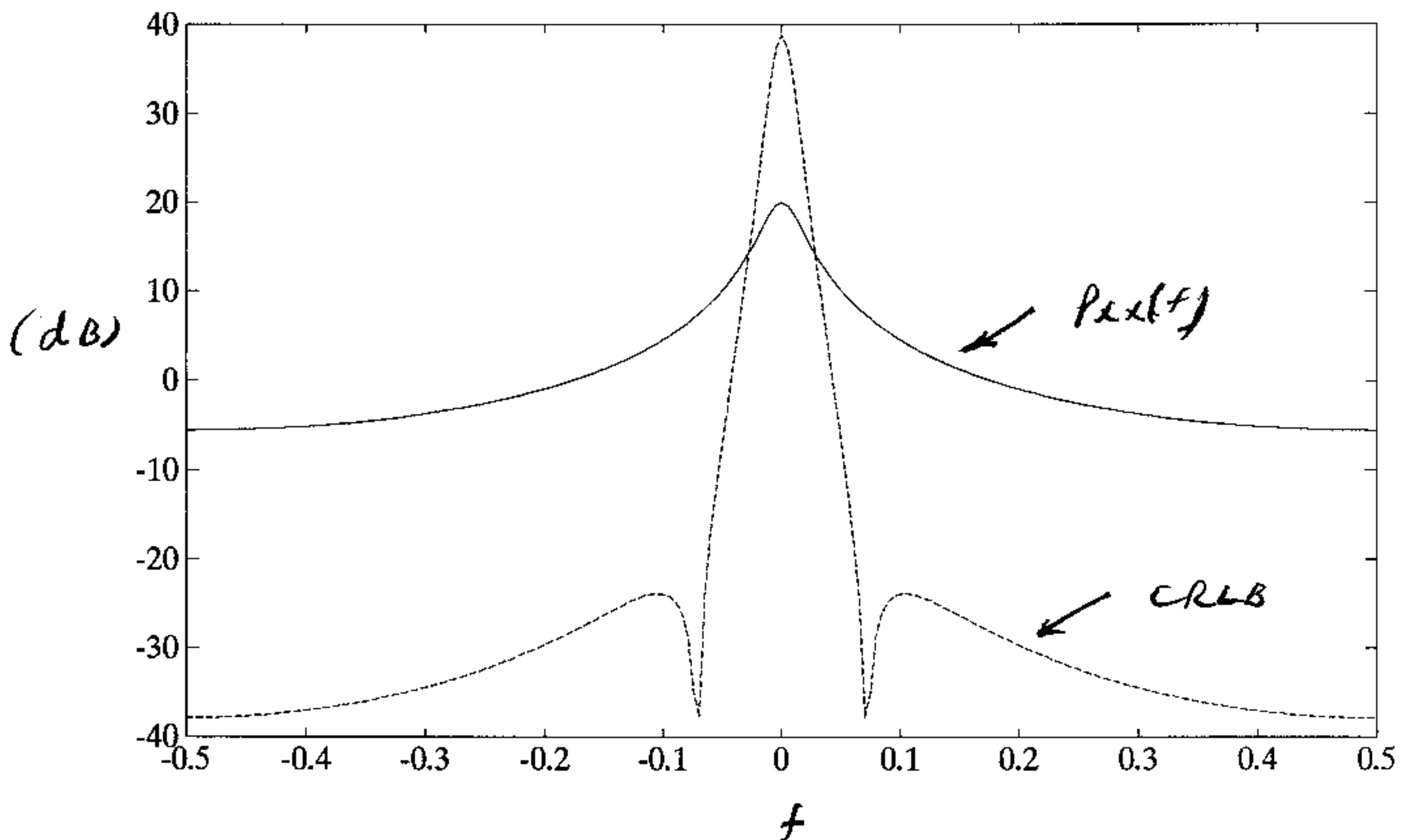
$$\operatorname{var}(\hat{P}_{xx}(f)) \geq \frac{4\sigma_u^4(1-a_{(1)}^2)}{N} \frac{(a_{(1)} + \cos 2\pi f)^2}{|A(f)|^8}$$

For the given values

$$\operatorname{var}(\hat{P}_{xx}(f)) \geq 0.0076 \frac{(a_{(1)} + \cos 2\pi f)^2}{|A(f)|^8}$$

$$= \frac{0.0076 (-0.9 + \cos 2\pi f)^2}{|1 - 0.9 e^{-j2\pi f}|^8}$$

See Figure.



Prob. 3.20

Because of the sensitivity of the PSD to small changes in  $a(1)$  for  $f$  near zero, the variance is highest at DC. Note that

$$\begin{aligned} \left. \frac{\partial P_{xx}(f)}{\partial a(1)} \right|_{f=0} &= -\frac{\sigma_u^2 \cdot 2(1+a(1))}{(1+a(1))^4} \\ &= \frac{-2\sigma_u^2}{(1+a(1))^3} \end{aligned}$$

and for this example

$$\left[ \left. \frac{\partial P_{xx}(f)}{\partial a(1)} \right|_{f=0} \right]^2 = 4 \times 10^6$$

## Chapter 4

1) This fits linear model form.

$$\underline{x} = \underline{H}\underline{\theta} + \underline{w}$$

$$\underline{H} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ r_1 & r_2 & \dots & r_p \\ \vdots & \vdots & & \vdots \\ r_1^{N-1} & r_2^{N-1} & \dots & r_p^{N-1} \end{bmatrix} \quad \underline{\theta} = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_p \end{bmatrix}$$

$$\hat{\underline{\theta}} = (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x} \quad C_{\hat{\theta}} = \sigma^2 (\underline{H}^T \underline{H})^{-1}$$

For  $p=2$ ,  $r_1 = 1$ ,  $r_2 = -1$  and  $N$  even

$$\underline{H} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ \vdots & \vdots \\ 1 & -1 \end{bmatrix} \Rightarrow \underline{H}^T \underline{H} = \begin{bmatrix} N & 0 \\ 0 & N \end{bmatrix} = N \underline{I}$$

since columns are orthogonal

$$\hat{\underline{\theta}} = \frac{1}{N} \underline{H}^T \underline{x} = \begin{bmatrix} \frac{1}{N} \sum_{n=0}^{N-1} x[n] \\ \frac{1}{N} \sum_{n=0}^{N-1} (-1)^n x[n] \end{bmatrix}$$

$$C_{\hat{\theta}} = \frac{\sigma^2}{N} \underline{I}$$

2) First assume columns of  $H$  are linearly independent  
 Then,  $\underline{H}\underline{x} = \sum_{i=1}^p x_i h_i \neq \underline{0}$  if  $\underline{x} \neq \underline{0}$

$$\Rightarrow \underline{x}^T \underline{H}^T \underline{H} \underline{x} = \| \underline{H}\underline{x} \|^2 > 0 \text{ for } \underline{x} \neq \underline{0}$$

$\Rightarrow \underline{H}^T \underline{H}$  is positive definite

Now assume  $\underline{H}^T \underline{H}$  is positive definite or

$$\underline{x}^T \underline{H}^T \underline{H} \underline{x} > 0 \quad \text{for all } \underline{x} \neq \underline{0}$$

$$\text{or} \quad \|H \underline{x}\|^2 > 0 \quad \text{for all } \underline{x} \neq \underline{0}$$

$$\Rightarrow H \underline{x} \neq \underline{0} \quad \text{for all } \underline{x} \neq \underline{0}$$

$\Rightarrow$  Columns of  $H$  are linearly independent

It can further be shown that for matrices of the form  $\underline{H}^T \underline{H}$ , invertibility is equivalent to being positive definite.

$$3) \quad \underline{H}^T \underline{H} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1+\epsilon \\ 1 & 1 & 1+\epsilon \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1+\epsilon \end{bmatrix} = \begin{bmatrix} 3 & 3+\epsilon \\ 3+\epsilon & 2+(1+\epsilon)^2 \end{bmatrix}$$

$$(H^T H)^{-1} = \frac{\begin{bmatrix} 2+(1+\epsilon)^2 & -(3+\epsilon) \\ -(3+\epsilon) & 3 \end{bmatrix}}{3[2+(1+\epsilon)^2] - (3+\epsilon)^2}$$

$$= \frac{\begin{bmatrix} 3+2\epsilon+\epsilon^2 & -(3+\epsilon) \\ -(3+\epsilon) & 3 \end{bmatrix}}{2\epsilon^2}$$

As  $\epsilon \rightarrow 0$ , all elements  $\rightarrow \infty$

$$\begin{aligned}
 \hat{\underline{\theta}} &= (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x} \\
 &= \frac{1}{2\epsilon^2} \begin{bmatrix} 3+2\epsilon+\epsilon^2 & -(3+\epsilon) \\ -3-\epsilon & 3 \end{bmatrix} \begin{bmatrix} 6 \\ 6+2\epsilon \end{bmatrix} \\
 &= \frac{1}{2\epsilon^2} \begin{bmatrix} 18+12\epsilon+6\epsilon^2-18-6\epsilon-6\epsilon-2\epsilon^2 \\ -18-6\epsilon+18+6\epsilon \end{bmatrix} \\
 &= \begin{pmatrix} 2 \\ 0 \end{pmatrix}
 \end{aligned}$$

Hence, even as  $\epsilon \rightarrow 0$ ,  $\hat{\underline{\theta}} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ . This is because  $\underline{x}$  lies in the subspace spanned by the first column of  $\underline{H}$ , which does not depend on  $\epsilon$ .

$$4) \quad \hat{\underline{\theta}} \sim N(\underline{\theta}, \sigma^2 (\underline{H}^T \underline{H})^{-1})$$

$$\hat{\underline{s}} = \underline{H} \hat{\underline{\theta}} \sim N(\underline{H} \underline{\theta}, \sigma^2 \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T)$$

Note that the Covariance matrix is singular since  $\underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T$  is a projection matrix of rank p. (See Chapter 8).

For Example 4.2

$$\hat{s}(n) = \sum_{k=1}^M \hat{a}_k \cos 2\pi \frac{k n}{N} + \sum_{k=1}^M \hat{b}_k \sin 2\pi \frac{k n}{N}$$

$$\text{where } \hat{a}_n = \frac{2}{N} \sum_{n=0}^{N-1} x(n) \cos 2\pi \frac{kn}{N}$$

$$\hat{b}_k = \frac{2}{N} \sum_{n=0}^{N-1} x(n) \sin 2\pi \frac{kn}{N}$$

$$\text{Also, } (\underline{H}^T \underline{H})^{-1} = \frac{2}{N} \underline{I} \Rightarrow$$

$$\hat{\xi} \sim N(\xi, \frac{2\sigma^2}{N} \underline{H} \underline{H}^T) \quad \text{where } \xi = H\theta.$$

5) Let  $\omega_k = \frac{2\pi}{N} kn$

$$\sum_{n=0}^{N-1} \cos \omega_k \cos \omega_\ell = \frac{1}{2} \sum_n [\cos(\omega_k + \omega_\ell) + \cos(\omega_k - \omega_\ell)]$$

$$= \frac{1}{2} \operatorname{Re} \sum_n e^{j(\omega_k + \omega_\ell)} + \frac{1}{2} \operatorname{Re} \sum_n e^{j(\omega_k - \omega_\ell)}$$

$$\text{But } \sum_{n=0}^{N-1} e^{j(\omega_k + \omega_\ell)} = \sum_{n=0}^{N-1} e^{j \frac{2\pi}{N} (k+\ell)n}$$

$$= \frac{1 - e^{j \frac{2\pi}{N} (k+\ell)N}}{1 - e^{j \frac{2\pi}{N} (k+\ell)}} = \frac{1 - e^{j 2\pi(k+\ell)}}{1 - e^{j \frac{2\pi}{N} (k+\ell)}} = 0$$

and similarly for  $\sum_n e^{j(\omega_k - \omega_\ell)}$ .

b) From Example 4.2

$\hat{a}_n \sim N(a_n, 2\sigma^2/N)$   $\hat{b}_k \sim N(b_k, 2\sigma^2/N)$   
and  $\hat{a}_n, \hat{b}_k$  are independent

$$\begin{aligned}
 E(\hat{P}) &= \frac{E(\hat{a}_k^2) + E(\hat{b}_k^2)}{2} \\
 &= \frac{\text{var}(\hat{a}_k) + E(\hat{a}_k)^2 + \text{var}(\hat{b}_k) + E(\hat{b}_k)^2}{2} \\
 &= \frac{a_k^2 + b_k^2 + 4\sigma^2/N}{2} = P + \frac{2\sigma^2}{N}
 \end{aligned}$$

$$\text{var}(\hat{P}) = \frac{\text{var}(\hat{a}_k^2) + \text{var}(\hat{b}_k^2)}{4}$$

since  $\hat{a}_k, \hat{b}_k$  are independent

But  $\text{var}(\hat{a}_k^2)$  can be found as follows:

$$\text{If } \hat{z} \sim N(u, \sigma^2)$$

$$\text{var}(\hat{z}^2) = 4u^2\sigma^2 + 2\sigma^4$$

(see development preceding (3.19))

$$\Rightarrow \text{var}(\hat{a}_k^2) = 4a_k^2 \frac{2\sigma^2}{N} + 2\left(\frac{2\sigma^2}{N}\right)^2$$

$$\text{var}(\hat{b}_k^2) = 4b_k^2 \frac{2\sigma^2}{N} + 2\left(\frac{2\sigma^2}{N}\right)^2$$

$$\text{var}(\hat{P}) = 2P \frac{2\sigma^2}{N} + \left(\frac{2\sigma^2}{N}\right)^2 = \frac{2\sigma^2}{N}(2P + \frac{2\sigma^2}{N})$$

$$\frac{E(\hat{P})^2}{\text{var}(\hat{P})} = \frac{(P + \frac{2\sigma^2}{N})^2}{\frac{2\sigma^2}{N}(2P + \frac{2\sigma^2}{N})}$$

When no signal is present or  $P = 0$ , the measure is one.

For example, if  $P \gg 4\sigma^2/N$  or  $NP \gg 4\sigma^2$ ,

$$\frac{E(\hat{P})^2}{\text{var}(\hat{P})} = \frac{P^2}{4P\sigma^2/N} = \frac{P}{\frac{4\sigma^2}{N}} \gg 1$$

and the sinusoid will be easily detectable.

$$7) (\underline{H}^T \underline{H})_{ij} = \sum_{n=1}^N u[n-i] u[n-j]$$

For large  $N$  this is  $\sum_{n=-\infty}^{\infty} u[n-i] u[n-j]$  since for  $u[n]=0$ ,  $n < 0$  and  $n > N-1$ , this will add only about  $2p$  additional terms. If  $N \gg p$ , these terms will be negligible.

Now assume  $i \geq j$  and let  $m = n - i$

$$\begin{aligned} (\underline{H}^T \underline{H})_{ij} &= \sum_{m=-\infty}^{\infty} u[m] u[m+i-j] \\ &= \sum_{m=0}^{N-1-(i-j)} u[m] u[m+i-j] \end{aligned}$$

Similarly for  $i < j$

$$\begin{aligned} (\underline{H}^T \underline{H})_{ij} &= \sum_{m=0}^{N-1-(j-i)} u[m] u[m+j-i] \\ \Rightarrow (\underline{H}^T \underline{H})_{ij} &= \sum_{m=0}^{N-1-(i-j)} u[m] u[m+i-j] \end{aligned}$$

$$8) \quad x[n] = \sum_{\ell=0}^{\infty} h[\ell] u[n-\ell]$$

$$\begin{aligned} r_{ux}[k] &= E(u[n]x[n+k]) \\ &= E(u[n] \sum_{\ell=0}^{\infty} h[\ell] u[n+k-\ell]) \\ &= \sum_{\ell=0}^{\infty} h[\ell] E(u[n]u[n+k-\ell]) \\ &= \sum_{\ell=0}^{\infty} h[\ell] r_{uu}[k-\ell] \\ &= h[k] * r_{uu}[k] \end{aligned}$$

$$\Rightarrow P_{ux}(f) = H(f) P_{uu}(f)$$

If  $P_{uu}(f) = \sigma^2$ , then  $H(f) = P_{ux}(f)/\sigma^2$   
or

$$\hat{H}(f) = \frac{\hat{P}_{ux}(f)}{r_{uu}[0]}$$

If  $P_{uu}(f) = 0$  over a band of frequencies, it would be impossible to estimate  $H(f)$  over that band. This is because there would be no power over that band at the output, leading to  $\hat{P}_{ux}(f) = 0$ , independent of  $H(f)$ . The TDL estimator of (4.23) is just the inverse Fourier transform of  $\hat{H}(f)$ .

$$9) P(\underline{x}; \underline{\theta}) = \frac{1}{(2\pi)^{n/2} \det(\underline{\Sigma})^{1/2}} e^{-\frac{1}{2} (\underline{x} - \underline{H}\underline{\theta})^T \underline{\Sigma}^{-1} (\underline{x} - \underline{H}\underline{\theta})}$$

$$\begin{aligned} \frac{\partial \ln P}{\partial \underline{\theta}} &= -\frac{1}{2} \frac{\partial}{\partial \underline{\theta}} \left[ \underline{x}^T \underline{\Sigma}^{-1} \underline{x} - 2 \underline{\theta}^T \underline{H}^T \underline{\Sigma}^{-1} \underline{x} + \underline{\theta}^T \underline{H}^T \underline{\Sigma}^{-1} \underline{H} \underline{\theta} \right] \\ &= -\frac{1}{2} \left[ -2 \underline{H}^T \underline{\Sigma}^{-1} \underline{x} + 2 \underline{H}^T \underline{\Sigma}^{-1} \underline{H} \underline{\theta} \right] \text{ using (4.3)} \\ &= \underline{H}^T \underline{\Sigma}^{-1} \underline{x} - \underline{H}^T \underline{\Sigma}^{-1} \underline{H} \underline{\theta} \\ &= \underbrace{\underline{H}^T \underline{\Sigma}^{-1} \underline{H}}_{I(\underline{\theta})} \underbrace{((\underline{H}^T \underline{\Sigma}^{-1} \underline{H})^{-1} \underline{H}^T \underline{\Sigma}^{-1} \underline{x} - \underline{\theta})}_{\hat{\underline{\theta}}} \end{aligned}$$

$\therefore \hat{\underline{\theta}} = (\underline{H}^T \underline{\Sigma}^{-1} \underline{H})^{-1} \underline{H}^T \underline{\Sigma}^{-1} \underline{x}$  is MVU estimator (and efficient)

$$C_{\hat{\underline{\theta}}} = (\underline{H}^T \underline{\Sigma}^{-1} \underline{H})^{-1}$$

$$10) \underline{\Sigma}^{-1} = \underline{D}^T \underline{D}$$

$$\text{Since } \underline{\Sigma} = \text{diag}(\sigma_0^2, \sigma_1^2, \dots, \sigma_{N-1}^2)$$

$$\underline{\Sigma}^{-1} = \text{diag}(\frac{1}{\sigma_0^2}, \frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_{N-1}^2})$$

$$\Rightarrow \underline{D} = \text{diag}(\frac{1}{\sigma_0}, \frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_{N-1}})$$

$$\hat{\underline{A}} = \sum_{n=0}^{N-1} d_n \underline{x}' \underline{E}_n$$

$$\text{where } d_n = \frac{(\underline{D}^{-1})_{nn}}{\underline{1}' \underline{D}^T \underline{D}^{-1}} = \frac{\frac{1}{\sigma_n}}{\sum_{m=0}^{N-1} \frac{1}{\sigma_m^2}}$$

Since  $\underline{\Sigma}$  is already diagonal or the

components of  $\underline{w}$  are uncorrelated, we need only form  $\underline{x}' = \underline{D}\underline{x}$  or  $x'(n) = x(n)/\sigma_n$ , so that the variances are one. Then, we "average" the  $x'(n)$  samples. Actually, we weight the samples since the DC level has been changed to a non-DC signal due to the prewhitening stage.

If all  $\sigma_n^2 = 0$ , say  $\sigma_m^2$ , then we cannot prewhiten the data. In this case, however, as  $\sigma_m^2 \rightarrow 0$ ,  $d_n \rightarrow 0$  for  $n \neq m$  and  $d_n \rightarrow \sigma_m$  for  $n = m$ . Thus,  $\hat{A} \rightarrow d_m x'(m) = x(m)$ , as expected.

$$(1) \quad \hat{A} = (\underline{H}^T \underline{C}^{-1} \underline{H})^{-1} \underline{H}^T \underline{C}^{-1} \underline{x} \quad \text{var}(\hat{A}) = (\underline{H}^T \underline{C}^{-1} \underline{H})^{-1}$$

$$\underline{C} = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \quad \underline{H} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\underline{C}^{-1} = \frac{1}{\sigma^2(1-\rho^2)} \begin{pmatrix} 1-\rho & \rho \\ -\rho & 1 \end{pmatrix}$$

$$\underline{H}^T \underline{C}^{-1} \underline{H} = \frac{2-2\rho}{\sigma^2(1-\rho^2)} = \frac{2(1-\rho)}{\sigma^2(1-\rho^2)} = \frac{2}{\sigma^2(1+\rho)}$$

$$\begin{aligned} \underline{H}^T \underline{C}^{-1} \underline{x} &= \frac{1}{\sigma^2(1-\rho^2)} \begin{pmatrix} 1-\rho & \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} x(0) \\ x(1) \end{pmatrix} \\ &= \frac{x(0)(1-\rho) + x(1)(1-\rho)}{\sigma^2(1-\rho^2)} \end{aligned}$$

$$= \frac{x(0) + x(1)}{\sigma^2(1+\rho)}$$

$$\hat{A} = \frac{\sigma^2(1+\rho)}{2} \frac{x(0) + x(1)}{\sigma^2(1+\rho)} = \frac{1}{2}(x(0) + x(1))$$

$$\text{var}(\hat{A}) = \frac{\sigma^2(1+\rho)}{2}$$

We don't need  $\text{prwhtness}$  here because  $H$  is an eigenvector of  $\Sigma$ . Hence,

$$(H^T \Sigma^{-1} H)^{-1} H^T \Sigma^{-1} = (H^T \frac{1}{\lambda} H)^{-1} \frac{1}{\lambda} H^T$$

$$= (H^T H)^{-1} H^T$$

As  $\rho \rightarrow 1$ ,  $\text{var}(\hat{A}) \rightarrow \sigma^2$

As  $\rho \rightarrow -1$ ,  $\text{var}(\hat{A}) \rightarrow 0$ .

See Prob 3.9 for explanation.

(2) If  $\underline{x} = H\underline{\theta} + \underline{w}$ ,  $\underline{x}' = A(H^T H)^{-1} H^T \underline{x}$

$$\underline{x}' = A(H^T H)^{-1} H^T H \underline{\theta} + A(H^T H)^{-1} H^T \underline{w}$$

$$\begin{matrix} \uparrow \\ rx1 \end{matrix} = \begin{matrix} A \underline{\theta} \\ \uparrow \\ rx1 \end{matrix} + \underline{w}' = \begin{matrix} H' \underline{\alpha} \\ \uparrow \\ rx1 \end{matrix} + \underline{w}' \quad \text{where } H' = I$$

$$\begin{aligned} \underline{\epsilon}' &= E(\underline{w}' H'^T) = E(A(H^T H)^{-1} H^T \underline{w} \underline{w}^T H (H^T H)^{-1} A^T) \\ &= \sigma^2 A(H^T H)^{-1} A^T \end{aligned}$$

Since  $\underline{A}$  is full rank,  $\Sigma'$  is positive definite and  $\Sigma^{-1}$  exists.

$$\Rightarrow \hat{\underline{\theta}} = (\underline{H}^T \Sigma'^{-1} \underline{H})^{-1} \underline{H}^T \Sigma'^{-1} \underline{x}' \\ = \Sigma' \underline{\varepsilon}'^{-1} \underline{x}' = \underline{A} (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x} = \underline{A} \hat{\underline{\theta}}$$

$$13) E(\hat{\underline{\theta}}) = E((\underline{H}^T \underline{A})^{-1} \underline{H}^T (\underline{H} \underline{\theta} + \underline{w})) \\ = \underline{\theta} + E((\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{w}) \\ = \underline{\theta} + E((\underline{H}^T \underline{H})^{-1} \underline{H}^T) E(\underline{w}) \\ = \underline{\theta} \quad \text{since } E(\underline{w}) = \underline{0}.$$

$$\Sigma_{\hat{\underline{\theta}}} = E((\hat{\underline{\theta}} - \underline{\theta})(\hat{\underline{\theta}} - \underline{\theta})^T) \\ = E[((\underline{H}^T \underline{H})^{-1} \underline{H}^T (\underline{x} - \underline{H} \underline{\theta})) \\ ((\underline{H}^T \underline{H})^{-1} \underline{H}^T (\underline{x} - \underline{H} \underline{\theta}))^T] \\ = E[(\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{w} \underline{w}^T \underline{H} (\underline{H}^T \underline{H})^{-1}] \\ = E_{H,W} E_{\underline{w}} [(\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{w} \underline{w}^T \underline{H} (\underline{H}^T \underline{H})^{-1}] \\ = E_{H,W} [(\underline{H}^T \underline{H})^{-1} \underline{H}^T \sigma^2 \underline{I} \underline{H} (\underline{H}^T \underline{H})^{-1}] \\ = E_{H,W} [\sigma^2 (\underline{H}^T \underline{H})^{-1}] = \sigma^2 E_H [(\underline{H}^T \underline{H})^{-1}]$$

Since  $\underline{H}$  and  $\underline{w}$  are independent. If  $\underline{H}$  and  $\underline{w}$  are not independent  $\hat{\underline{\theta}}$  may be biased.

$$14) \underline{H} = ! \text{ with probability } 1-\epsilon$$

$$\underline{H} = \left[ \underbrace{1 \dots 1}_{M} \underbrace{0 \dots 0}_{N-M} \right]^T \text{ with probability } \epsilon$$

$$\hat{A} = (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x}$$

$$= \frac{1}{N} \sum_{n=0}^{N-1} \underline{x}(n) \quad \text{no fade}$$

$$\frac{1}{M} \sum_{n=0}^{M-1} \underline{x}(n) \quad \text{fade}$$

$$\text{var}(\hat{A}) = \sigma^2 E_H [(\underline{H}^T \underline{H})^{-1}]$$

$$= \sigma^2 \left[ \frac{1}{N} (1-\epsilon) + \frac{\epsilon}{M} \right]$$

$$= \frac{\sigma^2}{N} \left[ 1 - \epsilon + \frac{\epsilon}{M} \right]$$

$$= \frac{\sigma^2}{N} \left[ 1 + \left( \frac{N}{M} - 1 \right) \epsilon \right] > \frac{\sigma^2}{N}$$

Clearly, the variances are the same only if  $M = N$  or  $\epsilon = 0$ . Otherwise, it is increased.

## Chapter 5

$$1) \quad p(\underline{x} | T(\underline{x}) = T_0; \sigma^2) = \frac{p(\underline{x}; \sigma^2) \delta(T(\underline{x}) - T_0)}{p(T(\underline{x}) = T_0; \sigma^2)}$$

$$= \frac{\frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} \sum_n x^2(n)}}{p\left(\sum_n x^2(n) = T_0; \sigma^2\right)}$$

But  $\sum_n x^2(n) = 5\sigma^2$

$$p\left(\sum_n x^2(n)\right) = \frac{1}{\sigma^2} p_s\left(\frac{\sum_n x^2(n)}{\sigma^2}\right)$$

$$p(\underline{x} | T(\underline{x}) = T_0; \sigma^2) = \frac{\frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} T_0}}{\frac{1}{\sigma^2} \frac{1}{2^{N/2} \Gamma(N/2)} e^{-T_0/2\sigma^2} \left(\frac{T_0}{\sigma^2}\right)^{\frac{N}{2}-1}}$$

$$= \frac{\frac{1}{\pi^{N/2}} \delta(T(\underline{x}) - T_0)}{\frac{1}{\Gamma(N/2)} T_0^{\frac{N}{2}-1}}$$

$$2) \quad p(\underline{x}; \sigma^2) = \prod_{n=0}^{N-1} \frac{x(n)}{\sigma^2} e^{-\frac{1}{2} \sum_n x^2(n)/\sigma^2} \quad \text{all } x(n) > 0$$

$$= u(\min x(n)) \prod_{n=0}^{N-1} x(n) \underbrace{\frac{1}{\sigma^{2N}} e^{-\frac{1}{2} \sum_n \frac{x^2(n)}{\sigma^2}}}_{h(\underline{x}, g(T(\underline{x}), \sigma^2))}$$

where  $u(x)$  is the unit step function

$T(\underline{x}) = \sum_{n=0}^{N-1} x^2(n)$  is a sufficient statistic

$$3) p(\underline{x}; \lambda) = \prod_{n=0}^{N-1} \lambda e^{-\lambda x(n)} \quad \text{all } x(n) > 0$$

$$= \underbrace{\lambda^N e^{-\lambda \sum_n x(n)}}_{g(T(\underline{x}), \lambda)} \cdot \underbrace{u(\min x(n))}_{h(\underline{x})}$$

$T(\underline{x}) = \sum_{n=0}^{N-1} x(n)$  is a sufficient statistic

$$4) p(\underline{x}; \theta) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2\theta}(x(n)-\theta)^2}$$

$$= \frac{1}{(2\pi\theta)^{N/2}} e^{-\frac{1}{2\theta} \sum_n (x(n)-\theta)^2}$$

But  $\sum_n (x(n)-\theta)^2 = \sum_n x^2(n) - 2\theta \sum_n x(n) + N\theta^2$

$$p(\underline{x}; \theta) = \underbrace{\frac{1}{(2\pi\theta)^{N/2}} e^{-\frac{1}{2\theta} \sum_n x^2(n) - \frac{1}{2} N\theta}}_{g(T(\underline{x}), \theta)} \underbrace{e^{\sum_n x(n)}}_{h(\underline{x})}$$

$T(\underline{x}) = \sum_{n=0}^{N-1} x^2(n)$  is a sufficient statistic

$$5) p(x(n); \theta) = \frac{1}{2\theta} (u(x(n+\theta)) - u(x(n-\theta)))$$

$$p(\underline{x}; \theta) = \frac{1}{(2\theta)^N} \prod_{n=0}^{N-1} [u(x(n+\theta)) - u(x(n-\theta))]$$

But the product is zero unless  $-\theta \leq x(n) \leq \theta$   
 for all  $x(n)$  or  $\min x(n) \geq -\theta$ ,  $\max x(n) \leq \theta$

or  $\max |x[n]| \leq \theta$  so that

$$p(\underline{x}; \theta) = \underbrace{\frac{1}{(2\theta)^N} u(\theta - \max |x[n]|)}_{g(T(\underline{x}), \theta)} \cdot \underbrace{1}_{h(\underline{x})}$$

and  $T(\underline{x}) = \max |x[n]|$  is the sufficient statistic.

$$b) p(\underline{x}; \sigma^2) = \underbrace{\frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2}}_{g(T(\underline{x}), \sigma^2)} \cdot \underbrace{1}_{h(\underline{x})}$$

$T(\underline{x}) = \sum_{n=0}^{N-1} (x[n] - A)^2$  is a sufficient statistic

To make it unbiased, divide by  $N$ . Thus,

$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - A)^2$  is the MVU estimator.

$$7) p(\underline{x}; f_0) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - \cos 2\pi f_0 n)^2}$$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} [\sum_n x^2(n) - 2 \sum_n x(n) \cos 2\pi f_0 n + \sum_n \cos^2 2\pi f_0 n]}$$

Because of the  $\sum_n x(n) \cos 2\pi f_0 n$  term, which cannot be separated into a statistic, there does not appear to be a sufficient statistic. Note that  $\sum_n x(n) \cos 2\pi f_0 n$  is not a statistic

since it depends on  $f_0$ .

$$8) P(\underline{x}; r) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - r^n)^2}$$

$$\text{But } \sum_n (x[n] - r^n)^2 = \sum_n x^2[n] - 2 \sum_n x[n] r^n + \sum_n r^{2n}$$

Again, the term  $\sum_n x[n] r^n$  cannot be separated into a single sufficient statistic and a function of  $r$ .

$$9) P_r \{ \underline{x} \} = \theta^{x[0]} (1-\theta)^{1-x[0]} \quad x[0] = 0, 1$$

$$\begin{aligned} P_r \{ \underline{x} \} &= \prod_{n=0}^{N-1} \theta^{x[n]} (1-\theta)^{1-x[n]} \\ &= \underbrace{\theta^{\sum_n x[n]} (1-\theta)^{N - \sum_n x[n]}}_{g(\tau(\underline{x}), \theta)} \cdot \underbrace{1}_{h(\underline{x})} \end{aligned}$$

where  $\tau(\underline{x}) = \sum_{n=0}^{N-1} x[n]$  is a sufficient statistic. To make it unbiased divide by  $N$  so that

$$\hat{\theta} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \text{ is the MVU estimator.}$$

10) At high SNR we ignore the noise so that

$$T_1(\underline{x}) \approx \sum_n A \cos(2\pi f_0 n + \phi) \cos 2\pi f_0 n$$

$$= \sum_n \frac{A}{2} [\cos \phi + \cos(4\pi f_{on} + \phi)]$$

$$\hat{\phi} \approx NA/2 \cos \phi \quad \text{using results of Prob 3.7}$$

Also,

$$T_2(\underline{x}) \approx \sum_n A \cos(2\pi f_{on} + \phi) \sin 2\pi f_{on}$$

$$= \sum_n \frac{A}{2} [-\sin \phi + \sin(4\pi f_{on} + \phi)]$$

$$\hat{\phi} \approx -\frac{NA}{2} \sin \phi$$

$$\text{Thus, } \hat{\phi} = -\arctan \frac{T_2(\underline{x})}{T_1(\underline{x})} = -\arctan \frac{-\frac{NA}{2} \sin \phi}{\frac{NA}{2} \cos \phi} \\ = \phi$$

This is not the MVU estimator since it is not unbiased. To see this note that even if we could assume  $E(T_1(\underline{x})) = \frac{NA}{2} \cos \phi$  and

$$E(T_2(\underline{x})) = -\frac{NA}{2} \sin \phi$$

(which will not be true in general - only at high SNR), it is not true that

$$E(\hat{\phi}) = E\left(-\arctan \frac{T_2(\underline{x})}{T_1(\underline{x})}\right)$$

$$= -\arctan \frac{E[T_2(\underline{x})]}{E[T_1(\underline{x})]}$$

↑  
incorrect

$$(1) \quad \theta = 2A + 1 \Rightarrow A = \frac{\theta - 1}{2}$$

$$\begin{aligned}
 p(\underline{x}; A) &= \frac{1}{(2\pi\sigma^2/N)^{1/2}} e^{-\frac{1}{2\sigma^2} \sum_n (x(n)-A)^2} \\
 p'(\underline{x}; \theta) &= \frac{1}{(2\pi\sigma^2/N)^{1/2}} e^{-\frac{1}{2\sigma^2} \sum_n (x(n) - \theta \frac{-1}{2})^2} \\
 &= \frac{1}{(2\pi\sigma^2/N)^{1/2}} e^{-\frac{1}{2\sigma^2} \left[ \sum_n x^2(n) - (\theta - 1) \sum_n x(n) + N \left( \frac{\theta - 1}{2} \right)^2 \right]} \\
 &= \underbrace{\frac{1}{(2\pi\sigma^2/N)^{1/2}} e^{-\frac{1}{2\sigma^2} \left[ N \left( \frac{\theta - 1}{2} \right)^2 - (\theta - 1) \sum_n x(n) \right]}}_{g(\tau(\underline{x}), \theta)} e^{-\frac{1}{2\sigma^2} \sum_n x^2(n)} \underbrace{e^{-\frac{1}{2\sigma^2} \sum_n x(n)}}_{h(\underline{x})}
 \end{aligned}$$

Thus  $\sum_{n=0}^{N-1} x(n)$  is a sufficient statistic for  $\theta$ .

To make it unbiased

$$E\left(\sum_n x(n)\right) = NA = N\left(\frac{\theta - 1}{2}\right) = N \frac{\theta}{2} - \frac{N}{2}$$

$$\text{Let } \hat{\theta} = \frac{2}{N} \left( \sum_n x(n) + \frac{N}{2} \right) = \frac{2}{N} \sum_n x(n) + 1$$

This is the MVU estimator. Note that

$\hat{\theta} = 2\hat{A} + 1$ , where  $\hat{A}$  is MVU estimator for  $A$ .

For  $\theta = A^3$  the sufficient statistic is again  $\sum_n x(n)$ . To make it unbiased we need a function  $g$  such that

$$E(g(\sum_n x(n))) = A^3$$

or  $E(h(\bar{x})) = A^3$  where  $\bar{x} \sim N(A, \sigma^2/N)$   
since  $\bar{x}$  is also a sufficient statistic for  $\theta$ .

Examining  $\bar{x}^3$  we have

$$\begin{aligned} E((\bar{x}-A)^3) &= E(\bar{x}^3) - 3AE(\bar{x}^2) + 3A^2E(\bar{x}) - A^3 \\ &= 0 \end{aligned}$$

since all odd-order moments are zero.

$$\begin{aligned} E(\bar{x}^3) &= 3A(A^2 + \sigma^2/N) - 3A^3 + A^3 \\ &= A^3 + 3A\sigma^2/N \end{aligned}$$

Try  $\hat{\theta} = \bar{x}^3 - 3\bar{x}\sigma^2/N$ . This is unbiased  
and is a function of the sufficient statistic.  
Thus, it is the MVU estimator.

$$12) g_1(u) = \frac{1}{N}u \Rightarrow E\left[\frac{1}{N}\sum_n x(n)\right] = A$$

$$g_2(u) = \frac{1}{N}u^{1/3} \Rightarrow E\left[\frac{1}{N}((\sum_n x(n))^3)^{1/3}\right] = A$$

Also, note that  $T_2 = T_1^3$ , which is a one-to-one transformation. For  $T_3$  there is no function that would make it unbiased. Also,  $T_3$  is not a one-to-one transformation of  $T_1$ .

$$\begin{aligned} 13) p(\underline{x}; \theta) &= e^{-\sum_n (\min x(n) - \theta)} u(\min x(n) - \theta) \\ &= \underbrace{e^{-\sum_n x(n)}}_{h(\underline{x})} \underbrace{e^{N\theta} u(\min x(n) - \theta)}_{g(T(\underline{x}), \theta)} \end{aligned}$$

where  $T(\underline{x}) = \min x(n)$  is the sufficient statistic. To find the MVR we proceed as in Example 5.8.

$$\begin{aligned} \Pr\{T \leq z\} &= 1 - \Pr\{T \geq z\} \\ &= 1 - \Pr\{x(0) \geq z, \dots, x(N-1) \geq z\} \\ &= 1 - \prod_{n=0}^{N-1} \Pr\{x(n) \geq z\} \\ &= 1 - \Pr^N\{x(n) \geq z\} \end{aligned}$$

$$\begin{aligned} p_T(z) &= \frac{d \Pr\{T \leq z\}}{dz} = -N \Pr\{x(n) \geq z\}^{N-1} \\ &\quad \cdot \frac{d \Pr\{x(n) \geq z\}}{dz} \end{aligned}$$

$$\begin{aligned} \text{But } \frac{d \Pr\{x(n) \geq z\}}{dz} &= \frac{d [1 - \Pr\{x(n) \leq z\}]}{dz} \\ &= - \frac{d \Pr\{x(n) \leq z\}}{dz} = -e^{-(z-\theta)} & z > \theta \\ &= 0 & z < \theta \end{aligned}$$

and

$$\Pr\{x(N) \geq z\} = \begin{cases} 1 - \int_0^z e^{-(x-\theta)} dx & \text{for } z > \theta \\ 1 & \text{for } z \leq \theta \end{cases}$$

$$\begin{aligned} \text{For } z > \theta &= 1 + e^\theta e^{-z} / 0 = 1 + e^{-(z-\theta)} - 1 \\ &= e^{-(z-\theta)} \end{aligned}$$

$$p_T(z) = \begin{cases} 0 & z < \theta \\ -N [e^{-(z-\theta)}]^{N-1} [-e^{-(z-\theta)}] & z > \theta \\ N e^{-N(z-\theta)} & z > \theta \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} E(T) &= \int_0^\infty z N e^{-N(z-\theta)} dz \\ &= N e^{N\theta} \int_0^\infty z e^{-Nz} dz \\ &= N e^{N\theta} \left[ -\frac{z}{N} e^{-Nz} - \frac{1}{N^2} e^{-Nz} \right]_0^\infty \\ &= N e^{N\theta} \left( \frac{\theta}{N} e^{-N\theta} + \frac{1}{N^2} e^{-N\theta} \right) = \theta + \frac{1}{N} \end{aligned}$$

$\Rightarrow \hat{\theta} = T - \frac{1}{N} = \min x_i \{n\} - \frac{1}{N}$  is the MLE estimator.

$$\begin{aligned} 14) \text{ a) } p(x; \mu) &= \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}x^2 + x\mu - \frac{1}{2}\mu^2} \\ &= e^{x\mu - \frac{1}{2}x^2 + (-\frac{1}{2}\mu^2 + \ln 1/\sqrt{2\pi})} \\ &\quad \uparrow \quad \uparrow \quad \uparrow \\ &\quad A(\mu) \quad B(x) \quad C(x) \quad D(\mu) \end{aligned}$$

$$\text{b) } p(x; \sigma^2) = \frac{x}{\sigma^2} e^{-\frac{1}{2}x^2/\sigma^2} u(x)$$

$$= e^{\ln x/\sigma^2 - \frac{1}{2}x^2/\sigma^2 + \ln u(x)}$$

$$= e^{\underbrace{-\frac{1}{2}x^2/\sigma^2}_{A(\sigma^2)} + \underbrace{\ln x u(x)}_{B(x)} - \underbrace{\ln \sigma^2}_{D(\sigma^2)}}$$

$$\text{c) } p(x; \lambda) = d e^{-\lambda x} u(x)$$

$$= e^{\underbrace{-\lambda x}_{A(\lambda)} + \underbrace{\ln u(x)}_{B(\lambda)} + \underbrace{\ln \lambda}_{D(\lambda)}}$$

$$A(\lambda) B(\lambda) C(x) D(\lambda)$$

$$\text{15') } p(\underline{x}; \theta) = \prod_{n=0}^{N-1} e^{A(\theta) B(x(n)) + C(x(n)) + D(\theta)}$$

$$= e^{A(\theta) \sum_n B(x(n)) + \sum_n C(x(n)) + N D(\theta)}$$

$$= \underbrace{e^{A(\theta) \sum_n B(x(n)) + N D(\theta)}}_{g(T(\underline{x}), \theta)} \underbrace{e^{\sum_n C(x(n))}}_{h(\underline{x})}$$

where  $T(\underline{x}) = \sum_{n=0}^{N-1} B(x(n))$

a)  $B(x) = x \Rightarrow T(\underline{x}) = \sum_n x(n)$

b)  $B(x) = x^2 \Rightarrow T(\underline{x}) = \sum_n x^2(n)$

c)  $B(x) = x \Rightarrow T(\underline{x}) = \sum_n x(n)$

Need only make  $T(\underline{x})$  unbiased

a)  $\hat{u} = \frac{1}{N} \sum_n x(n)$

b)  $E(x^2) = \int_0^\infty \frac{x^2}{\sigma^2} e^{-\frac{1}{2} x^2/\sigma^2} dx$

$$= 2\sigma^2$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{2N} \sum_n x^2 - \bar{x}^2$$

$$(c) E(x) = \int_0^\infty x \lambda e^{-\lambda x} dx = 1/\lambda$$

It is not obvious how to make  $\tau$  unbiased for this PDF. However, if we reparameterize the PDF by  $\theta = 1/\lambda$ , the MVU of  $\theta$  is easily found.

$$(b) p(x; \theta) = \frac{1}{(2\pi)^N \det \Sigma(\underline{x})} e^{-\frac{1}{2}(\underline{x}-\underline{m})^T \Sigma^{-1}(\underline{x}-\underline{m})}$$

$$\text{If we have } \Sigma = \begin{pmatrix} a & \underline{b}^T \\ \underline{b} & \underline{\underline{\Sigma}} \end{pmatrix} \quad a = N-1 + \sigma^2 \\ \underline{b} = -\underline{1}$$

$$\begin{aligned} \det(\Sigma) &= \det(\underline{\underline{\Sigma}}) \det(a - \underline{b}^T \underline{\underline{\Sigma}}^{-1} \underline{b}) \\ &= a - \underline{b}^T \underline{b} = N-1 + \sigma^2 - (-1)^T (-1) \\ &= N-1 + \sigma^2 - (N-1) = \sigma^2 \end{aligned}$$

$$\begin{aligned} \Sigma^{-1} &= \begin{bmatrix} (a - \underline{b}^T \underline{b})^{-1} & - (a - \underline{b}^T \underline{b})^{-1} \underline{b}^T \\ - \underline{b} (a - \underline{b}^T \underline{b})^{-1} & (\underline{\underline{\Sigma}} - \frac{\underline{b} \underline{b}^T}{a})^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\sigma^2} & \frac{1}{\sigma^2} \underline{1}^T \\ \frac{1}{\sigma^2} \underline{1} & (\underline{\underline{\Sigma}} - \frac{\underline{1} \underline{1}^T}{N-1+\sigma^2})^{-1} \end{bmatrix} \end{aligned}$$

$$\text{But } (\underline{\underline{\Sigma}} - \frac{\underline{1} \underline{1}^T}{N-1+\sigma^2})^{-1} = \underline{\underline{\Sigma}} + \frac{\underline{1} \underline{1}^T}{N-1+\sigma^2} \\ 1 - \frac{\underline{1}^T}{N-1+\sigma^2}$$

$$= \underline{\mathbf{I}} + \frac{\underline{\mathbf{I}}\underline{\mathbf{I}}^T}{\sigma^2}$$

$$\underline{\mathbf{C}}^{-1} = \frac{1}{\sigma^2} \begin{bmatrix} \underline{\mathbf{I}} & \underline{\mathbf{I}}^T \\ \underline{\mathbf{I}} & \sigma^2 \underline{\mathbf{I}} + \underline{\mathbf{I}}\underline{\mathbf{I}}^T \end{bmatrix}$$

$$(\underline{\mathbf{x}} - \underline{\mathbf{u}})^T \underline{\mathbf{C}}^{-1} (\underline{\mathbf{x}} - \underline{\mathbf{u}}) =$$

$$\frac{1}{\sigma^2} \left[ \underbrace{\mathbf{x}(0) - N\underline{\mathbf{u}} \quad \mathbf{x}(1) \dots \mathbf{x}(N-1)}_{\underline{\mathbf{x}}'}^T \right] \begin{bmatrix} \underline{\mathbf{I}} & \underline{\mathbf{I}}^T \\ \underline{\mathbf{I}} & \sigma^2 \underline{\mathbf{I}} + \underline{\mathbf{I}}\underline{\mathbf{I}}^T \end{bmatrix} \begin{pmatrix} \mathbf{x}(0) - N\underline{\mathbf{u}} \\ \underline{\mathbf{x}}' \end{pmatrix}$$

$$= \frac{1}{\sigma^2} \left( (\mathbf{x}(0) - N\underline{\mathbf{u}})^2 + (\mathbf{x}(0) - N\underline{\mathbf{u}}) \underline{\mathbf{I}}^T \underline{\mathbf{x}}' + (\mathbf{x}(0) - N\underline{\mathbf{u}}) \underline{\mathbf{x}}'^T \underline{\mathbf{I}} + \sigma^2 \underline{\mathbf{x}}'^T \underline{\mathbf{x}}' + (\underline{\mathbf{x}}'^T \underline{\mathbf{I}})^2 \right)$$

$$= \frac{1}{\sigma^2} \left( (\mathbf{x}(0) - N\underline{\mathbf{u}} + \underline{\mathbf{I}}^T \underline{\mathbf{x}}')^2 + \sigma^2 \underline{\mathbf{x}}'^T \underline{\mathbf{x}}' \right)$$

$$= \frac{1}{\sigma^2} \left[ \left( \sum_{n=0}^{N-1} \mathbf{x}(n) - N\underline{\mathbf{u}} \right)^2 + \sigma^2 \sum_{n=1}^{N-1} \mathbf{x}^2(n) \right]$$

$$= \frac{N^2}{\sigma^2} (\bar{\mathbf{x}} - \underline{\mathbf{u}})^2 + \sum_{n=1}^{N-1} \mathbf{x}^2(n)$$

$$p(\underline{\mathbf{x}}; \underline{\theta}) = \underbrace{\frac{1}{(2\pi)^{N/2}\sigma}}_{g(\tau(\underline{\mathbf{x}}), \underline{\theta})} e^{-\frac{N^2}{2\sigma^2}(\bar{\mathbf{x}} - \underline{\mathbf{u}})^2} \underbrace{e^{-\frac{1}{2} \sum_{n=1}^{N-1} \mathbf{x}^2(n)}}_{h(\underline{\mathbf{x}})}$$

$\Rightarrow \bar{\mathbf{x}}$  is a sufficient statistic for  $\underline{\theta}$ .

$$17) \text{ as } p(\underline{x}; A) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x(n) - A \cos 2\pi f_0 n)^2}$$

$$\text{But } \sum_n (x(n) - A \cos 2\pi f_0 n)^2 =$$

$$\sum_n x^2(n) - 2A \sum_n x(n) \cos 2\pi f_0 n + A^2 \sum_n \cos^2 2\pi f_0 n$$

$$p(\underline{x}; A) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} (A^2 \sum_n \cos^2 2\pi f_0 n - 2A \sum_n x(n) \cos 2\pi f_0 n)}$$

$\underbrace{\cdot e^{-\frac{1}{2\sigma^2} \sum_n x^2(n)}}$

where  $T(\underline{x}) = \sum_{n=0}^{N-1} x(n) \cos 2\pi f_0 n$  is the sufficient statistic

$$E(T(\underline{x})) = \sum_{n=0}^{N-1} A \cos^2 2\pi f_0 n$$

$$\Rightarrow \hat{A} = \frac{\sum_{n=0}^{N-1} x(n) \cos 2\pi f_0 n}{\sum_{n=0}^{N-1} \cos^2 2\pi f_0 n}$$

b) Let  $\underline{\sigma} = [\hat{A} \ \sigma_2]$

From part a we have

$$T(\underline{x}) = \begin{bmatrix} \sum_n x(n) \cos 2\pi f_0 n \\ \sum_n x^2(n) \end{bmatrix}$$

is a sufficient statistic

To make  $T_1(x)$  unbiased let

$$\hat{A} = \frac{\sum_n x(n) \cos 2\pi f_0 n}{\sum_n \cos^2 2\pi f_0 n} \quad \text{as before.}$$

To make  $T_2(x)$  unbiased :

$$\begin{aligned} E(T_2(x)) &= \sum_n E(x^2(n)) \\ &= \sum_n E((A \cos 2\pi f_0 n + W(n))^2) \\ &= \sum_n A^2 \cos^2 2\pi f_0 n + N \sigma^2 \end{aligned}$$

From Example 5.11 we expect that we will have to subtract out the squared mean to generate an unbiased estimator of  $\sigma^2$ .

But

$$\begin{aligned} E(\hat{A}^2) &= \frac{E\left[\left(\sum_n x(n) \cos 2\pi f_0 n\right)^2\right]}{\left(\sum_n \cos^2 2\pi f_0 n\right)^2} \\ &= \frac{\sum_m \sum_n E(x(m)x(n)) \cos 2\pi f_0 m \cos 2\pi f_0 n}{\left(\sum_n \cos^2 2\pi f_0 n\right)^2} \\ &= \frac{\sum_{m,n} E(A^2 \cos 2\pi f_0 m \cos 2\pi f_0 n + \ell^2 \delta_{mn}) \cdot \frac{(\cos 2\pi f_0 m)}{(\cos 2\pi f_0 n)}}{\left(\sum_n \cos^2 2\pi f_0 n\right)^2} \end{aligned}$$

$$= \frac{A^2 \left( \sum_n \cos^2 2\pi f_0 n \right)^2 + \sigma^2 \sum_n \cos^2 2\pi f_0 n}{\left( \sum_n \cos^2 2\pi f_0 n \right)^2}$$

$$= A^2 + \frac{\sigma^2}{\sum_n \cos^2 2\pi f_0 n}$$

so that if  $T_2'(\underline{x}) = T_2(\underline{x}) - \sum_n \cos^2 2\pi f_0 n \hat{A}^2$

$$E(T_2'(\underline{x})) = A^2 \sum_n \cos^2 2\pi f_0 n + N \sigma^2$$

$$- A^2 \sum_n \cos^2 2\pi f_0 n \cdot \sigma^2$$

$$= (N-1) \sigma^2$$

$$\Rightarrow \text{Let } T_2''(\underline{x}) = \frac{1}{N-1} T_1'(\underline{x})$$

$$= \frac{1}{N-1} \left[ \sum_n x^2(n) - \sum_n \cos^2 2\pi f_0 n \hat{A}^2 \right]$$

$$\therefore \hat{\underline{\theta}} = \begin{pmatrix} \hat{A} \\ \hat{\sigma}^2 \end{pmatrix}$$

$$= \begin{bmatrix} \sum_{n=0}^{N-1} x(n) \cos 2\pi f_0 n \\ \frac{1}{N-1} \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n \\ \frac{1}{N-1} \left[ \sum_{n=0}^{N-1} x^2(n) - \hat{A}^2 \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n \right] \end{bmatrix}$$

$$18) p(x(n)) = \begin{cases} \frac{1}{\theta_2 - \theta_1}, & \theta_1 \leq x(n) \leq \theta_2 \\ 0 & \text{otherwise} \end{cases}$$

$$p(\underline{x}; \underline{\theta}) = \begin{cases} \frac{1}{(\theta_2 - \theta_1)^n} & \text{all } x(n) \text{ satisfy} \\ & \theta_1 \leq x(n) \leq \theta_2 \\ 0 & \text{otherwise} \end{cases}$$

Alternatively, for the PDF to be nonzero

$$\min x(n) \geq \theta_1, \quad \max x(n) \leq \theta_2 \quad \text{so that}$$

$$p(\underline{x}; \underline{\theta}) = \underbrace{\frac{1}{(\theta_2 - \theta_1)^n} u(\min x(n) - \theta_1) u(\max x(n) - \theta_2)}_{g(\tau(\underline{x}), \underline{\theta})}$$

$$\underbrace{\cdot \cdot \cdot}_{h(\underline{x})}$$

$$\Rightarrow \underline{T}(\underline{x}) = \begin{bmatrix} \min x(n) \\ \max x(n) \end{bmatrix} \quad \text{is a sufficient statistic}$$

$$\begin{aligned} 19) & (\underline{x} - \underline{H}\hat{\underline{\theta}})^T (\underline{x} - \underline{H}\hat{\underline{\theta}}) + (\underline{\theta} - \hat{\underline{\theta}})^T \underline{H}^T \underline{H} (\underline{\theta} - \hat{\underline{\theta}}) \\ &= (\underline{x} - \underline{H}(\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x})^T (\underline{x} - \underline{H}(\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x}) \\ &+ (\underline{\theta} - (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x})^T \underline{H}^T \underline{H} (\underline{\theta} - (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x}) \\ &= \underline{x}^T (\underline{I} - \underline{H}(\underline{H}^T \underline{H})^{-1} \underline{H}^T) (\underline{I} - \underline{H}(\underline{H}^T \underline{H})^{-1} \underline{H}^T) \underline{x} \\ &+ \underline{\theta}^T \underline{H}^T \underline{H} \underline{\theta} - \underline{\theta}^T \underline{H}^T \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x} \\ &- \underline{x}^T \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{\theta} + \underline{x}^T \underline{H}^T (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x} \\ &= \underline{x}^T \underline{x} - \underline{x}^T \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x} + \underline{\theta}^T \underline{H}^T \underline{H} \underline{\theta} - 2\underline{\theta}^T \underline{H}^T \underline{x} \\ &+ \underline{x}^T \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x} = (\underline{x} - \underline{H}\underline{\theta})^T (\underline{x} - \underline{H}\underline{\theta}) \end{aligned}$$

$$\begin{aligned}
 p(\underline{x}; \theta) &= \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} (\underline{x} - H\theta)^T (\underline{x} - H\theta)} \\
 &= \underbrace{\frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} (\underline{\theta} - \hat{\theta})^T H^T H (\underline{\theta} - \hat{\theta})}}_g(\tau(\underline{x}), \underline{\theta}) \underbrace{e^{-\frac{1}{2\sigma^2} (\underline{x} - H\hat{\theta})^T (\underline{x} - H\hat{\theta})}}_h(\underline{x})
 \end{aligned}$$

where  $\tau(\underline{x}) = \hat{\theta}$  = sufficient statistic  
 Since we already know that  $\hat{\theta}$  is unbiased,  
 it is the MVU estimator. This is not  
 unexpected since we saw in Chapter 3  
 that  $\hat{\theta}$  is efficient.

## Chapter 6

1)  $\hat{A} = (\underline{H}^T \underline{\Sigma}^{-1} \underline{H})^{-1} \underline{H}^T \underline{\Sigma}^{-1} \underline{x}$

where  $\underline{H} = \begin{bmatrix} 1 \\ r \\ \vdots \\ r^{N-1} \end{bmatrix} \quad \underline{\Sigma} = \sigma^2 \underline{I}$

$$\hat{A} = \left( \frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n} \right)^{-1} \frac{1}{\sigma^2} \sum_{n=0}^{N-1} x[n] r^n$$

$$= \frac{\sum_{n=0}^{N-1} x[n] r^n}{\sum_{n=0}^{N-1} r^{2n}}$$

$$\text{var}(\hat{A}) = \frac{1}{\underline{H}^T \underline{\Sigma}^{-1} \underline{H}} = \frac{\sigma^2}{\sum_{n=0}^{N-1} r^{2n}}$$

$\text{var}(\hat{A}) \rightarrow 0$  if  $|r| \geq 1$ .

2)  $\text{var}(\hat{A}) = \frac{1}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}}$

If  $\sigma_n^2 = n+1$ ,  $\text{var}(\hat{A}) = \frac{1}{\sum_{n=0}^{N-1} \frac{1}{n+1}}$

As  $N \rightarrow \infty$ ,  $\sum_{n=0}^{N-1} \frac{1}{n+1} \rightarrow \infty$  since this is a harmonic series  $\Rightarrow \text{var}(\hat{A}) \rightarrow 0$

If  $\sigma_n^2 = (n+1)^2$ , as  $N \rightarrow \infty$   $\sum_{n=0}^{N-1} \frac{1}{(n+1)^2} \rightarrow \text{constant}$   
 $\Rightarrow \text{var}(\hat{A}) \not\rightarrow 0$ .

In this case the noise samples have such a large variance that the estimator

variance does not go to zero.

$$3) \hat{A} = \frac{\underline{1}^T \underline{C}^{-1} \underline{x}}{\underline{1}^T \underline{C}^{-1} \underline{1}}$$

$$\underline{C}^{-1} = \frac{1}{\sigma^2} \begin{bmatrix} \underline{B} & & \\ & \underline{B} & \underline{0} \\ & & \ddots \\ & \underline{0} & & \underline{B} \end{bmatrix} \quad \text{where } \underline{B} = \frac{\begin{bmatrix} 1-p \\ -p & 1 \end{bmatrix}}{1-p^2}$$

$\underline{1}^T \underline{C}^{-1} \underline{1}$  = sum of all elements in  $\underline{C}^{-1}$

$$= \frac{N/2}{\sigma^2} \frac{2-2p}{1-p^2} = \frac{N}{\sigma^2(1+p)}$$

Let  $\underline{x} = [\underline{x}_1^T \underline{x}_2^T \dots \underline{x}_{N/2}^T]^T$  where each  $\underline{x}_i$  is  $2 \times 1$ .

$$\underline{1}^T \underline{C}^{-1} \underline{x} = \frac{1}{\sigma^2} \underline{1}^T \begin{bmatrix} \underline{B} \underline{x}_1 \\ \vdots \\ \underline{B} \underline{x}_{N/2} \end{bmatrix}$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^{N/2} \underset{2 \times 1}{\underset{\uparrow}{\underline{1}^T \underline{B} \underline{x}_i}}$$

$$\text{But } \underline{B}^T \underline{1} = \frac{\begin{bmatrix} 1-p \\ 1-p \end{bmatrix}}{1-p^2} \Rightarrow \underline{1}^T \underline{B} \underline{x}_i = \underline{x}_i^T \underline{B}^T \underline{1}$$

$$= \frac{(1-p)(\underline{x}_i)_1 + (1-p)(\underline{x}_i)_2}{1-p^2}$$

$$= \frac{(\underline{x}_i)_1 + (\underline{x}_i)_2}{1-p^2}$$

$$\underline{1}^T \underline{C}^{-1} \underline{x} = \frac{1}{\sigma^2} \frac{1}{1+p} \sum_{i=1}^{N/2} (\underline{x}_i)_1 + (\underline{x}_i)_2 \frac{1+p}{2}$$

$$= \frac{1}{\sigma^2} \frac{1}{1+p} \sum_{n=0}^{N-1} x(n)$$

$$\hat{A} = \frac{\frac{1}{\sigma^2} \frac{N \bar{x}}{1+p}}{\frac{N}{\sigma^2(1+p)}} = \bar{x}$$

$$\text{var}(\hat{A}) = \frac{\frac{1}{N}}{\frac{\sigma^2(1+p)}{N}} = \frac{\sigma^2(1+p)}{N}$$

Since the subvectors  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_{N/2}$  are uncorrelated, we average them. For a single subvector we also average the samples (see Probs. 3.9 and 4.11). Hence, we obtain  $\bar{x}$ . The variance is  $\sigma^2/N$  for  $p=0$  (our usual case),  $2\sigma^2/N$  for  $p \rightarrow 1$ , since the samples of each subvector are equal and hence we have only  $N/2$  uncorrelated samples, and  $\rightarrow 0$  for  $p \rightarrow -1$ , since then the noise samples cancel (see Probs. 3.9 and 4.11).

- 4) In either case we have the model

$$\underline{x} = \underline{L} \underline{u} + \underline{w} \quad \text{where } E(\underline{w}) = \underline{0}$$

$$\text{and } E(\underline{w}\underline{w}^T) = \text{var}(\underline{w}) \underline{I}$$

The BLUE for each case is

$$\hat{u} = \frac{\underline{L}^T \underline{L}^{-1} \underline{x}}{\underline{L}^T \underline{L}^{-1} \underline{L}} = \frac{\underline{L}^T \underline{x}}{\underline{L}^T \underline{L}} = \underline{x}$$

But in the Gaussian case the BLUE is also the MVU estimator - not so for the Laplacian PDF.

$$(5) E(x) = \int_0^\infty x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\ln x - \theta)^2} dx$$

$$\text{Let } y = \ln x \quad dy/dx = 1/x \quad \Rightarrow \quad dx = e^y dy$$

$$\begin{aligned} E(x) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\theta)^2} e^y dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y^2 - 2y(\theta+1) + (\theta+1)^2)} \\ &\quad \cdot e^{-\frac{1}{2}(\theta^2 - (\theta+1)^2)} dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-(\theta+1))^2} dy \\ &\quad \cdot e^{-\frac{1}{2}(-2\theta-1)} \\ &= e^{\theta+\frac{1}{2}} \end{aligned}$$

$$\text{Now let } y = \ln x, \quad dy/dx = 1/x = e^{-y}$$

$$\begin{aligned} p(y) &= \frac{p(x(y))}{|dy/dx|} = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\theta)^2}}{e^{-y}} \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\theta)^2} \\ \Rightarrow y &\sim N(\theta, 1) \end{aligned}$$

$$\text{The BLUE is } \hat{\theta} = \frac{1}{N} \sum_{n=0}^{N-1} y[n] \\ = \frac{1}{N} \sum_{n=0}^{N-1} a_n x[n]$$

b) For  $\hat{\theta}$  to be unbiased

$$\begin{aligned} E(\hat{\theta}) &= E\left(\sum_n a_n x[n] + b\right) \\ &= \sum_n a_n (E[x[n]] + \rho) + b \\ &= \theta \\ \Rightarrow \sum_n a_n s[n] &= 1, \quad \beta \sum_n a_n = -b \end{aligned}$$

$$\begin{aligned} \text{or } \hat{\theta} &= \sum_n a_n x[n] - \beta \sum_n a_n \\ &= \sum_n a_n (x[n] - \rho) \end{aligned}$$

Let  $x'[n] = x[n] - \rho$ . Then, we have the same problem as before. Thus,

$$\hat{\theta} = \frac{\underline{s}^T C^{-1} x'}{\underline{s}^T C^{-1} \underline{s}} = \frac{\underline{s}^T C^{-1} (x - \rho)}{\underline{s}^T C^{-1} \underline{s}}$$

and since the covariance for  $x - \rho$  is  $C$ ,

$$\text{var}(\hat{\theta}) = \frac{1}{\underline{s}^T C^{-1} \underline{s}}$$

$$7) \quad \hat{A} = \frac{\underline{s}^T C^{-1} x}{\underline{s}^T C^{-1} \underline{s}}$$

Assume  $C^{-1} \underline{s} = \lambda \underline{s}$ . Then,  $C^{-1} \underline{s} = 1/\lambda \underline{s}$

$$\hat{A} = \frac{\underline{x}^T \underline{\zeta}^{-1} \underline{\zeta}}{\underline{\zeta}^T \underline{\zeta}^{-1} \underline{\zeta}} = \frac{\underline{x}^T / \lambda \underline{\zeta}}{\underline{\zeta}^T / \lambda \underline{\zeta}} = \frac{\underline{x}^T \underline{\zeta}}{\underline{\zeta}^T \underline{\zeta}}$$

$$\text{var}(\hat{A}) = \frac{1}{\underline{\zeta}^T \underline{\zeta}^{-1} \underline{\zeta}} = \frac{1}{\underline{\zeta}^T \underline{\zeta}}$$

In this case we obtain same result as if  $\underline{\zeta} = \sigma^2 \underline{I}$ . We do not need a prewhitener.

$$\begin{aligned} \text{a) } \text{var}(\hat{A}) &= \frac{1}{\underline{\zeta}^T \underline{\zeta}^{-1} \underline{\zeta}} \\ &= \frac{1}{(\sum_i \alpha_i v_i)^T \underline{\zeta}^{-1} (\sum_j \alpha_j v_j)} \\ &= \frac{1}{\sum_{i,j} \alpha_i \alpha_j \underbrace{v_i^T \underline{\zeta}^{-1} v_j}_{V_i^T / \lambda_j V_j} \underbrace{\lambda_j}_{1/\lambda_j \delta_{ij}}} \\ &= \frac{1}{\sum_i \alpha_i^2 / \lambda_i} \end{aligned}$$

$$\begin{aligned} \underline{\zeta} &= \underline{\zeta}^T \underline{\zeta} = (\sum_i \alpha_i v_i)^T (\sum_j \alpha_j v_j) \\ &= \sum_{i,j} \sum_i \alpha_i \alpha_j \underbrace{v_i^T v_j}_{\delta_{ij}} = \sum_i \alpha_i^2 \end{aligned}$$

Must minimize  $\frac{1}{\sum \alpha_i^2/\lambda_i}$  subject

to constraint  $\sum_i \alpha_i^2 = \epsilon_0$ . Equivalently,  
we must maximize  $\sum_i \alpha_i^2/\lambda_i$ . Using  
Lagrange multipliers

$$F = \sum_i \alpha_i^2/\lambda_i + \lambda / (\sum_i \alpha_i^2 - \epsilon_0)$$

$$\frac{\partial F}{\partial \alpha_k} = \frac{2\alpha_k}{\lambda_k} + \lambda 2\alpha_k = 0$$

$$\Rightarrow \alpha_k = 0 \text{ or}$$

$$\lambda = -1/\lambda_k \text{ all } k$$

Clearly, we cannot have  $\alpha_k = 0$  for all  $k$ , since  
then constraint could not be satisfied.  
Since the eigenvalues are distinct, we  
also cannot have  $\lambda = -1/\lambda_k$  for all  $k$ .  
Thus, we must have

$$\alpha_k = 0 \text{ except for } k=j$$

$$\lambda = -1/\lambda_j \text{ and } \alpha_j \neq 0.$$

Hence,  $\underline{\alpha} = \alpha_j v_j$ . To determine which  
eigenvector to use

$$\text{var}(\underline{\alpha}) = \frac{1}{\alpha_j^2/\lambda_j} = \lambda_j/\alpha_j^2$$

And since  $\sigma_0^2 = \lambda_j^2$ ,  $\text{var}(\hat{A}) = \lambda_j / \sigma_0^2$   
so that  $\lambda_j$  should be the minimum  
eigenvalue. Hence, the optimal signal is

$$\underline{s} = c \underline{V}_{MN} = \sqrt{\sigma_0^2} \underline{V}_{MN}$$

where  $\underline{V}_{MN}$  is the eigenvector associated  
with the smallest eigenvalue. Intuitively,  
we place signal along direction where  
there is the least amount of noise.

9)  $\theta = A$

$$\underline{s} = [1, \cos 2\pi f_1, \dots, \cos 2\pi f_{(N-1)}]^T$$

$$\hat{A} = \frac{\underline{s}^T \underline{C}^{-1} \underline{x}}{\underline{s}^T \underline{C}^{-1} \underline{s}} = \frac{\underline{s}^T \underline{x}}{\underline{s}^T \underline{s}}$$

$$= \frac{\sum_{n=0}^{N-1} x(n) \cos 2\pi f_n n}{\sum_{n=0}^{N-1} \cos^2 2\pi f_n n}$$

This is a scaled Fourier coefficient since

$$\hat{A} = \frac{N/2}{\sum_n \cos^2 2\pi f_n n} \underbrace{\frac{2}{N} \sum_n x(n) \cos 2\pi f_n n}_{\text{Fourier coefficient}}$$

and for large  $N$  and  $f_i$  not near 0 or  $1/2$   
the scale factor is one.

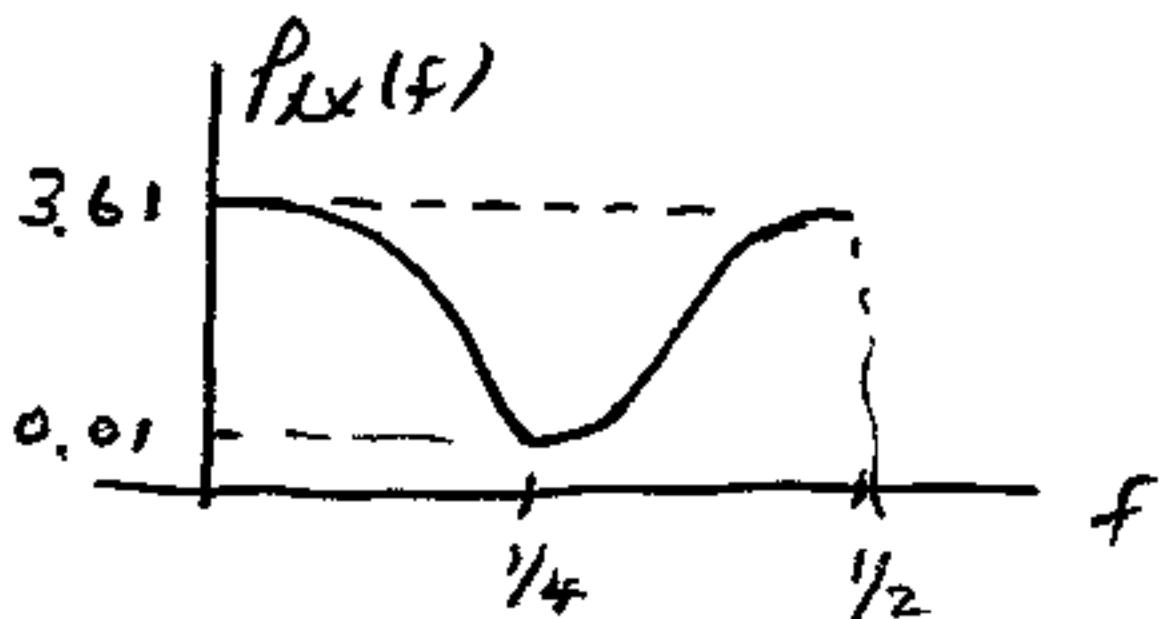
$$\text{var}(\hat{A}) = \frac{\underline{s^T C^{-1} s}}{N} = \frac{\sigma^2}{\sum_{n=0}^{N-1} \cos^2 2\pi f_n}$$

$$= \frac{\sigma^2}{\sum_{n=0}^{N-1} \cos^2 2\pi f_n} \geq \frac{\sigma^2}{N}$$

Since  $\sum_{n=0}^{N-1} \cos^2 2\pi f_n \leq N$

$\Rightarrow$  variance is minimized if  $f_i = 0$ .  
 Note that this choice maximizes the signal energy.

$$\begin{aligned} 10) \quad P_{xx}(f) &= r_{xx}(0) + r_{xx}(2) e^{-j4\pi f} \\ &\quad + r_{xx}(2) e^{j4\pi f} \\ &= 1.81 + 2(0.9) \cos 4\pi f \\ &= 1.81 + 1.8 \cos 4\pi f \end{aligned}$$

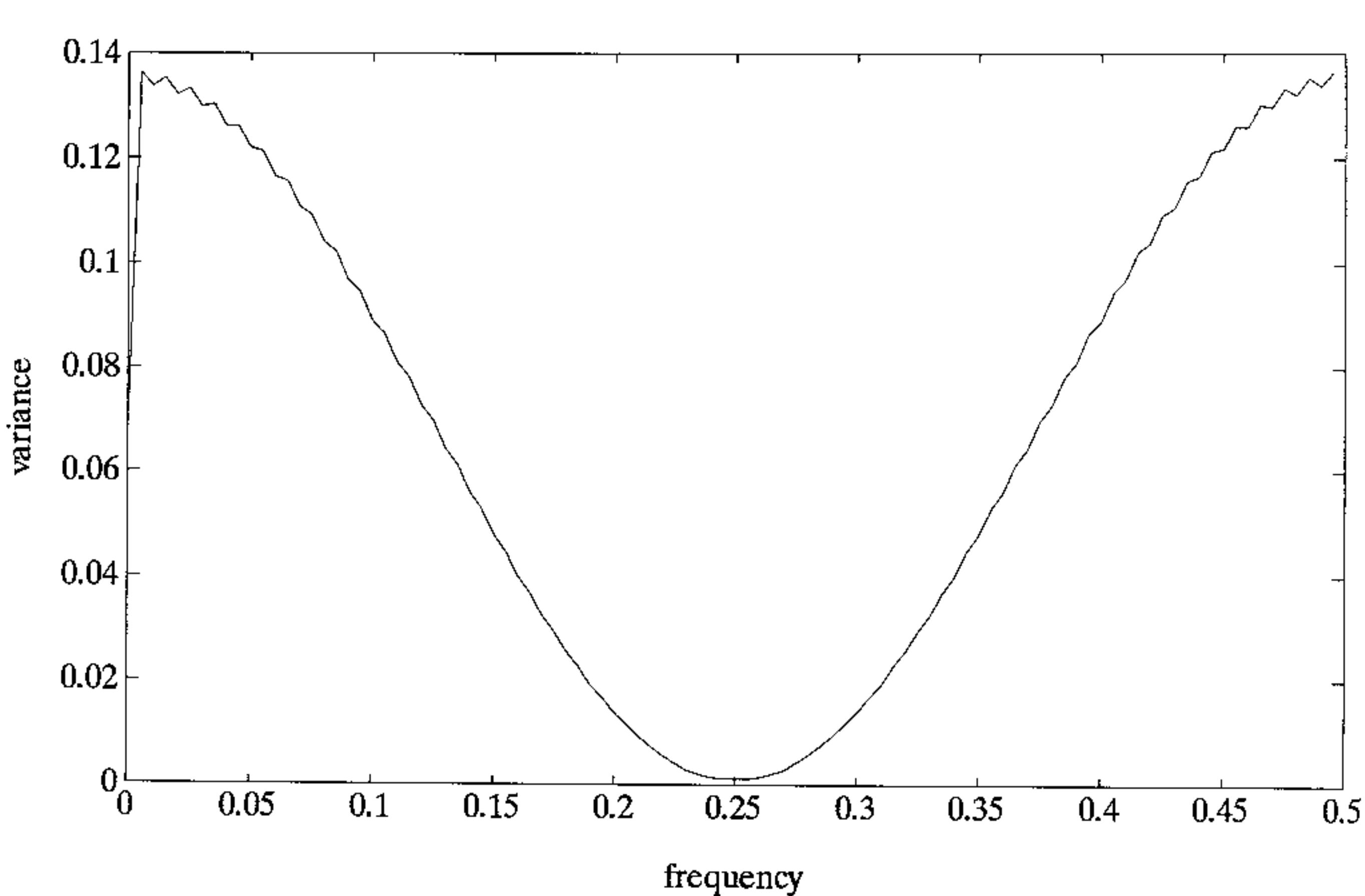


Need to minimize  $\frac{\underline{s^T C^{-1} s}}{N} = \text{var}(\hat{A})$   
 where  $\underline{s} = [1 \cos 2\pi f_1 \dots \cos 2\pi f_{(N-1)}]^T$   
 and

$$\underline{C} = \begin{bmatrix} r_{xx}(0) & r_{xx}(1) & \dots & r_{xx}(N-1) \\ r_{xx}(1) & r_{xx}(0) & & \\ \vdots & & \ddots & \\ r_{xx}(N-1) & & & r_{xx}(0) \end{bmatrix}$$

$$= \begin{bmatrix} 1.81 & 0 & 0.9 & 0 & \dots & 0 \\ 0 & 1.81 & 0 & 0.9 & \dots & 0 \\ & & \ddots & & & \\ & & & & & 1.81 \end{bmatrix}$$

As seen in the following graph, the variance is minimized for  $f_1 = 0.25$  or where the PSD is minimum.



$$(1) \quad H(e^{j2\pi f}) = \sum_{n=0}^{N-1} h[n] e^{-j2\pi f n}$$

$$H(e^{j0}) = \sum_{n=0}^{N-1} h[n] = 1$$

The noise power at the output at  $n = N-1$  is

$$E\left[\left(\sum_k h[k] w[N-1-k]\right)^2\right]$$

$$= \sum_k \sum_l h[k] h[l] E\left[w[N-1-k] w[N-1-l]\right] \\ r_{ww[k-1]}$$

$$= \underline{h}^T \underline{\underline{C}} \underline{h} \quad \text{where } (\underline{\underline{C}})_{kl} = r_{ww[k-1]}$$

Thus, to find the FIR filter coefficients we need to minimize  $\underline{h}^T \underline{\underline{C}} \underline{h}$  subject to the constraint  $\|\underline{h}\|^2 = 1$ . This is just the DLE setup so that

$$\underline{h}_{opr} = \frac{\underline{\underline{C}}^{-1} \underline{I}}{\underline{I}^T \underline{\underline{C}}^{-1} \underline{I}}$$

The noise power at the output is

$$\underline{h}_{opr}^T \underline{\underline{C}} \underline{h}_{opr} = \frac{\underline{I}^T \underline{\underline{C}}^{-1} \underline{\underline{C}} \underline{C}^{-1} \underline{I}}{(\underline{I}^T \underline{\underline{C}}^{-1} \underline{I})^2}$$

$$= \frac{1}{\text{Tr}(\Sigma^{-1})}$$

which is just the variance of  $\hat{A}$  since

$$\begin{aligned}\text{var}(\hat{A}) &= E[(\hat{A} - E(\hat{A}))^2] \\ &= E\left[\left(\sum_k h[k] \times [x_{N-i-k} - \sum_k h[k] A]\right)^2\right]\end{aligned}$$

since  $\sum_k h[k] = 1$

$$\begin{aligned}&= E\left[\left(\sum_k h[k] (x_{N-i-k} - A)\right)^2\right] \\ &= E\left[\left(\sum_k h[k] w_{N-i-k}\right)^2\right]\end{aligned}$$

The BLUE may be viewed as the output of a linear filter constrained to pass the DC signal and whose coefficients are chosen to minimize the noise at the filter output.

12) Since  $\underline{x} = \underline{H}\underline{\theta} + \underline{w}$  and  $\underline{B}$  is invertible,  
 $\underline{\theta} = \underline{B}^{-1}(\underline{\alpha} - \underline{b}) \Rightarrow \underline{x} = \underline{H}\underline{B}^{-1}(\underline{\alpha} - \underline{b}) + \underline{w}$

or  $\underline{x} = \underline{H}\underline{B}^{-1}\underline{\alpha} - \underline{H}\underline{B}^{-1}\underline{b} + \underline{w}$

$$\underbrace{\underline{x} + \underline{H}\underline{B}^{-1}\underline{b}}_{\underline{x}'} = \underbrace{\underline{H}\underline{B}^{-1}\underline{\alpha} + \underline{w}}_{\underline{w}'}$$

$$\begin{aligned}
 \Rightarrow \hat{\theta} &= (\underline{H}^T \underline{C}^{-1} \underline{H})^{-1} \underline{H}^T \underline{C}^{-1} \underline{x} \\
 &= (\underline{B}^{-1}{}^T \underline{H}^T \underline{C}^{-1} \underline{H} \underline{B}^{-1})^{-1} \underline{B}^{-1}{}^T \underline{H}^T \underline{C}^{-1} (\underline{x} + \underline{H} \underline{B}^{-1} \underline{b}) \\
 &= \underline{B} (\underline{H}^T \underline{C}^{-1} \underline{H})^{-1} \underline{H}^T \underline{C}^{-1} (\underline{x} + \underline{H} \underline{B}^{-1} \underline{b}) \\
 &= \underline{B} \hat{\theta} + \underline{B} \underline{B}^{-1} \underline{b} = \underline{B} \hat{\theta} + \underline{b}
 \end{aligned}$$

13)  $\mathcal{T} = \underline{x}^T \underline{C}^{-1} \underline{x} - \underline{x}^T \underline{C}^{-1} \underline{H} \underline{\theta} - \underline{\theta}^T \underline{H}^T \underline{C}^{-1} \underline{x}$   
 $+ \underline{\theta}^T \underline{H}^T \underline{C}^{-1} \underline{H} \underline{\theta}$   
 $= \underline{x}^T \underline{C}^{-1} \underline{x} - 2 \underline{\theta}^T \underline{H}^T \underline{C}^{-1} \underline{x} + \underline{\theta}^T \underline{H}^T \underline{C}^{-1} \underline{H} \underline{\theta}$

Using (4.3) we have

$$\frac{\partial \mathcal{T}}{\partial \underline{\theta}} = -2 \underline{H}^T \underline{C}^{-1} \underline{x} + 2 \underline{H}^T \underline{C}^{-1} \underline{H} \underline{\theta} = 0$$

$$\Rightarrow \hat{\underline{\theta}} = (\underline{H}^T \underline{C}^{-1} \underline{H})^{-1} \underline{H}^T \underline{C}^{-1} \underline{x}$$

14) Since  $p(w_{(n)})$  is even,  $E(w_{(n)}) = 0$ .

$$\text{var}(w_{(n)}) = E(w^2_{(n)})$$

$$\begin{aligned}
 &= \int_{-\infty}^{\infty} w^2 \left( \frac{1-\epsilon}{\sqrt{2\pi\sigma_\theta^2}} e^{-\frac{1}{2} w^2/\sigma_\theta^2} \right. \\
 &\quad \left. + \frac{\epsilon}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{1}{2} w^2/\sigma_x^2} \right) dw
 \end{aligned}$$

$$\begin{aligned}
 &= (1-\epsilon) \int_{-\infty}^{\infty} w^2 \frac{1}{\sqrt{2\pi\sigma_B^2}} e^{-\frac{w^2}{2\sigma_B^2}} dw \\
 &\quad + \epsilon \int_{-\infty}^{\infty} w^2 \frac{1}{\sqrt{2\pi\sigma_I^2}} e^{-\frac{w^2}{2\sigma_I^2}} dw \\
 &= (1-\epsilon)\sigma_B^2 + \epsilon\sigma_I^2
 \end{aligned}$$

The BLUE of  $\sigma^2$  is  $\hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} y(n)^2$  (see section 6.3). This is because the  $w(n)$ 's are independent and thus  $y(n) = w^2 L(n)$ 's are independent. The mean is  $E(y(n)) = \sigma^2$  and covariance matrix is  $\Sigma = \text{var}(y(n)) \equiv \Sigma$  so that

$$\hat{\sigma}^2 = \frac{\underline{y}^T \underline{\Sigma}^{-1} \underline{y}}{\underline{y}^T \underline{\Sigma}^{-1} \underline{y}} = \frac{\underline{I}^T \underline{y}}{\underline{I}^T \underline{I}} = \frac{1}{N} \sum_{n=0}^{N-1} y(n)^2$$

Now using results from Prob. 6.12

$$\sigma_I^2 = \frac{\sigma^2 - (1-\epsilon)\sigma_B^2}{\epsilon}$$

$$\begin{aligned}
 \Rightarrow \hat{\sigma}_I^2 &= \frac{\hat{\sigma}^2 - (1-\epsilon)\sigma_B^2}{\epsilon} \\
 &= \frac{\frac{1}{N} \sum_{n=0}^{N-1} y(n)^2 - (1-\epsilon)\sigma_B^2}{\epsilon}
 \end{aligned}$$

$$15) \quad \underbrace{\underline{x} - \underline{s}}_{\underline{x}'} = \underline{H}\underline{\theta} + \underline{w}$$

$$\Rightarrow \hat{\underline{\theta}} = (\underline{H}^T \underline{C}^{-1} \underline{H})^{-1} \underline{H}^T \underline{C}^{-1} \underline{x}'$$

$$= (\underline{H}^T \underline{C}^{-1} \underline{H})^{-1} \underline{H}^T \underline{C}^{-1} (\underline{x} - \underline{s})$$

$$16) \quad \hat{A} = (\underline{I}^T \hat{\underline{C}}^{-1} \underline{I})^{-1} \underline{I}^T \hat{\underline{C}}^{-1} \underline{x}$$

$$E(\hat{A}) = (\underline{I}^T \hat{\underline{C}}^{-1} \underline{I})^{-1} \underline{I}^T \hat{\underline{C}}^{-1} E(\underline{x}) \underbrace{=} A$$

$\Rightarrow$  unbiased for any  $\hat{\underline{C}}$

$$\text{var}(\hat{A}) = E[(\hat{A} - A)^2] = E[((\underline{I}^T \hat{\underline{C}}^{-1} \underline{I})^{-1} \underline{I}^T \hat{\underline{C}}^{-1} \underline{w})^2]$$

$$= \frac{\underline{I}^T \hat{\underline{C}}^{-1} E(\underline{w} \underline{w}^T) \hat{\underline{C}}^{-1} \underline{I}}{(\underline{I}^T \hat{\underline{C}}^{-1} \underline{I})^2}$$

$$= \frac{\underline{I}^T \hat{\underline{C}}^{-1} \underline{C} \hat{\underline{C}}^{-1} \underline{I}}{(\underline{I}^T \hat{\underline{C}}^{-1} \underline{I})^2}$$

$$\text{If } \underline{C} = \underline{\Sigma}, \quad \hat{\underline{C}} = \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix} \Rightarrow \hat{\underline{C}}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1/\alpha \end{pmatrix}$$

$$\text{var}(\hat{A}) = \frac{\underline{I}^T \begin{pmatrix} 1 & 0 \\ 0 & 1/\alpha \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1/\alpha \end{pmatrix} \underline{I}}{(\underline{I}^T \begin{pmatrix} 1 & 0 \\ 0 & 1/\alpha \end{pmatrix} \underline{I})^2}$$

$$= \frac{1 + \frac{1}{\alpha^2}}{(1 + \frac{1}{\alpha})^2}$$

Clearly, by the minimum variance property of the BLUE

$$\text{var}(\hat{\alpha})_{\text{MIN}} = \frac{1}{2} \quad \text{for } \alpha = 1$$

$$\begin{aligned} \text{Let } f &= \frac{\text{var}(\hat{\alpha})}{\text{var}(\hat{\alpha})_{\text{MIN}}} = \frac{2(1 + \frac{1}{\alpha^2})}{(1 + \frac{1}{\alpha})^2} \\ &= \frac{2(1 + \alpha^2)}{(1 + \alpha)^2} \end{aligned}$$



For  $\alpha \rightarrow 0$  or  $\alpha \rightarrow \infty$  we desired one data sample and thus the variance doubles. For  $\alpha = 1$  we have the BLUE and hence the minimum variance.

Chapter 7

$$1) \quad p(\underline{x}; A) = \frac{1}{(2\pi A)^{N/2}} e^{-\frac{1}{2A} \sum_n (x_{(n)} - A)^2}$$

$$\frac{\partial \ln p}{\partial A} = -\frac{N}{2A} + \frac{1}{A} \sum_n (x_{(n)} - A) + \frac{1}{2A^2} \sum_n (x_{(n)} - A)^2$$

$$\begin{aligned} \frac{\partial^2 \ln p}{\partial A^2} &= \frac{N}{2A^2} - \frac{1}{A^2} \sum_n x_{(n)} - \frac{1}{A^2} \sum_n (x_{(n)} - A) \\ &\quad - \frac{1}{A^3} \sum_n (x_{(n)} - A)^2 \end{aligned}$$

$$\begin{aligned} E\left(\frac{\partial^2 \ln p}{\partial A^2}\right) &= \frac{N}{2A^2} - \frac{NA}{A^2} - 0 - \frac{1}{A^3} (NA) \\ &= -\frac{N}{2A^2} - \frac{N}{A} \end{aligned}$$

$$I(A) = \frac{N}{2A^2} + \frac{N}{A} \Rightarrow \text{var}(\hat{A}) \geq \frac{1}{\frac{N}{2A^2} + \frac{N}{A}}$$

$$\text{or } \text{var}(\hat{A}) \geq \frac{A^2}{N(A + \frac{1}{2})}$$

$$2) \quad \text{var}(\bar{x}) = \sigma^2/N = A/N$$

$$\text{But } \text{var}(\hat{A}) \geq \frac{A}{N} \left( \frac{A}{A+1/2} \right) < A/N$$

Even as  $N \rightarrow \infty$ ,  $\bar{x}$  does not attain CRLB. Thus, MLE is better (at least for large data records). For finite data records we would need to determine the exact mean and variance of  $\hat{A}$  and compare them to  $\bar{x}$ .

3) a)  $p(\underline{x}; \mu) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x(n)-\mu)^2}$

 $= \frac{1}{(2\pi)^N h} e^{-\frac{1}{2} \sum_{n=0}^{N-1} (x(n)-\mu)^2}$

To maximize  $p$ , we minimize  $\sum_n (x(n)-\mu)^2$ . Since it is a quadratic in  $\mu$ , differentiation produces a global minimum.

$\Rightarrow \sum_n (x(n)-\mu) = 0 \Rightarrow \hat{\mu} = \bar{x}$

(This is just a DC level,  $\mu$ , in WGN).

b)  $p(\underline{x}; \lambda) = \lambda^N e^{-\lambda \sum_n x(n)}$

all  $x(n) > 0$   
otherwise

Assuming all  $x(n) > 0$  we have

$p = \lambda^N e^{-\lambda N \bar{x}}$

$\frac{dp}{d\lambda} = N \lambda^{N-1} e^{-\lambda N \bar{x}} + \lambda^N (-N \bar{x}) e^{-\lambda N \bar{x}} = 0$

$\Rightarrow \hat{\lambda} = 1/\bar{x}$

To verify that  $\hat{\lambda}$  yields the global maximum we can consider  $\ln p$ , which is a monotonic function of  $p$ .

$$\ln p = N \ln \lambda - N \bar{x}$$

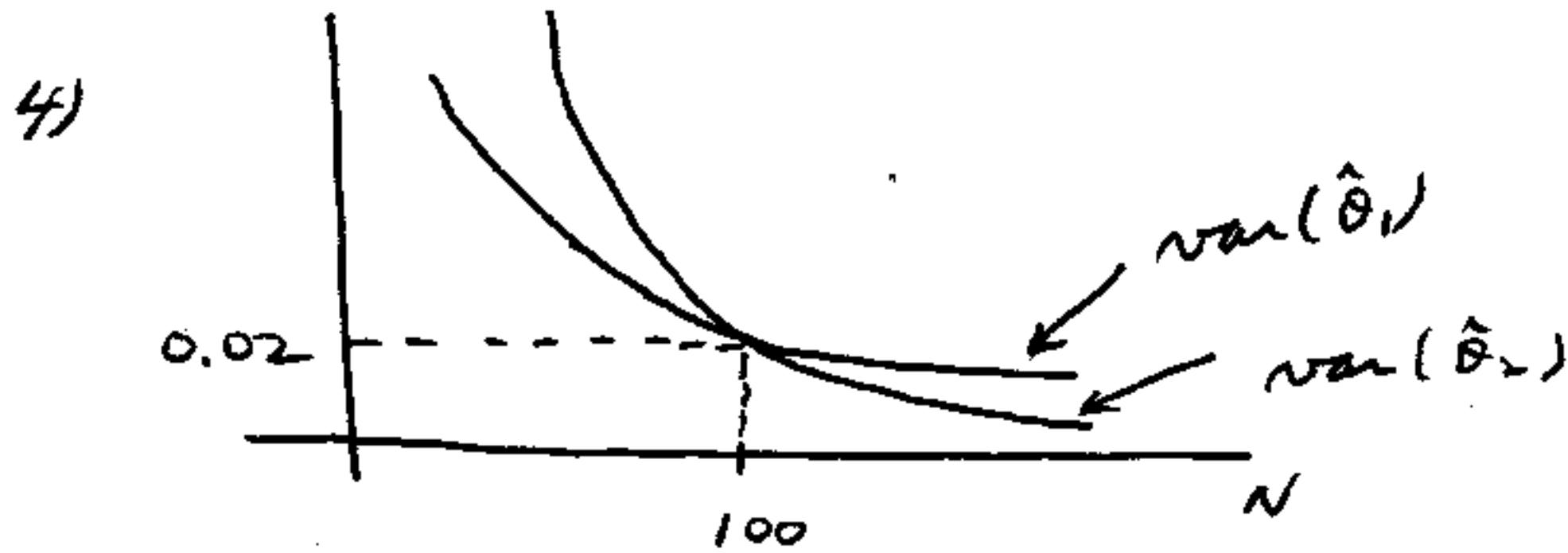
$$\frac{d \ln p}{d \lambda} = N/\lambda - N \bar{x}$$

$$\frac{d^2 \ln p}{d \lambda^2} = -N/\lambda^2 < 0 \quad \text{for all } \lambda$$

$\Rightarrow \ln p$  is concave function

$\Rightarrow \hat{\lambda}$  is global maximum solution

This result is reasonable since the mean of  $x$  is ' $\lambda$ '.



$\hat{\theta}_1$  better for  $N < 100$

$\hat{\theta}_2$  better for  $N > 100$

5)  $\Pr \{ |\hat{A} - A| > \epsilon \} = \Pr \left\{ \left| \frac{\bar{x} - A}{\sigma/\sqrt{N}} \right| > \frac{\epsilon}{\sigma/\sqrt{N}} \right\}$

$$\leq \frac{1}{(\epsilon/\sigma/\sqrt{N})^2}$$

$$= \frac{\sigma^2}{N\epsilon^2} \rightarrow 0 \text{ as } N \rightarrow \infty$$

$\Rightarrow \bar{x}$  is consistent

6) Linearizing about the true value of  $\theta$ , ie,  $\theta_0$  yields

$$\alpha = g(\theta) = g(\theta_0) + \frac{dg}{d\theta} \Big|_{\theta=\theta_0} (\theta - \theta_0)$$

$$\begin{aligned} \text{But } \hat{\alpha} - \alpha &= g(\hat{\theta}) - g(\theta_0) \\ &\approx \left[ g(\theta_0) + \frac{dg}{d\theta} \Big|_{\theta=\theta_0} (\hat{\theta} - \theta_0) \right] - g(\theta_0) \\ &= \frac{dg}{d\theta} \Big|_{\theta=\theta_0} (\hat{\theta} - \theta_0) \end{aligned}$$

$$\begin{aligned} \Pr \{ |\hat{\alpha} - \alpha| > \epsilon \} &= \Pr \{ \left| \frac{dg}{d\theta} \Big|_{\theta=\theta_0} (\hat{\theta} - \theta_0) \right| > \epsilon \} \\ &= \Pr \left\{ |\hat{\theta} - \theta_0| > \frac{\epsilon}{\left| \frac{dg}{d\theta} \Big|_{\theta=\theta_0} \right|} \right\} \rightarrow 0 \text{ as } N \rightarrow \infty \end{aligned}$$

since  $\hat{\theta}$  is consistent for  $\theta$  (as long as  $dg/d\theta$  is bounded).

$$\begin{aligned} 7) p(x; \theta) &= \prod_{n=0}^{N-1} e^{A(\theta)B(x_{(n)}) + C(x_{(n)}) + D(\theta)} \\ &= e^{A(\theta) \sum_n B(x_{(n)}) + \sum_n C(x_{(n)}) + ND(\theta)} \end{aligned}$$

To maximize  $p$  we must minimize  
 $A(\theta) \sum_n B(x_{(n)}) + ND(\theta)$

Differentiating produces the necessary condition

$$\frac{dA(\theta)}{d\theta} \sum_{n=0}^N B(x(n)) + N \frac{dD(\theta)}{d\theta} = 0$$

For the PDFs of Prob. 7.2

a)  $p(x; \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x^2 - 2\mu x + \mu^2)}$

$$\Rightarrow A(\theta) = \mu \quad B(x) = x \quad D(\theta) = -\frac{1}{2}\mu^2$$

(1)  $\sum_n x(n) + N(-\mu) = 0 \Rightarrow \hat{\mu} = \bar{x}$

b)  $p(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$

$$= \begin{cases} e^{-\lambda x + \ln \lambda} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

$$\Rightarrow A(\theta) = -\lambda \quad B(x) = x \quad D(\theta) = \ln \lambda$$

(-1)  $\sum_n x(n) + N(-\lambda) = 0 \Rightarrow \hat{\lambda} = \bar{x}$

8) For discrete random variables the ML principle applies to the probability function

$$Pr\{x\} = \prod_{n=0}^{N-1} p^{x(n)} (1-p)^{1-x(n)}$$

$$= p^{\sum_n x(n)} (1-p)^{N - \sum_n x(n)}$$

$$= p^{N\bar{x}} (1-p)^{N-N\bar{x}}$$

Maximizing  $\ln \Pr\{\underline{x}\}$  over  $p$

$$\frac{d \ln \Pr\{\underline{x}\}}{dp} = \frac{d}{dp} [N\bar{x} \ln p + (N-N\bar{x}) \ln (1-p)]$$

$$= \frac{N\bar{x}}{p} + \frac{N-N\bar{x}}{1-p} (-1) = 0$$

$$N\bar{x}(1-p) - (N-N\bar{x})p = 0 \Rightarrow \hat{p} = \bar{x}$$

9)  $p(x) = \begin{cases} 1/\theta & 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}$

$$p(\underline{x}) = \prod_{n=0}^{N-1} p(x(n)) = \begin{cases} \frac{1}{\theta^N} & 0 < \text{all } x(n) < \theta \\ 0 & \text{otherwise} \end{cases}$$

Clearly,  $p(\underline{x})$  is maximized over  $\theta$  when  $\theta$  is as small as possible. But  $\theta > x(n)$  for all  $n$ . Thus,  $\theta_{\min} = \max x(n)$  and thus  $\hat{\theta} = \max x(n)$ .

10)  $p(\underline{x}; A) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} \sum_n (x(n) - Ax(n))^2}$

Minimize  $\sum_n (x(n) - Ax(n))^2$  to find MLE.

$$-2 \sum_n (x[n] - A s[n]) s[n] = 0$$

$$\Rightarrow \hat{A} = \frac{\sum_{n=0}^{N-1} x[n] s[n]}{\sum_{n=0}^{N-1} s^2[n]}$$

$$E(\hat{A}) = \frac{\sum_n E(x[n]) s[n]}{\sum_n s^2[n]} = \frac{\sum_n A s[n] s[n]}{\sum_n s^2[n]} = A$$

$$\begin{aligned} \text{var}(\hat{A}) &= \frac{\sum_n \text{var}(x[n]) s^2[n]}{\left(\sum_n s^2[n]\right)^2} \\ &= \sigma^2 \frac{\sum_n s^2[n]}{\left(\sum_n s^2[n]\right)^2} = \frac{\sigma^2}{\sum_{n=0}^{N-1} s^2[n]} \\ &= I^{-1}(A) \quad (\text{see Theorem 4.1}) \end{aligned}$$

$\hat{A}$  is Gaussian since it is linear in the data. Hence,  $\hat{A} \sim N(A, I^{-1}(A))$  and thus asymptotic PDF holds for finite data records. This problem is special case of linear model.

$$\begin{aligned}
 \text{II) } p(\underline{x}; \rho) &= \prod_{n=0}^{N-1} \frac{1}{2\pi \det(\Sigma)} e^{-\frac{1}{2} \underline{x}^T(n) \Sigma^{-1} \underline{x}(n)} \\
 &= \frac{1}{(2\pi)^N [\det(\Sigma)]^{N/2}} e^{-\frac{1}{2} \sum_n \underline{x}^T(n) \Sigma^{-1} \underline{x}(n)}
 \end{aligned}$$

But  $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \Rightarrow \Sigma^{-1} = \frac{\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}}{1-\rho^2}$   
 $\det(\Sigma) = 1-\rho^2$

$$\begin{aligned}
 \ln p &= -N \ln 2\pi - \frac{N}{2} \ln (1-\rho^2) \\
 &\quad - \frac{1}{2(1-\rho^2)} \underbrace{\sum_n \underline{x}^T(n) \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \underline{x}(n)}_Q
 \end{aligned}$$

$$\begin{aligned}
 Q &= \sum_n (x_1(n) x_2(n)) \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} \\
 &= \sum_n [x_1^2(n) + x_2^2(n) - 2\rho x_1(n) x_2(n)]
 \end{aligned}$$

$$\text{Let } C_{ij} = \sum_n x_i(n) x_j(n)$$

$$\Rightarrow Q = C_{11} + C_{22} - 2\rho C_{12}$$

$$\begin{aligned}
 \ln p &= -N \ln 2\pi - \frac{N}{2} \ln (1-\rho^2) \\
 &\quad - \frac{1}{2(1-\rho^2)} (C_{11} + C_{22} - 2\rho C_{12})
 \end{aligned}$$

$$\frac{d \ln p}{d\rho} = \frac{-N C_{12} (-2\rho)}{1-\rho^2} - \frac{1}{2(1-\rho^2)} (-2 C_{12})$$

$$-\frac{1}{2}(c_{11} + c_{22} - 2\rho c_{12}) \left[ \frac{2\rho}{(1-\rho^2)^2} \right] = 0$$

$$N\rho(1-\rho^2) + c_{12}(1-\rho^2) - \rho(c_{11} + c_{22} - 2\rho c_{12}) = 0$$

$$\rho^3 - \frac{1}{N} c_{12} \rho^2 + (\frac{1}{N} c_{11} + \frac{1}{N} c_{22} - 1) \rho - \frac{1}{N} c_{12} = 0$$

As  $N \rightarrow \infty$ ,  $\frac{c_{11}}{N} \rightarrow 1$ ,  $\frac{c_{22}}{N} \rightarrow 1$

$$\rho^3 - \frac{1}{N} c_{12} \rho^2 + \rho - \frac{1}{N} c_{12} = 0$$

for which a solution is  $\hat{\rho} = \frac{1}{N} c_{12}$   
 $= \frac{1}{N} \sum_{n=1}^{N-1} x_1(n)x_2(n)$

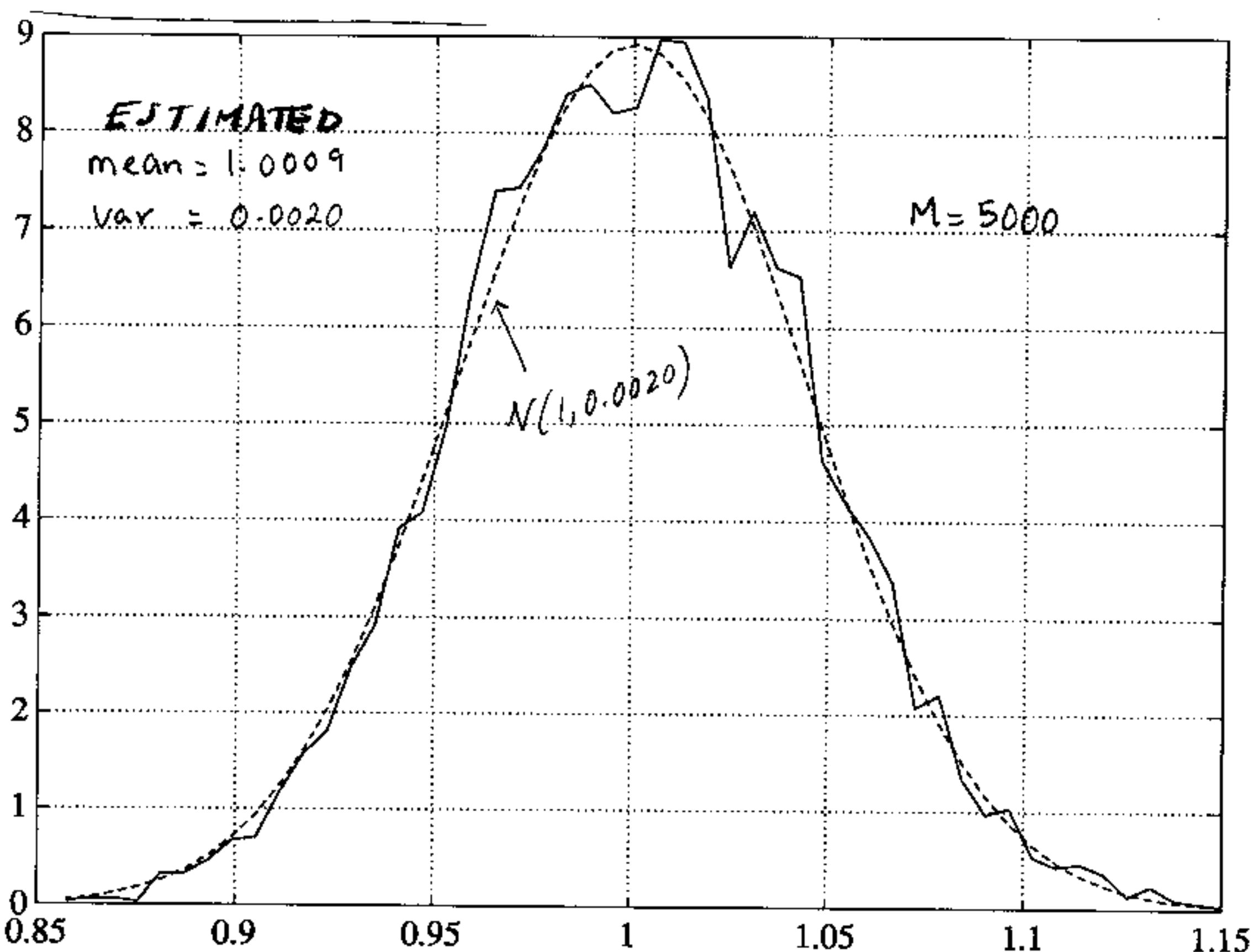
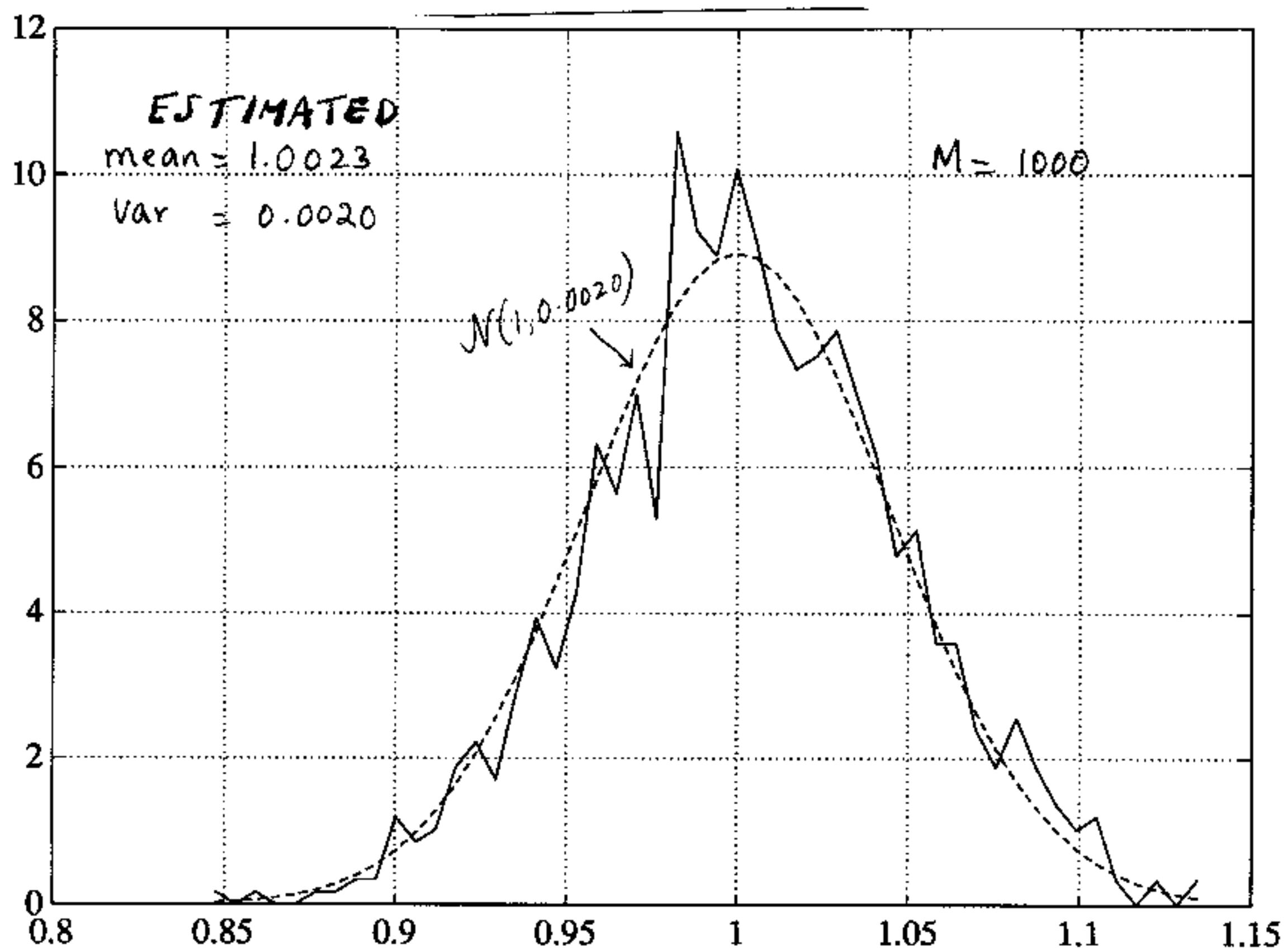
(2) To find the MLE we maximize  $\rho$  or equivalently  $\ln \rho$

$$\Rightarrow \frac{\partial \ln \rho}{\partial \theta} = 0$$

$$\text{But } \frac{\partial \ln \rho}{\partial \theta} = I(\theta)(\hat{\theta} - \theta)$$

Since  $I(\theta) > 0$  for all  $\theta$  (we always assume this - otherwise PDF does not depend on  $\theta$ ), the only solution is  $\theta = \hat{\theta}$  or MLE is just  $\hat{\theta}$ , the efficient estimator.

13) See plots below.



We have a better fit as  $N$  increases.

14) See plots on next two pages.

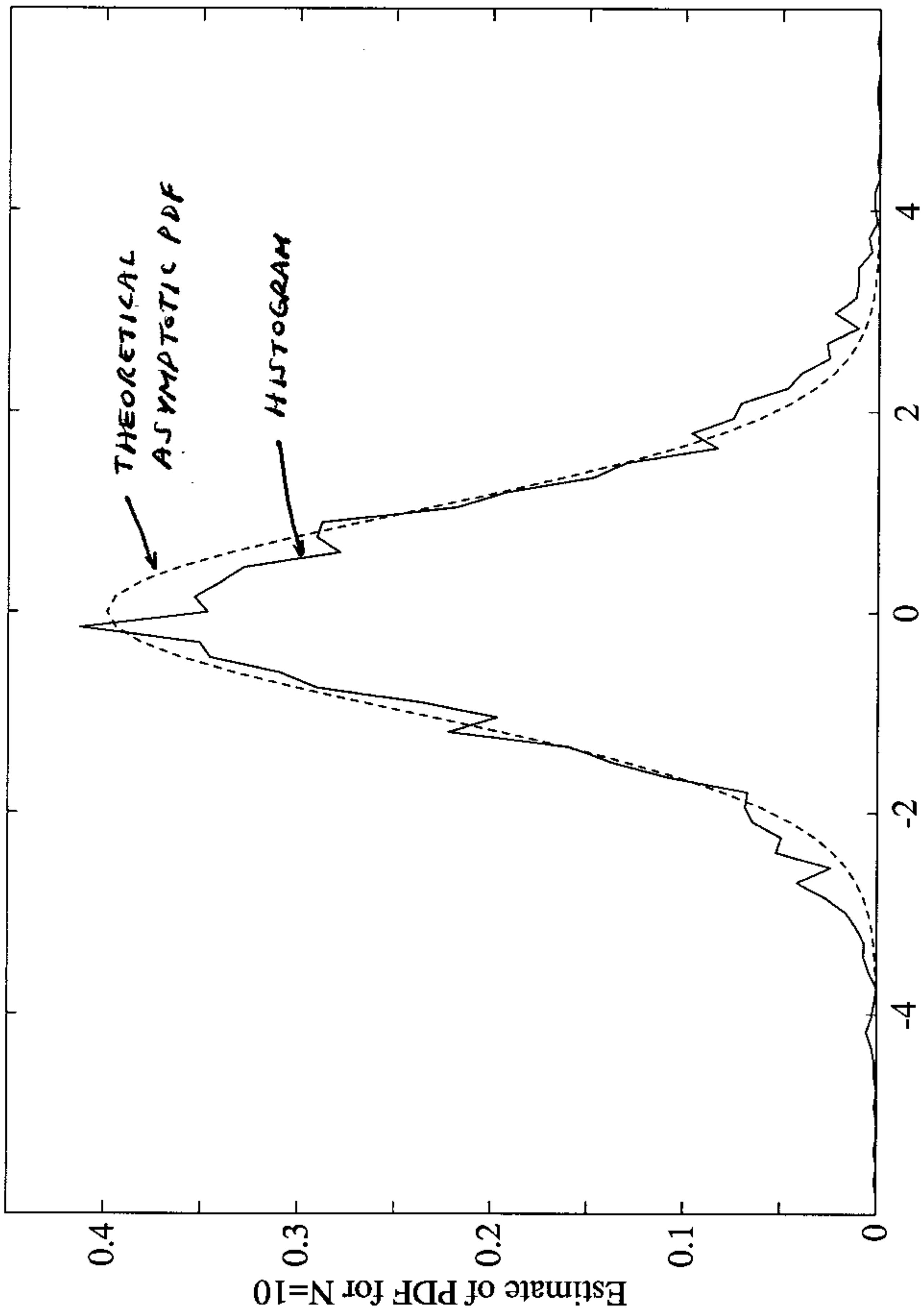
15)  $E_S, E_C$  are zero mean since  $W(n)$  is zero mean.

$$\begin{aligned} E(E_S E_C) &= -\frac{4}{N^2 A^2} \sum_m \sum_n E(W(m) W(n)) \\ &\quad \cdot \sin 2\pi f_0 m \cos 2\pi f_0 n \\ &= -\frac{4\sigma^2}{N^2 A^2} \sum_n \sin 2\pi f_0 n \cos 2\pi f_0 n \\ &= -\frac{2\sigma^2}{NA^2} \sum_n \sin 4\pi f_0 n \approx 0 \end{aligned}$$

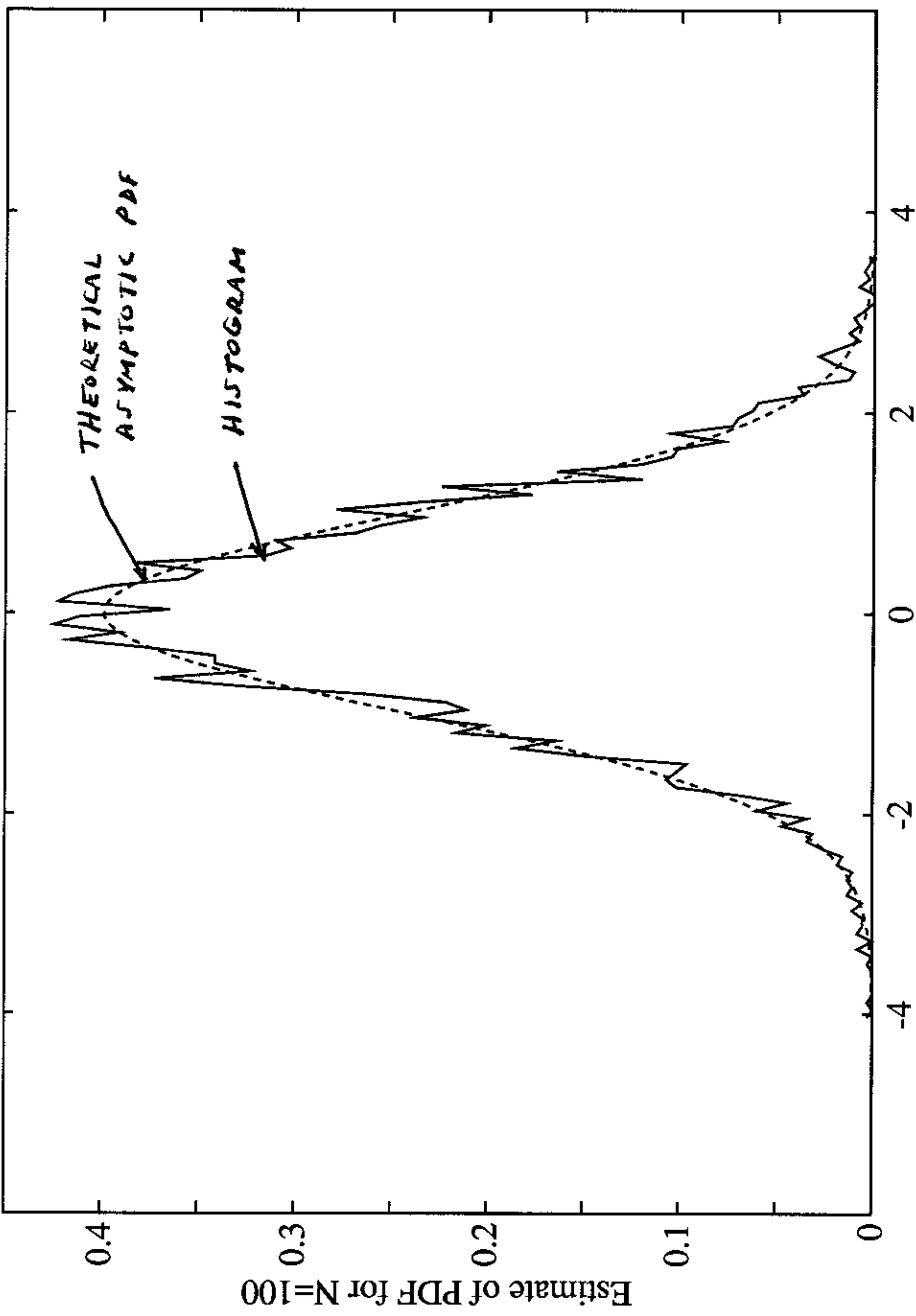
Thus,  $E_S, E_C$  are independent Gaussian random variables with zero means and variances given as follows:

$$\begin{aligned} \text{var}(E_S) &= E(E_S^2) = \frac{4}{N^2 A^2} \sum_m \sum_n E(W(m) W(n)) \\ &\quad \cdot \sin 2\pi f_0 m \sin 2\pi f_0 n \\ &= \frac{4\sigma^2}{N^2 A^2} \sum_n \sin^2 2\pi f_0 n \\ &= \frac{4\sigma^2}{N^2 A^2} \sum_n \left( \frac{1}{2} - \frac{1}{2} \cos 4\pi f_0 n \right) \\ &\approx \frac{4\sigma^2}{N^2 A^2} (N/2) = \frac{2\sigma^2}{NA^2} \end{aligned}$$

Problem 7.14: Verification of Slutsky's theorem



Problem 7.14: Verification of Slutsky's theorem



and similarly,  $\text{var}(\epsilon_c) \approx 2\sigma^2/NA^2$ .

$$g(\epsilon_s, \epsilon_c) = \arctan \frac{\sin \phi + \epsilon_s}{\cos \phi + \epsilon_c}$$

$$\frac{\partial g}{\partial \epsilon_s} = \frac{1}{1 + \left( \frac{\sin \phi + \epsilon_s}{\cos \phi + \epsilon_c} \right)^2} \frac{1}{\cos \phi + \epsilon_c}$$

$$\begin{aligned} \frac{\partial g}{\partial \epsilon_s} \Big|_{\epsilon_s = \epsilon_c = 0} &= \frac{\cos^2 \phi}{\cos^2 \phi + \sin^2 \phi} \frac{1}{\cos \phi} \\ &= \cos \phi \end{aligned}$$

$$\frac{\partial g}{\partial \epsilon_c} = \frac{1}{1 + \left( \frac{\sin \phi + \epsilon_s}{\cos \phi + \epsilon_c} \right)^2} \frac{\sin \phi + \epsilon_s}{(\cos \phi + \epsilon_c)^2}$$

$$\frac{\partial g}{\partial \epsilon_c} \Big|_{\epsilon_s = \epsilon_c = 0} = \cos^2 \phi \frac{\sin \phi}{\cos^2 \phi} = \sin \phi$$

$$\hat{\phi} \approx \phi + \cos \phi \epsilon_s + \sin \phi \epsilon_c$$

$$\Rightarrow \hat{\phi} \sim N(\phi, \underbrace{\cos^2 \phi \text{var}(\epsilon_s) + \sin^2 \phi \text{var}(\epsilon_c)}_{\frac{2\sigma^2}{NA^2}})$$

The variance is just the CRLD.

$$\begin{aligned} 16) \quad \text{var}(\hat{A}) &= \text{var}(x(0)) \\ &= \text{var}(W(0)) \\ &= \int_{-\infty}^{\infty} u^2 \frac{1}{2} e^{-tu^2} du \end{aligned}$$

$$= \int_0^\infty u^2 e^{-u} du$$

$$= - (u^2 + 2u + 2) e^{-u} \Big|_0^\infty = 2$$

$$\text{var}(\hat{\lambda}) \geq \int_{-\infty}^{\infty} \left( \frac{dp(u)}{du} \right)^2 / p(u) du$$

$$= \int_{-\infty}^{\infty} \left( \frac{1}{2} e^{-|u|} \right)^2 / \frac{1}{2} e^{-|u|} du$$

since  $\frac{dp}{du} = \begin{cases} -\frac{1}{2} e^{-u} & u > 0 \\ \frac{1}{2} e^u & u < 0 \end{cases}$

$$\text{var}(\hat{\lambda}) \geq \int_{-\infty}^{\infty} \frac{1}{2} e^{-|u|} du = 1$$

No, the MLE has variance 2 for all  $N$ .

- 17) Let  $\alpha = 1/\theta$  so that by covariance of the MLE,  $\hat{\alpha} = 1/\hat{\theta}$ . But  $\hat{\alpha} = 1/N \sum_{n=0}^{N-1} x^2(n)$   
 since
- $$p(x; \alpha) = \frac{1}{(2\pi\alpha)^{N/2}} e^{-\frac{1}{2\alpha} \sum_{n=0}^{N-1} x^2(n)}$$

$$\frac{\partial \log p}{\partial \alpha} = -\frac{N}{2} 1/\alpha + \frac{1}{2\alpha^2} \sum_{n=0}^{N-1} x^2(n) = 0$$

$$\Rightarrow \hat{\alpha} = 1/N \sum_n x^2(n)$$

$$\text{Thus, } \hat{\theta} = \frac{1}{N} \overbrace{\sum_{n=0}^{N-1} x^2(n)}^1$$

From Chapter 3

$$I(\alpha) = N/2\alpha^2$$

$$\text{and } I^{-1}(\alpha) = I^{-1}(0)(\frac{\partial \alpha}{\partial \theta})^2$$

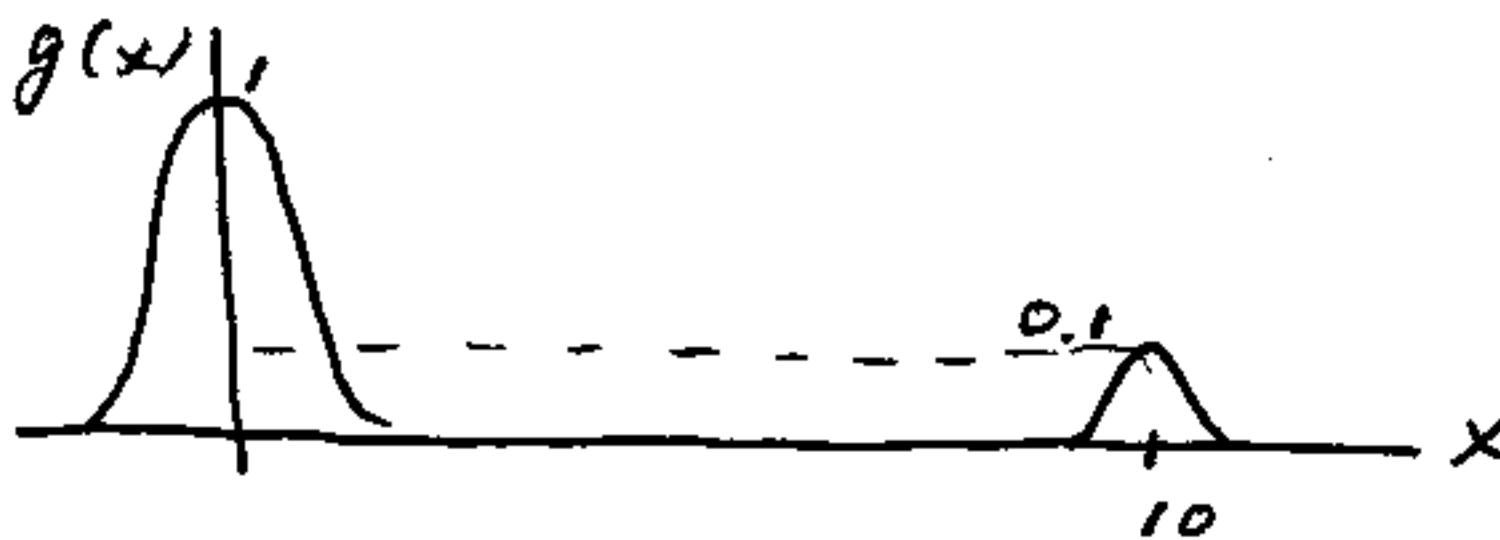
$$I^{-1}(0) = \frac{2\alpha^2}{N} \left( -\frac{\partial \theta}{\partial \alpha} \right)^2$$

$$= \frac{2\alpha^2}{N} (-1/\alpha^2)^2 = \frac{2}{N\alpha^2}$$

$$= 2\theta^2/N$$

$$\Rightarrow \hat{\theta} \approx N(\theta, 2\theta^2/N)$$

18)



$$g'(x) = -x e^{-\frac{1}{2}x^2} - 0.1(x-10) e^{-\frac{1}{2}(x-10)^2}$$

$$g''(x) = x^2 e^{-\frac{1}{2}x^2} + 0.1(x-10)^2 e^{-\frac{1}{2}(x-10)^2} - 0.1 e^{-\frac{1}{2}(x-10)^2} - e^{-\frac{1}{2}x^2}$$

$$x_{k+1} = x_k - \frac{dg/dx}{d^2g/dx^2} \Big|_{x=x_k}$$

Using a computer it is found that for  $x_0 = 0.5$ , the iteration converges to  $x = 0$  and for  $x_0 = 0.95$  it converges to  $x = 10$ . For other initial values such as  $x_0 = 1$  it does not converge. To attain the maximum (global),  $x_0$  must be close to the true value.

$$19) p(x; f_0) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} \sum_n (x[n] - \cos 2\pi f_0 n)^2}$$

To find NLE minimize

$$\begin{aligned} \sum_n (x[n] - \cos 2\pi f_0 n)^2 &= \\ &\sum_n x^2[n] - 2 \sum_n x[n] \cos 2\pi f_0 n \\ &\quad + \sum_n \cos^2 2\pi f_0 n \\ &\approx \sum_n x^2[n] - 2 \sum_n x[n] \cos 2\pi f_0 n + N/2 \end{aligned}$$

$$\Rightarrow \text{maximize } \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n$$

over  $0 < f_0 < 1/2$

Using a Newton-Raphson iteration

$$g(f_0) = \sum_n x[n] \cos 2\pi f_0 n$$

$$\frac{dg}{df_0} = - \sum_n 2\pi n x[n] \sin 2\pi f_0 n$$

$$\frac{d^2g}{df_0^2} = - \sum_n (2\pi n)^2 x[n] \cos 2\pi f_0 n$$

$$f_{0,k+1} = f_{0,k} - \frac{\sum_n (2\pi n) x[n] \sin 2\pi f_{0,k} n}{\sum_n (2\pi n)^2 x[n] \cos 2\pi f_{0,k} n} \quad |_{f_0 = f_{0,k}}$$

The function to be maximized is shown on following page for 10 different realizations of  $w[n]$ . Next, for 500 realizations we plot the results for a grid search and a Newton-Raphson iteration with 10 iterations. In (a), the grid search produces:

$$E(\hat{f}_0) = 0.25 = f_0$$

$$\text{var}(\hat{f}_0) = 1.3 \times 10^{-6}$$

In (b), (c), (d) the Newton-Raphson produces

Initial freq. ( $f_{0,i}$ )	$E(\hat{f}_{0,i})$	$\text{var}(\hat{f}_{0,i})$
0.24	0.25	$1.2 \times 10^{-6}$
0.26	0.25	$1.2 \times 10^{-6}$
0.28	0.16	10

$$20) p(x; \xi) = \frac{1}{(2\pi\sigma^2)^N} e^{-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - s[n])^2}$$

Maximized when  $\sum (x[n] - s[n])^2$  is minimized.

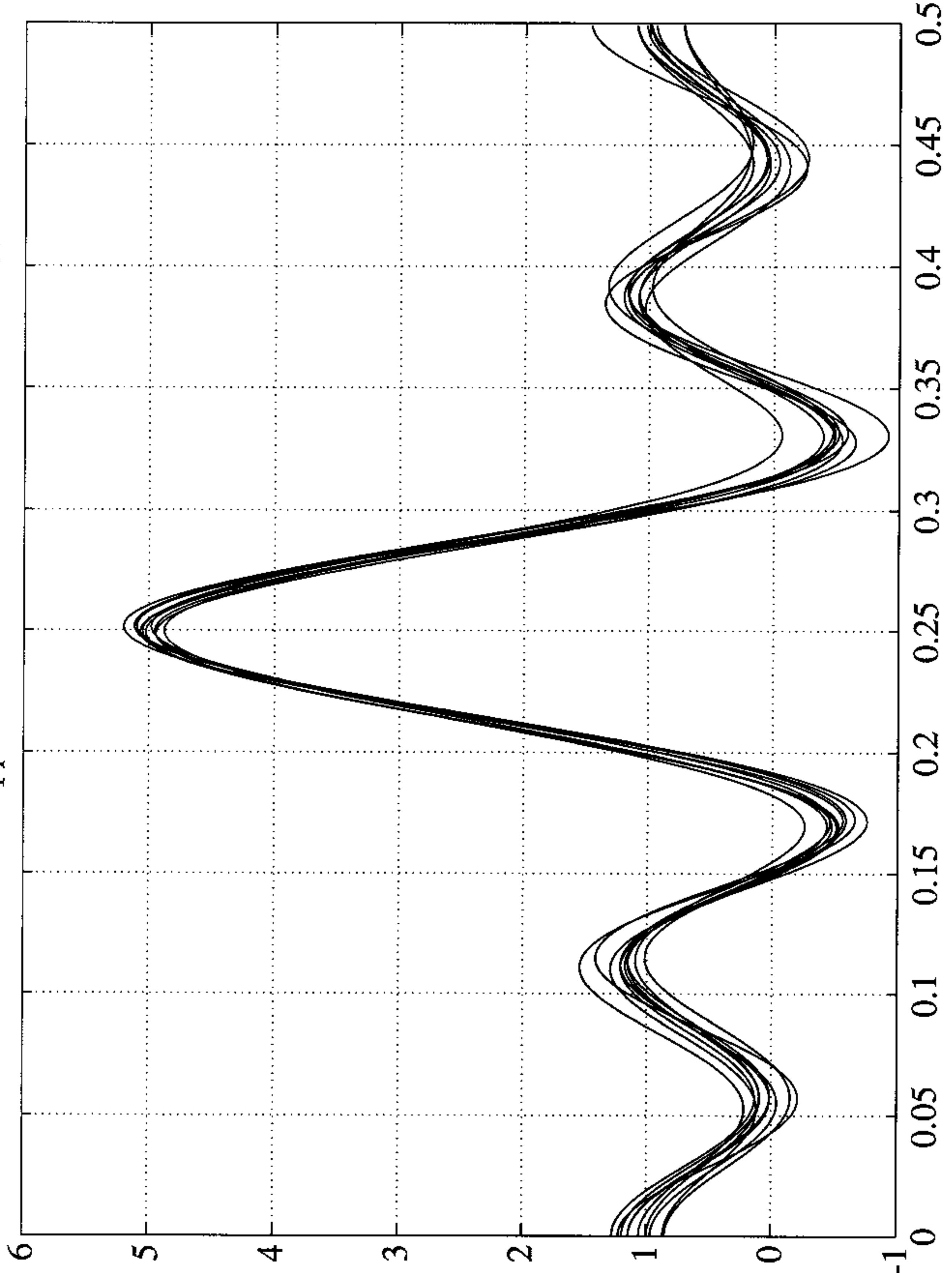
Clearly, then  $\hat{s}[n] = x[n]$ .

But  $\hat{s} = x \sim N(s, \sigma^2 I)$  and thus  $\hat{s}$  is unbiased and Gaussian. To determine if it is efficient we find the CRLB.

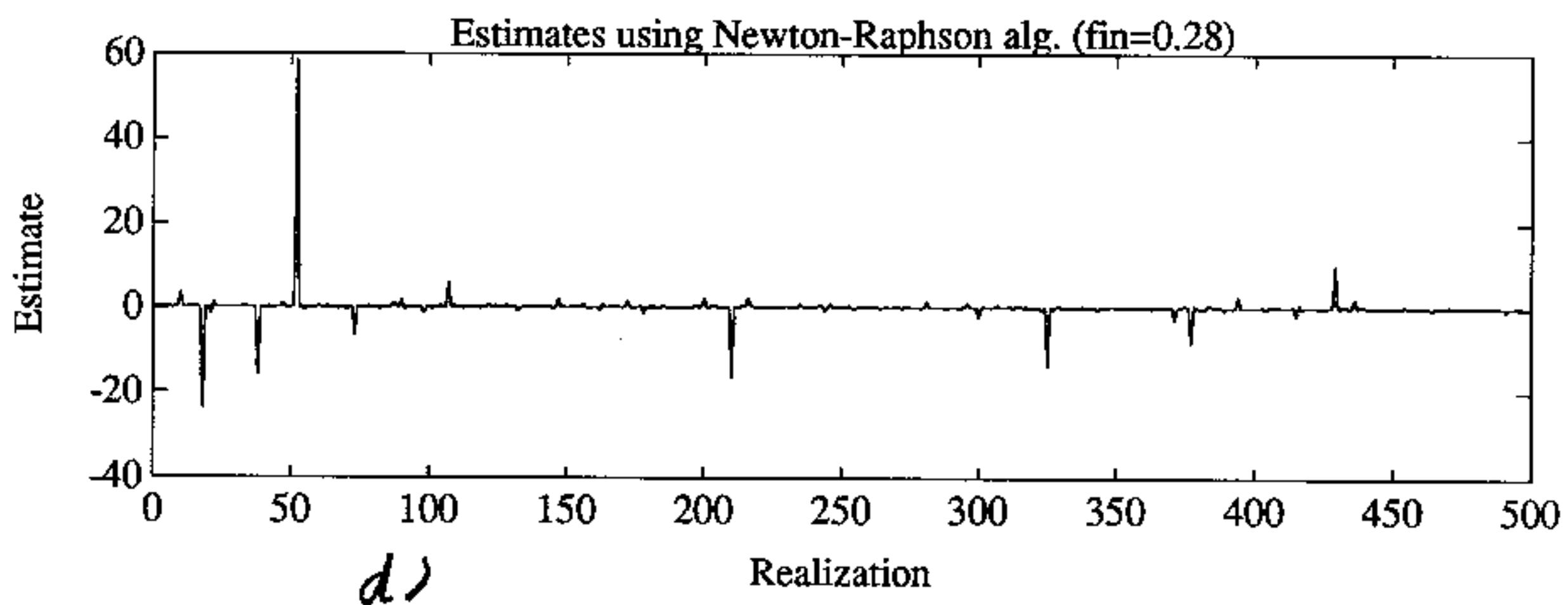
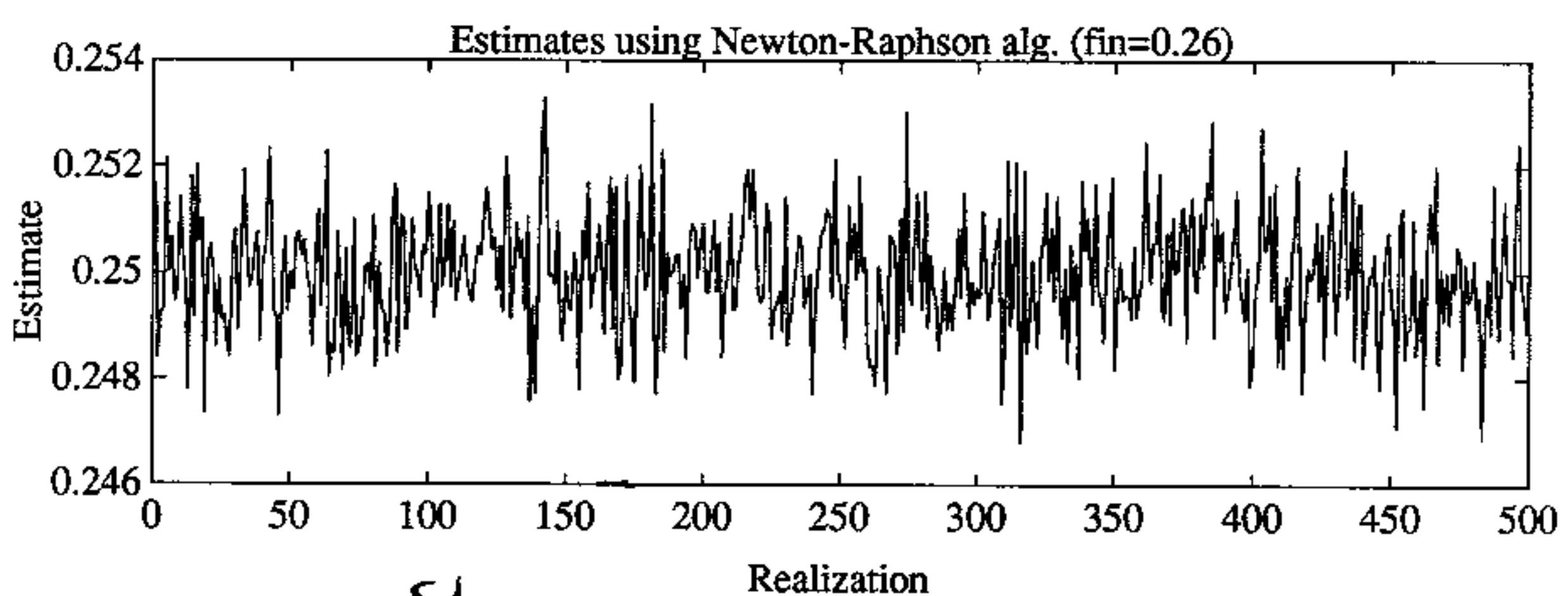
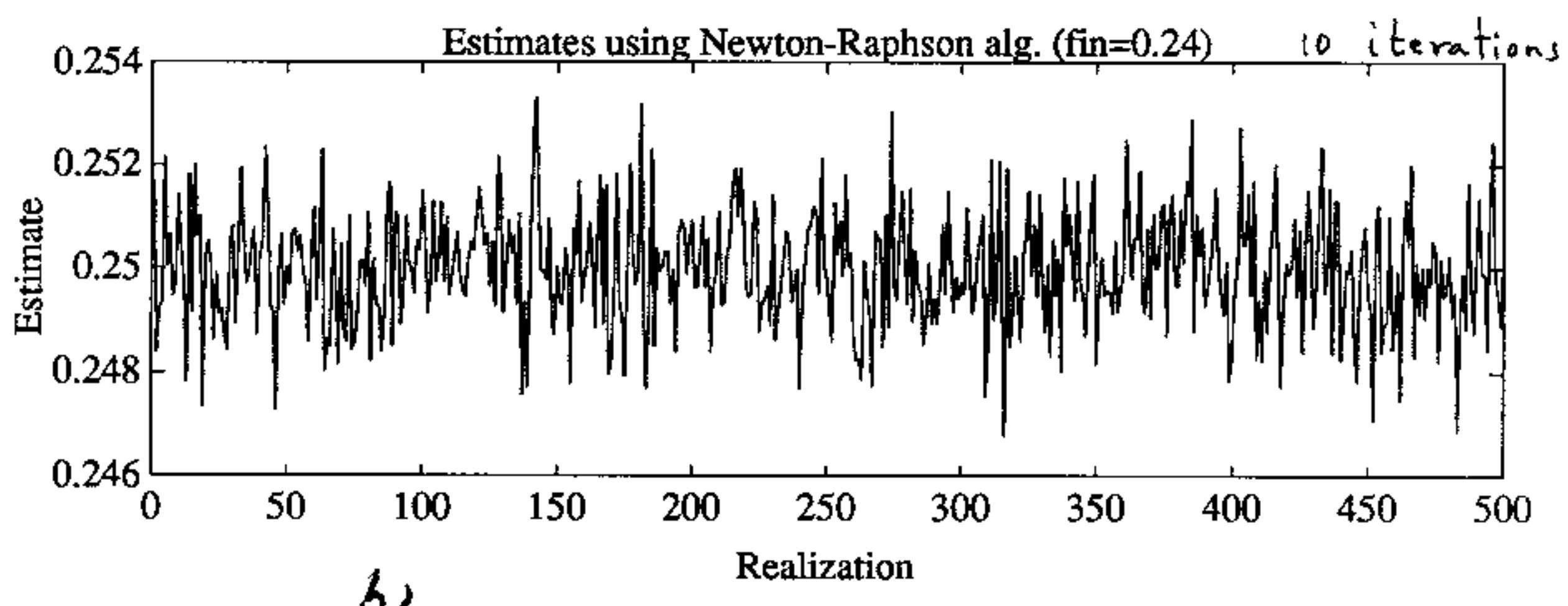
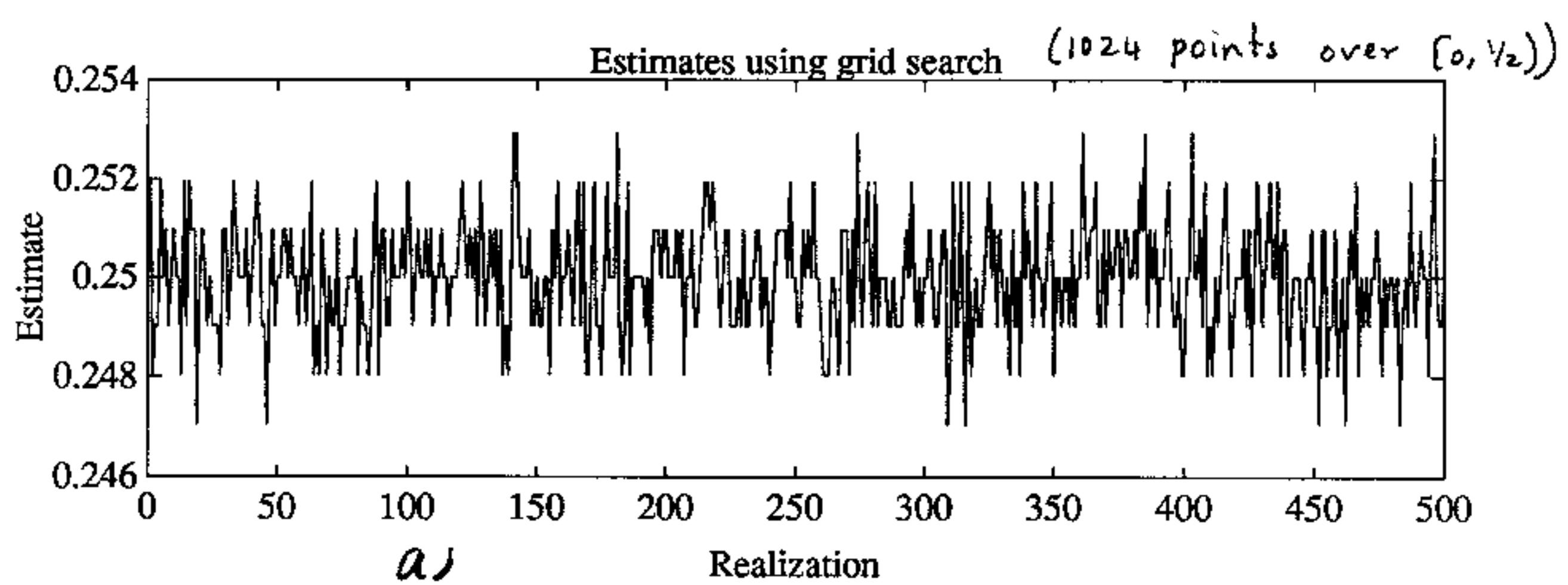
$$\frac{\partial \ln p}{\partial s[k]} = \frac{1}{\sigma^2} (x[k] - s[k])$$

$$\frac{\partial^2 \ln p}{\partial s[k] \partial s[l]} = -\frac{1}{\sigma^2} \delta_{kl}$$

Problem 7.19: Approx. MLE function to be maximized



objective function for 10 trials



$$\Rightarrow \hat{I}(s) = \frac{1}{\hat{\sigma}^2} I$$

$$\text{or } \hat{s} \sim N(s, I^{-1}(s))$$

Hence,  $\hat{s}$  is efficient. However, it is not consistent since as  $n \rightarrow \infty$ ,  $\text{var}(\hat{s}_n) = \sigma^2 \not\rightarrow 0$ .

21) From Example 7.1.2  $\hat{A} = \bar{x}$   
 $\hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} (x_n - \bar{x})^2$

Hence, by the invariance

property  $\hat{\lambda} = \frac{\hat{A}^2}{\hat{\sigma}^2} = \frac{\bar{x}^2}{\frac{1}{N} \sum_{n=0}^{N-1} (x_n - \bar{x})^2}$

22)  $I = \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln |A(t)|^2 dt = \int_{-\frac{1}{2}}^{\frac{1}{2}} [\ln A(s) + \ln A^*(s)] dt$

$$= 2 \operatorname{Re} \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln A(s) ds$$

Now let

$$z = e^{j2\pi t}, dz = j2\pi e^{j2\pi t} dt = j2\pi z dt$$

$$\Rightarrow I = 2 \operatorname{Re} \int \frac{1}{j2\pi} \frac{dz}{z} \ln A(z)$$

$$= 2 \operatorname{Re} \left\{ \frac{1}{j2\pi} \int \ln A(z) dz \right\}$$

$$= 2 \operatorname{Re} \left\{ \left. z^{-1} \{ \ln A(z) \} \right|_{z=0} \right\}$$

Since  $A(z)$  converges for all  $z \neq 0$ , it

converges for  $|z| \geq 1$  and since all of its zeros are within the unit circle,  $\ln A(z)$  converges on and outside the unit circle. Thus,  $\ln A(z)$  has a causal inverse.

The sample at  $n = \infty$  is found from the initial value theorem or

$$2^{-1} \{ \ln A(z) \}_{n=0} = \lim_{z \rightarrow \infty} \ln A(z)$$

$$= \lim_{z \rightarrow \infty} \ln \left[ 1 + \sum_{k=1}^{\infty} a[k] z^{-k} \right] = \ln 1 = 0$$

23) We use (7.60) which when differentiated produces

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \left( \frac{1}{P_{xx}(f)} - \frac{I(f)}{P_{dx}(f)} \right) \frac{\partial P_{dx}(f)}{\partial P_0} df = 0$$

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \left( \frac{1}{P_0 Q(f)} - \frac{I(f)}{P_0^2 Q^2(f)} \right) Q(f) df = 0$$

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \left( P_0 - \frac{I(f)}{Q(f)} \right) df = 0$$

$$\Rightarrow \hat{P}_0 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{I(f)}{Q(f)} df$$

If  $Q(f) = 1$  for all  $f$ ,

$$\hat{P}_0 = \int_{-\frac{1}{2}}^{\frac{1}{2}} I(f) df$$

$$= \frac{1}{N} \sum_{n=0}^{N-1} x^2(n) \text{ from Prob 7.25}$$

24) See plots on next page. For  $f_0 = 0.25$  the peak of the periodogram and that of the exact function  $\underline{x}^T H (H^T H)^{-1} H^T \underline{x}$  are at 0.25. But for  $f_0 = 0.05$  the peak of the periodogram is shifted away ( $= 0.068$ ) from the true value. This is due to the interaction of the complex sinusoids at  $f_0 = 0.05$  and  $-f_0 = -0.05$ , which are not adequately resolved by the periodogram.

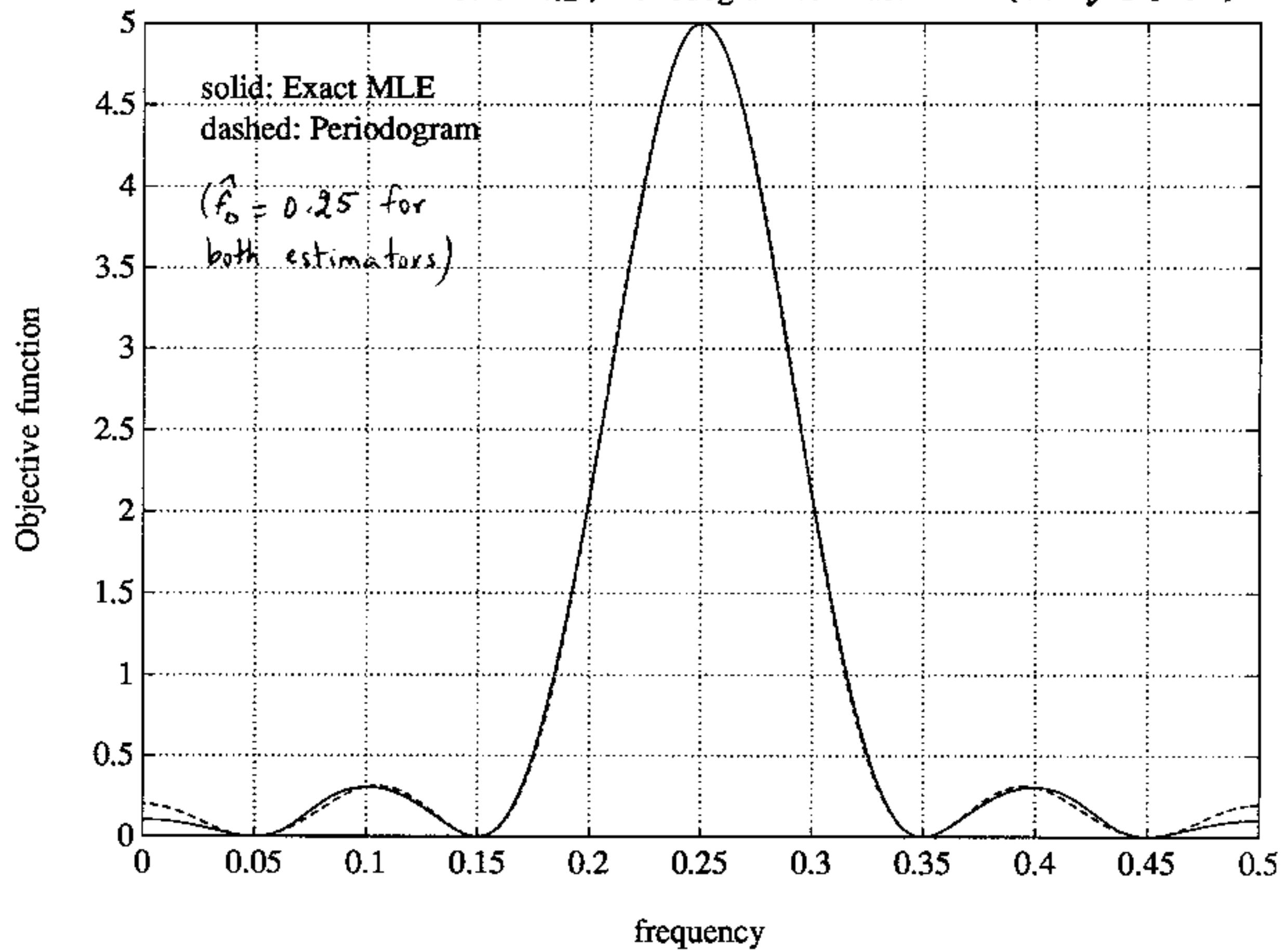
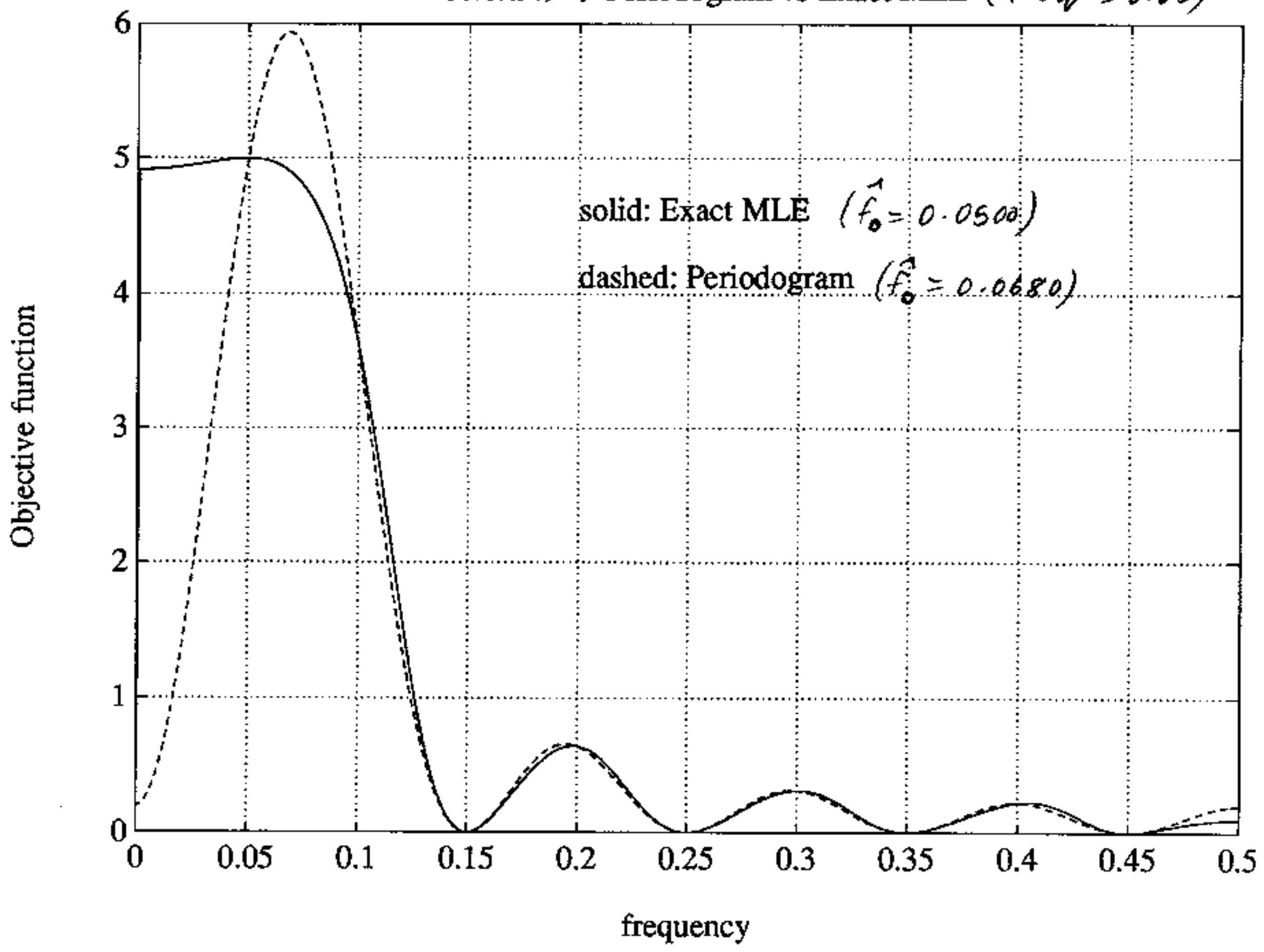
25)  $\mathcal{F}^{-1}\{I(f)\} = \frac{1}{N} \mathcal{F}^{-1}\{x'(f) x'(f)^*\}$   
 But if  $x[n] \xrightarrow{\mathcal{F}} x'(f)$   
 $x'[-n] \xrightarrow{\mathcal{F}} x'(f)^*$  for  $x[n]$  real

$$\begin{aligned} \Rightarrow \mathcal{F}^{-1}\{I(f)\} &= \frac{1}{N} \sum_{n=-\infty}^{\infty} x'[n] \star x'[-n] \\ &= \frac{1}{N} \sum_{n=-\infty}^{\infty} x'[-n] x'[k-n] \\ &= \frac{1}{N} \sum_{n=-\infty}^{\infty} x'[n] x'[n+k] \end{aligned}$$

For  $k \geq 0$

$$\mathcal{F}^{-1}\{I(f)\} = \frac{1}{N} \sum_{n=0}^{N-1-k} x[n] x[n+k]$$

since  $x'[n] = 0$  for  $n < 0$  or  $n > N-1$   
 $= x[n]$  otherwise

Problem 7.24 Periodogram vs Exact MLE ( $\text{freq} = 0.25$ )Problem 7.24 Periodogram vs Exact MLE ( $\text{freq} = 0.05$ )

For  $k < 0$

$$\mathcal{F}^{-1}\{\mathcal{I}(f)\} = \frac{1}{N} \sum_{n=-k}^{N-1} x[n] x[n+k]$$

Let  $\ell = n+k$

$$= \frac{1}{N} \sum_{\ell=0}^{N-1+k} x[\ell-k] x[\ell]$$

$$= \frac{1}{N} \sum_{n=0}^{N-1+k} x[n] x[n-k]$$

Combining the results we have our solution.

26)  $\hat{\alpha}[1] = -\hat{r}_{xx}[1]/\hat{r}_{xx}[0]$

$$\hat{\sigma}^2 = \hat{r}_{xx}[0] + \hat{\alpha}[1] \hat{r}_{xx}[1]$$

By maximizing the MLE of  $P_{xx}(f_0)$  we

$$\hat{P}_{xx}(f_0) = \frac{\hat{\sigma}^2}{(1 + \hat{\alpha}[1] e^{-j2\pi f_0})^2}$$

But

$$C_2 = \frac{\partial \mathcal{J}}{\partial \underline{\theta}} \mathcal{I}^{-1}(\underline{\theta}) \frac{\partial \mathcal{J}}{\partial \underline{\theta}}^T \geq 0$$

$$\mathcal{I}(\underline{\theta}) = \begin{bmatrix} \frac{N r_{xx}[0]}{\sigma_u^2} & 0 \\ 0 & \frac{N}{2\sigma_u^4} \end{bmatrix}$$

$$\text{and } g(\alpha[1], \sigma_u^2) = \frac{\sigma_u^2}{(1 + \alpha[1] e^{-j2\pi f_0})^2}$$

so that

$$\frac{\partial g}{\partial \alpha[1]} = \frac{-\sigma_u^2}{|A(f)|^4} (A(f) e^{j2\pi f_0} + A^*(f) e^{-j2\pi f_0})$$

$$= -\frac{\sigma_u^2}{|A(f)|^4} 2Re(A(f) e^{j2\pi f_0})$$

$$\frac{\partial g}{\partial \sigma_u^2} = \frac{1}{|A(f)|^2}$$

$$\text{var}(\hat{P}_{xx}(f_0)) = \frac{\partial g}{\partial \underline{\theta}} I^{-1}(\underline{\theta}) \frac{\partial g}{\partial \underline{\theta}}^T$$

$$= \frac{\sigma_u^2}{N r_{xx}(0)} \frac{\sigma_u^4}{|A(f)|^8} 4Re^2(A(f) e^{j2\pi f_0})$$

$$+ \frac{2\sigma_u^4}{N} \frac{1}{|A(f)|^4}$$

$$= \frac{4 \hat{P}_{xx}^4(f_0)}{N r_{xx}(0) \sigma_u^2} Re^2(A(f) e^{j2\pi f_0})$$

$$+ \frac{2}{N} \hat{P}_{xx}^2(f_0)$$

## Chapter 8

1) Nonlinear LS due to  $f_0$  and  $r$ .

Yes, quadratic in  $A, B$ .

Analytically, we could find the values of  $A$  and  $B$  that minimize  $J$  for given  $f_0$  and  $r$ . Then, plug these into  $J$ , which will now be a nonquadratic function of  $f_0$  and  $r$ . Next, use a grid search over  $0 < r < 1$  and  $0 \leq f_0 \leq \frac{1}{2}$ .

$$2) J_{\text{MIN}} = \sum_{n=0}^{N-1} x^2(n) - n \bar{x}^2 \leq \sum_{n=0}^{N-1} x^2(n)$$

$$\text{Also, } J_{\text{MIN}} = \sum_{n=0}^{N-1} (x(n) - \bar{x})^2 \geq 0$$

$$3) J = \sum_{n=0}^{N-1} (x(n) - A)^2 + \sum_{n=M}^{N-1} (x(n) + A)^2$$

$$\frac{\partial J}{\partial A} = -2 \sum_{n=0}^{N-1} (x(n) - A) + 2 \sum_{n=M}^{N-1} (x(n) + A) = 0$$

$$-2 \sum_{n=0}^{M-1} x(n) + 2MA + 2 \sum_{n=M}^{N-1} x(n) + 2(N-M)A = 0$$

$$\hat{A} = \frac{1}{N} \left( \sum_{n=0}^{M-1} x(n) - \sum_{n=M}^{N-1} x(n) \right)$$

$$J_{\text{MIN}} = \sum_{n=0}^{M-1} (x(n) - \hat{A})(x(n) - \hat{A}) + \sum_{n=M}^{N-1} (x(n) + \hat{A})(x(n) + \hat{A})$$

$$= \sum_{n=0}^{M-1} x(n)(x(n) - \hat{A}) + \sum_{n=M}^{N-1} x(n)(x(n) + \hat{A})$$

using the  $\partial J/\partial A = 0$  equation.

$$\begin{aligned} J_{M,N} &= \sum_0^{N-1} x^2(n) - \hat{A} \left( \sum_0^{M-1} x(n) - \frac{\sum_0^{N-1} x(n)}{N} \right) \\ &= \sum_0^{N-1} x^2(n) - N \hat{A}^2 \end{aligned}$$

For  $w(n)$ ,  $w_{6N}$  we have

$$\begin{aligned} E(\hat{A}) &= \frac{1}{N} [MA - (N-M)A] = A \\ \text{var}(\hat{A}) &= \frac{1}{N^2} \left( \text{var}\left(\sum_0^{M-1} x(n)\right) \right. \\ &\quad \left. + \text{var}\left(\frac{\sum_0^{N-1} x(n)}{N}\right) \right) \\ &= \frac{1}{N^2} \left( M\sigma^2 + (N-M)\sigma^2 \right) = \sigma^2/N \end{aligned}$$

$\Rightarrow \hat{A} \sim N(A, \sigma^2/N)$  since  $\hat{A}$  is a linear function of the  $x(n)$ 's.

4) See solution for Prob 5, 14

$$5) \quad \underline{x} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \cos 2\pi f_1 n, \cos 2\pi f_2 n, \dots, \cos 2\pi f_p n \\ \vdots \\ \underbrace{\cos 2\pi f_1(N-1), \cos 2\pi f_2(N-1), \dots, \cos 2\pi f_p(N-1)}_H \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_p \end{bmatrix} \underbrace{\theta}_{\Omega}$$

$$\underline{H}^T \underline{\underline{\theta}} = \underline{H}^T \underline{x} \quad \text{are normal equations}$$

If  $f_i = i/N$ , the column vectors of  $\underline{H}$  are orthogonal (see (4.13)). Thus,

$$\underline{H}^T \underline{H} = \frac{N}{2} \underline{\underline{I}}$$

$$\Rightarrow \hat{\underline{\theta}} = \frac{2}{N} \underline{H}^T \underline{x} \Rightarrow \hat{\theta}_i = \frac{2}{N} \sum_{n=0}^{N-1} x(n) \cos 2\pi f_i n$$

$$T_{MIN} = \underline{x}^T (\underline{\underline{I}} - \underline{H}(\underline{H}^T \underline{H})^{-1} \underline{H}^T) \underline{x}$$

$$= \underline{x}^T (\underline{\underline{I}} - \frac{2}{N} \underline{H} \underline{H}^T) \underline{x}$$

$$= \underline{x}^T \underline{x} - \frac{2}{N} \| \underline{H}^T \underline{x} \|^2$$

$$= \underline{x}^T \underline{x} - \frac{2}{N} \left( \frac{N}{2} \right)^2 \| \hat{\underline{\theta}} \|^2$$

$$= \sum_{n=0}^{N-1} x^2(n) - \frac{N}{2} \sum_{i=1}^p \hat{\theta}_i^2$$

For  $w[n], w \in \mathbb{N}$  the PDF is

$$\hat{\underline{\theta}} \sim N(\underline{\theta}, \sigma^2 (\underline{H}^T \underline{H})^{-1}) \text{ since}$$

$$E(\hat{\underline{\theta}}) = (\underline{H}^T \underline{H})^{-1} \underline{H}^T E(\underline{x}) = (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{\theta} = \underline{\theta}$$

$$C\hat{\underline{\theta}} = E[(\hat{\underline{\theta}} - \underline{\theta})(\hat{\underline{\theta}} - \underline{\theta})^T]$$

$$= E[(\underline{H}^T \underline{H})^{-1} \underline{H}^T (\underbrace{\underline{x} - \underline{H}\underline{\theta}}_w) (\underbrace{\underline{x} - \underline{H}\underline{\theta}}_w)^T \underline{H} (\underline{H}^T \underline{H})^{-1}]$$

$$= (\underline{H}^T \underline{H})^{-1} \underline{H}^T \sigma^2 \underline{\underline{\Xi}} \underline{H} (\underline{H}^T \underline{H})^{-1} = \sigma^2 (\underline{H}^T \underline{H})^{-1}$$

$$\text{or } \underline{\hat{\theta}} = 2\sigma^2/N \underline{\underline{\Xi}}$$

and since  $\underline{\hat{\theta}}$  is a linear function of  $\underline{x}$ , we have a Gaussian PDF or

$$\underline{\hat{\theta}} \sim N(\underline{\theta}, 2\sigma^2/N \underline{\underline{\Xi}})$$

6)  $\hat{\underline{\theta}} = (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x}$

From Prob 8.5 we have  $\hat{\underline{\theta}} \sim N(\underline{\theta}, \sigma^2 (\underline{H}^T \underline{H})^{-1})$   
Yes, it is unbiased.

7)  $E(\hat{\sigma}^2) = \frac{1}{N} E [\underline{x}^T (\underline{\underline{\Xi}} - \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T) \underline{x}]$   
 $= \frac{1}{N} \text{tr} [(\underline{\underline{\Xi}} - \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T) E(\underline{x} \underline{x}^T)]$   
 since  $E(\underline{x} \underline{x}^T) = \text{tr}(E(\underline{x} \underline{x}^T))$

$$E(\hat{\sigma}^2) = \frac{1}{N} \text{tr} [(\underline{\underline{\Xi}} - \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T) E((H\underline{\theta} + \underline{w})(H\underline{\theta} + \underline{w})^T)]$$

$$= \frac{1}{N} \text{tr} [(\underline{\underline{\Xi}} - \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T) (H\underline{\theta} \underline{\theta}^T H^T + \sigma^2 \underline{\underline{\Xi}})]$$

$$= \frac{1}{N} \text{tr} [H \underline{\theta} \underline{\theta}^T H^T + \sigma^2 \underline{\underline{\Xi}} - \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T H \underline{\theta} \underline{\theta}^T H^T - \sigma^2 \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T]$$

$$\begin{aligned}
 &= \frac{\sigma^2}{N} \text{tr} (\underline{\mathbf{I}} - \underline{\mathbf{H}}(\underline{\mathbf{H}}^T \underline{\mathbf{H}})^{-1} \underline{\mathbf{H}}^T) \\
 &= \sigma^2 - \frac{\sigma^2}{N} \text{tr} [\underline{\mathbf{H}}(\underline{\mathbf{H}}^T \underline{\mathbf{H}})^{-1} \underline{\mathbf{H}}^T] \\
 &= \sigma^2 - \sigma^2/N \text{tr} [\underbrace{(\underline{\mathbf{H}}^T \underline{\mathbf{H}})^{-1} \underline{\mathbf{H}}^T}_{\mathbf{I} (p \times p)}] \\
 &= \sigma^2 - \sigma^2 p/N = \sigma^2 \frac{N-p}{N}
 \end{aligned}$$

$\hat{\sigma}^2$  is biased. To make it unbiased use

$$\hat{\sigma}^2 = \frac{1}{N-p} J_{ML}$$

It is said that we lose  $p$  degrees of freedom in estimating  $\underline{\theta}$ .

$$\begin{aligned}
 \text{Q) } J(A) &= \sum_{n=0}^{N-1} \frac{1}{\sigma_n^2} (x(n)-A)^2 \\
 \frac{\partial J}{\partial A} &= -2 \sum_n \frac{1}{\sigma_n^2} (x(n)-A) = 0 \\
 \Rightarrow \hat{A} &= \frac{\sum_{n=0}^{N-1} x(n)/\sigma_n^2}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}}
 \end{aligned}$$

$$E(A) = \frac{\sum_n E(x(n))/\sigma_n^2}{\sum_n 1/\sigma_n^2} = A$$

$$\begin{aligned}
 \text{var}(\hat{\theta}) &= \frac{1}{\left(\sum_n \sigma_n^2\right)^2} \underbrace{\sum_n \text{var}(x_{n1})/\sigma_n^2}_{= \sum_n 1/\sigma_n^2 \text{ var}(x_{n1})} \\
 &= \sum_{n=0}^{N-1} 1/\sigma_n^2
 \end{aligned}$$

$$\begin{aligned}
 9) \quad J &= (\underline{x} - \underline{H}\underline{\theta})^T \underline{W} (\underline{x} - \underline{H}\underline{\theta}) \\
 &= (\underline{x} - \underline{H}\underline{\theta})^T \underline{D}^T \underline{D} (\underline{x} - \underline{H}\underline{\theta}) \\
 &= (\underline{D}\underline{x} - \underline{D}\underline{H}\underline{\theta})^T (\underline{D}\underline{x} - \underline{D}\underline{H}\underline{\theta}) \\
 &\quad \underline{x}' \quad \underline{H}' 
 \end{aligned}$$

$$\begin{aligned}
 \Rightarrow \hat{\underline{\theta}} &= (\underline{H}'^T \underline{H}')^{-1} \underline{H}'^T \underline{x}' \\
 &= (\underline{H}^T \underline{D}^T \underline{D} \underline{H})^{-1} \underline{H}^T \underline{D}^T \underline{D} \underline{x} \\
 &= (\underline{H}^T \underline{W} \underline{H})^{-1} \underline{H}^T \underline{W} \underline{x}
 \end{aligned}$$

$$\begin{aligned}
 J_{MIN} &= \underline{x}'^T (\underline{I} - \underline{H}' (\underline{H}'^T \underline{H}')^{-1} \underline{H}'^T) \underline{x}' \\
 &= \underline{x}'^T \underline{D}' (\underline{I} - \underline{D} \underline{H} (\underline{H}^T \underline{D}^T \underline{D} \underline{H})^{-1} \underline{H}^T \underline{D}^T) \underline{D} \underline{x} \\
 &= \underline{x}'^T (\underline{H} - \underline{W} \underline{H} (\underline{H}^T \underline{W} \underline{H})^{-1} \underline{H}^T \underline{W}) \underline{x}
 \end{aligned}$$

10)  $\hat{\underline{x}} = \underline{H}\hat{\underline{\theta}} = \underline{H}(\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x} = \underline{P} \underline{x}$

$$\underline{x} - \hat{\underline{x}} = \underline{x} - \underline{P}\underline{x} = (\underline{I} - \underline{P})\underline{x}$$

$$\begin{aligned} \|\hat{\underline{x}}\|^2 + \|\underline{x} - \hat{\underline{x}}\|^2 &= \underline{x}^T \underline{P}^T \underline{P} \underline{x} \\ &\quad + \underline{x}^T (\underline{I} - \underline{P})^T (\underline{I} - \underline{P}) \underline{x} \\ &= \underline{x}^T \underline{P} \underline{x} + \underline{x}^T (\underline{I} - \underline{P}) \underline{x} = \underline{x}^T \underline{x} = \|\underline{x}\|^2 \end{aligned}$$

Since  $\underline{P}$  and  $(\underline{I} - \underline{P})$  are symmetric and idempotent.

$$(1) \quad \underline{x}_1 = \underline{\beta}_1 + \underline{\beta}_1^\perp \quad \underline{x}_2 = \underline{\beta}_2 + \underline{\beta}_2^\perp$$

$$\begin{aligned} l &= \underline{x}_1^T \underline{P} \underline{x}_2 - \underline{x}_2^T \underline{P} \underline{x}_1 = (\underline{\beta}_1 + \underline{\beta}_1^\perp)^T \underline{P} (\underline{\beta}_2 + \underline{\beta}_2^\perp) \\ &\quad - (\underline{\beta}_2 + \underline{\beta}_2^\perp)^T \underline{P} (\underline{\beta}_1 + \underline{\beta}_1^\perp) \\ &= (\underline{\beta}_1 + \underline{\beta}_1^\perp)^T \underline{P} \underline{\beta}_2 - (\underline{\beta}_2 + \underline{\beta}_2^\perp)^T \underline{P} \underline{\beta}_1 \end{aligned}$$

Since  $\underline{P} \underline{\beta}_1^\perp = 0$

$$l = (\underline{\beta}_1 + \underline{\beta}_1^\perp)^T \underline{\beta}_2 - (\underline{\beta}_2 + \underline{\beta}_2^\perp)^T \underline{\beta}_1$$

since  $\underline{P} \underline{\beta}_2 = \underline{\beta}_2$

$$l = \underline{\beta}_1^T \underline{\beta}_2 - \underline{\beta}_2^T \underline{\beta}_1 = 0 \quad \text{since } \underline{\beta}_1^T \underline{\beta}_1 = 0$$

Now let  $\underline{x}_1 = \underline{e}_i = [0 \ 0 \dots 0 \underset{i}{1} 0 \dots 0]^T$   
 $\underline{x}_2 = \underline{e}_j$   $\uparrow i \neq j$  place

$$\underline{e}_i^T \underline{P} \underline{e}_j = \underline{e}_j^T \underline{P} \underline{e}_i \quad \text{from previous result}$$

$$\Rightarrow (\underline{P})_{ij} = (\underline{P})_{ji}$$

$$\begin{aligned} 12) \text{ a) } \underline{P}^2 &= \underline{H}(\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{A}(\underline{H}^T \underline{H})^{-1} \underline{H}^T \\ &= \underline{H}(\underline{H}^T \underline{H})^{-1} \underline{H}^T = \underline{P} \end{aligned}$$

$$\begin{aligned} \text{b) } \underline{x}^T \underline{P} \underline{x} &= \underline{x}^T \underline{P} \underline{P} \underline{x} = \underline{x}^T \underline{P}^T \underline{P} \underline{x} \\ &= (\underline{P} \underline{x})^T \underline{P} \underline{x} = \|\underline{P} \underline{x}\|^2 \geq 0 \end{aligned}$$

$$\begin{aligned} \text{c) } \underline{P} \underline{x} &= \underline{\lambda} \underline{x} \Rightarrow \underline{P} \underline{x} = \underline{\lambda} \underline{P} \underline{x} = \underline{\lambda} \underline{P} \underline{x} \\ &= \underline{\lambda} (\underline{P} \underline{x}) = \underline{\lambda}^2 \underline{x} \end{aligned}$$

$\Rightarrow$  If  $\lambda$  is an eigenvalue, so is  $\lambda^2$ .

By uniqueness  $\lambda^2 = \lambda$  or  $\lambda = 0, 1$

d) Rank = sum of nonzero eigenvalues.

To show that there are  $p$  nonzero eigenvalues, we find  $\text{tr}(\underline{P}) = \sum_{i=1}^n \lambda_i$

$$\begin{aligned} \text{tr}(\underline{P}) &= \text{tr}(\underline{H}(\underline{H}^T \underline{H})^{-1} \underline{H}^T) \\ &= \text{tr}((\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{H}) \\ &= \text{tr}(\underline{I}) = p \end{aligned}$$

$$\begin{aligned} \Rightarrow \sum_{i=1}^p \lambda_i &= p \Rightarrow p \text{ eigenvalues} = 1 \\ &\Rightarrow \text{rank} = p \end{aligned}$$

$$(13) \quad \hat{\theta} = (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x}$$

where  $\underline{H} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & N-1 \end{bmatrix}$

$$\underline{H}^T \underline{H} = \begin{bmatrix} N & \sum n \\ \sum n & \sum n^2 \end{bmatrix} = \begin{bmatrix} N & N(N-1)/2 \\ N(N-1)/2 & \frac{N(N-1)(2N-1)}{6} \end{bmatrix}$$

using (3.22)

$$(\underline{H}^T \underline{H})^{-1} = \frac{1}{\frac{N^2(N-1)(2N-1)}{6} - \frac{N^2(N-1)^2}{4}} \begin{bmatrix} \frac{N(N-1)(2N-1)}{6} & -\frac{N(N-1)}{2} \\ -\frac{N(N-1)}{2} & N \end{bmatrix}$$

$$\hat{\theta} = \begin{bmatrix} \frac{2(2N-1)}{N(N+1)} & \frac{-6}{N(N+1)} \\ \frac{-6}{N(N+1)} & \frac{12}{N(N^2-1)} \end{bmatrix} \begin{bmatrix} \sum_n x(n) \\ \sum n x(n) \end{bmatrix}$$

which produces (8.23).

$$(14) \quad \underline{h}'_{k+1} \perp \{ \underline{h}_1, \underline{h}_2, \dots, \underline{h}_k \}$$

$$\text{Let } P_k = \underline{H}_k (\underline{H}_k^T \underline{H}_k)^{-1} \underline{H}_k^T$$

$$\text{where } \underline{H}_k = [ \underline{h}_1 \dots \underline{h}_k ]$$

$$\text{Then, } \underline{h}'_{k+1} = \underline{h}_{k+1} - P_k \underline{h}_{k+1} \quad (\text{See Fig 8.8})$$

$$= P_k^\perp \underline{h}_{k+1}$$

Since  $\underline{h}_{k+1} \perp \{\underline{h}_1, \dots, \underline{h}_k\}$  it is also  $\perp$  to  $\hat{s}_k \Rightarrow$  we can project  $\underline{x}$  onto  $\underline{h}_{k+1}$  and add the result to  $\hat{s}_k$ .

$$\alpha \underline{h}_{k+1} = \frac{\underline{x}^T \underline{h}_{k+1}}{\|\underline{h}_{k+1}\|} \frac{\underline{h}_{k+1}}{\|\underline{h}_{k+1}\|}$$

Since  $\frac{\underline{h}_{k+1}}{\|\underline{h}_{k+1}\|}$  is a unit vector in the  $\underline{h}_{k+1}$  direction

$$\text{or } \alpha = \frac{\underline{x}^T \underline{h}_{k+1}}{\|\underline{h}_{k+1}\|^2} = \frac{\underline{x}^T P_k^\perp \underline{h}_{k+1}}{\|P_k^\perp \underline{h}_{k+1}\|^2}$$

$$\hat{s}_{k+1} = \underline{H}_k \hat{\theta}_k + \frac{\underline{x}^T P_k^\perp \underline{h}_{k+1}}{\underline{h}_{k+1}^T P_k^\perp \underline{h}_{k+1}} P_k^\perp \underline{h}_{k+1}$$

$$= \underline{H}_k \hat{\theta}_k + \frac{(\underline{I} - P_k) \underline{h}_{k+1} \underline{h}_{k+1}^T P_k^\perp \underline{x}}{\underline{h}_{k+1}^T P_k^\perp \underline{h}_{k+1}}$$

$$= \underline{H}_k \hat{\theta}_k + \frac{\underline{h}_{k+1} \underline{h}_{k+1}^T P_k^\perp \underline{x}}{\underline{h}_{k+1}^T P_k^\perp \underline{h}_{k+1}} - \frac{\underline{H}_k (\underline{H}_k^T \underline{H}_k)^{-1} \underline{H}_k^T}{\cdot \underline{h}_{k+1}^T P_k^\perp \underline{h}_{k+1}} \cdot \frac{\underline{h}_{k+1} \underline{h}_{k+1}^T P_k^\perp \underline{x}}{\underline{h}_{k+1}^T P_k^\perp \underline{h}_{k+1}}$$

$$= [\underline{H}_k \underline{h}_{k+1}] \left[ \begin{array}{c} \hat{\theta}_k - \frac{(\underline{H}_k^T \underline{H}_k)^{-1} \underline{H}_k^T}{\cdot \underline{h}_{k+1}^T P_k^\perp \underline{h}_{k+1}} \cdot \frac{\underline{h}_{k+1} \underline{h}_{k+1}^T P_k^\perp \underline{x}}{\underline{h}_{k+1}^T P_k^\perp \underline{h}_{k+1}} \\ \hline \underline{h}_{k+1}^T P_k^\perp \underline{x} \end{array} \right]$$

$$= \underline{H}_{n+1} \hat{\underline{\theta}}_{k+1}$$

(15) Clearly,  $\hat{A}_1 = \bar{x}$ . From (8.28) with  
 $\underline{H}_1 = \underline{I}$      $\underline{h}_2 = [1 \ r \ \dots \ r^{N-1}]^T = \underline{h}$

$$\hat{A}_2 = \hat{A}_1 - \frac{\underline{1}^T \underline{h} \underline{h}^T \underline{P}_1^{-1} \underline{x}}{\underline{h}^T \underline{P}_1^{-1} \underline{h}}$$

$$\hat{B}_2 = \frac{\underline{h}^T \underline{P}_1^{-1} \underline{x}}{\underline{h}^T \underline{P}_1^{-1} \underline{h}}$$

$$\begin{aligned}\underline{P}_1^{-1} &= \underline{I} - \underline{1}(\underline{1}^T \underline{1})^{-1} \underline{1}^T = \underline{I} - \frac{1}{N} \underline{1} \underline{1}^T \\ \underline{h}^T \underline{P}_1^{-1} \underline{x} &= \underline{h}^T (\underline{x} - \bar{x} \underline{1}) \\ &= \sum_n x(n) r^n - \bar{x} \sum_n r^n\end{aligned}$$

$$\begin{aligned}\underline{h}^T \underline{P}_1^{-1} \underline{h} &= \underline{h}^T (\underline{I} - \frac{1}{N} \underline{1} \underline{1}^T) \underline{h} \\ &= \underline{h}^T \underline{h} - \frac{1}{N} (\underline{1}^T \underline{h})^2 \\ &= \sum_n r^{2n} - \frac{1}{N} (\sum_n r^n)^2\end{aligned}$$

$$\hat{A}_2 = \bar{x} - \frac{\frac{1}{N} \sum_n r^n \left[ \sum_n x(n) r^n - \bar{x} \sum_n r^n \right]}{\sum_n r^{2n} - \frac{1}{N} (\sum_n r^n)^2}$$

$$\hat{B}_2 = \frac{\sum_n x(n) r^n - \bar{x} \sum_n r^n}{\sum_n r^{2n} - \frac{1}{N} (\sum_n r^n)^2}$$

$$(16) \quad \underline{H}_1 = [1, 1, \dots, 1]^T \Rightarrow \hat{\underline{A}}_1 = (\underline{H}_1^T \underline{H}_1)^{-1} \underline{H}_1^T \underline{x} = \bar{x}$$

$$J_{M, \underline{A}_1} = \sum_{n=0}^{N-1} x^2(n) - N \bar{x}^2$$

$$\underline{H}_2 = \left[ \begin{array}{c} 1 \\ 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \end{array} \right]_{(N-M) \times 1}^T \quad \text{for } \underline{\theta}_2 = \begin{bmatrix} A \\ B-A \end{bmatrix}$$

$\uparrow \quad \uparrow$   
 $\underline{H}_1 \quad \underline{H}_2$

$$\underline{P}_1^\perp = \underline{I} - \frac{1}{N} \underline{1} \underline{1}^T$$

$$\underline{h}_2^T \underline{P}_1^\perp \underline{x} = \underline{h}_2^T (\underline{x} - \frac{1}{N} \underline{1} \underline{1}^T \underline{x})$$

$$= \underline{h}_2^T (\underline{x} - \bar{x} \underline{1})$$

$$= \sum_{n=M}^{N-1} (x(n) - \bar{x}) = \frac{1}{N-M} \sum_{n=M}^{N-1} x(n) - (N-M) \bar{x}$$

$$\underline{h}_2^T \underline{P}_1^\perp \underline{h}_2 = \underline{h}_2^T \underline{h}_2 - \underline{h}_2^T \underline{1} \underline{1}^T \underline{h}_2 / N$$

$$= (N-M) - (N-M)^2 / N$$

$$= (N-M) M / N$$

$$\Rightarrow \hat{\underline{A}}_2 = \hat{\underline{A}}_1 - \frac{\frac{1}{N} \underline{1}^T \underline{h}_2 [\sum_{n=M}^{N-1} x(n) - (N-M) \bar{x}]}{(N-M) M / N}$$

$$= \bar{x} - \frac{N-M}{N} \frac{[\sum_{n=M}^{N-1} x(n) - (N-M) \bar{x}]}{(N-M) M / N}$$

$$= \bar{x} - \frac{1}{M} \sum_{n=0}^{M-1} x(n) + \frac{N}{M} \bar{x} - \bar{x} = \frac{1}{M} \sum_{n=0}^{M-1} x(n)$$

$$\hat{B-A} = \frac{\sum_{n=0}^{M-1} x(n) - (N-M) \bar{x}}{(N-M) M / N}$$

$$\text{Let } \bar{x}_1 = \frac{1}{N-M} \sum_{n=0}^{N-1} x(n)$$

$$\hat{B}-A = \frac{N}{M} (\bar{x}_1 - \bar{x})$$

$$\hat{\underline{\theta}}_2 = \begin{bmatrix} \frac{1}{M} \sum_{n=0}^{N-1} x(n) \\ \frac{N}{M} (\bar{x}_1 - \bar{x}) \end{bmatrix}$$

The decrease in  $T_{N,N}$  is from (8.31)

$$\frac{(\underline{h}_2^T \underline{P}_1^{-1} \underline{x})^2}{\underline{h}_2^T \underline{P}_1^{-1} \underline{h}_2} = \underline{h}_2^T \underline{P}_1^{-1} \underline{x} (\hat{\underline{\theta}}_2)_2$$

$$= \left( \sum_{n=0}^{N-1} x(n) - (N-M)\bar{x} \right) (\hat{\underline{\theta}}_2)_2$$

$$= (N-M)(\bar{x}_1 - \bar{x}) \frac{N}{M} (\bar{x}_1 - \bar{x})$$

$$= \frac{N}{M} (N-M) (\bar{x}_1 - \bar{x})^2$$

To detect a jump (positive or negative) we test to see if the L.S. error decreases significantly for the second-order model or if  $(\bar{x}_1 - \bar{x})^2$  is large. For no jump we expect  $\bar{x}_1 \approx \bar{x}$  but for a jump or  $A \neq B$ ,  $(\bar{x}_1 - \bar{x})^2$  will be large. Also, note that the decrease in  $T_{N,N}$  is just  $(N-M) \frac{M}{N} (\hat{B}-A)^2$ .

- 17) Wish to show that  $\underline{h}_i^T \underline{L} \underline{k}^2 \underline{x} = 0$  for  $i = 1, 2, \dots, k$

$$\begin{bmatrix} \underline{h}_k^T \underline{P}_k^{-1} \underline{x} \\ \vdots \\ \underline{h}_{k+1}^T \underline{P}_k^{-1} \underline{x} \end{bmatrix} = \begin{bmatrix} \underline{h}_k^T \\ \vdots \\ \underline{h}_{k+1}^T \end{bmatrix} \underline{P}_k^{-1} \underline{x} = \underline{H}_k^T \underline{P}_k^{-1} \underline{x}$$

$$\begin{aligned} &= \underline{H}_k^T (\underline{\underline{I}} - \underline{H}_k (\underline{H}_k^T \underline{H}_k)^{-1} \underline{H}_k^T) \underline{x} \\ &= (\underline{H}_k^T - \underline{H}_k^T) \underline{x} = \underline{0} \end{aligned}$$

$$\begin{aligned} 18) \quad J_{MIN_{k+1}} &= \underline{x}^T \underline{P}_{k+1}^{-1} \underline{x} \\ &= \underline{x}^T \underline{x} - \underline{x}^T \underline{P}_{k+1} \underline{x} \\ &= \underline{x}^T \underline{x} - \underline{x}^T \underline{P}_k \underline{x} - \underline{x}^T \frac{\underline{P}_k^{-1} \underline{h}_{k+1} \underline{h}_{k+1}^T \underline{P}_k^{-1} \underline{x}}{\underline{h}_{k+1}^T \underline{P}_k^{-1} \underline{h}_{k+1}} \\ &= J_{MIN_k} - \frac{(\underline{h}_{k+1}^T \underline{P}_k^{-1} \underline{x})^2}{\underline{h}_{k+1}^T \underline{P}_k^{-1} \underline{h}_{k+1}} \end{aligned}$$

$$\begin{aligned} 19) \quad J_{MIN}[N] &= \sum_0^N \frac{1}{\sigma_n^2} (x[n] - \hat{A}[n])^2 \\ &= \sum_0^{N-1} \frac{1}{\sigma_n^2} [x[n] - \hat{A}[n-1] - K[N](x[N] - \hat{A}[N-1])]^2 \\ &\quad + \frac{1}{\sigma_N^2} (x[N] - \hat{A}[N])^2 \\ &= \sum_0^{N-1} \frac{1}{\sigma_n^2} (x[n] - \hat{A}[n-1])^2 - 2K[N](x[N] - \hat{A}[N-1]) \\ &\quad \cdot \sum_0^{N-1} \frac{1}{\sigma_n^2} (x[n] - \hat{A}[n-1]) \\ &\quad + K^2[N](x[N] - \hat{A}[N-1])^2 \sum_0^{N-1} \frac{1}{\sigma_n^2} \\ &\quad + \frac{1}{\sigma_N^2} [x[N] - \hat{A}[N-1] - K[N](x[N] - \hat{A}[N-1])]^2 \end{aligned}$$

But  $\sum_0^{N-1} \frac{1}{\sigma_n^2} (x[n] - \hat{A}[n-1]) = 0$

due to definition of  $\hat{A}[N-1]$ .

$$\begin{aligned}
 J_{N,N}[N] &= J_{N,N}[N-1] + \frac{K^2 L_N}{\text{var}(\hat{A}[N-1])} (x[N] - \hat{A}[N-1])^2 \\
 &\quad + \frac{1}{\sigma_N^2} (x[N] - \hat{A}[N-1])^2 - 2 \frac{KL_N}{\sigma_N^2} (x[N] - \hat{A}[N-1])^2 \\
 &\quad + \frac{1}{\sigma_N^2} K^2 L_N (x[N] - \hat{A}[N-1])^2 \\
 &= J_{N,N}[N-1] + (x[N] - \hat{A}[N-1])^2
 \end{aligned}$$

where  $J = \frac{K^2 L_N}{\text{var}(\hat{A}[N-1])} + \frac{1}{\sigma_N^2} - \frac{2KL_N}{\sigma_N^2} + \frac{K^2 L_N}{\sigma_N^2}$

but from (8.38),  $1 - K[N] = \frac{\sigma_N^2}{\text{var}(\hat{A}[N-1]) + \sigma_N^2}$

$$\begin{aligned}
 J &= \frac{\text{var}(\hat{A}[N-1])}{(\text{var}(\hat{A}[N-1]) + \sigma_N^2)^2} + \frac{1}{\sigma_N^2} \frac{\sigma_N^4}{(\text{var}(\hat{A}[N-1]) + \sigma_N^2)^2} \\
 &= \frac{1}{\text{var}(\hat{A}[N-1]) + \sigma_N^2}
 \end{aligned}$$

20)  $H[n] = \begin{bmatrix} 1 \\ r \\ \vdots \\ r^n \end{bmatrix} \Rightarrow h[n] = r^n$

$$\begin{aligned}
 \hat{A}[n] &= \hat{A}[n-1] + K[n] (x[n] - r^n \hat{A}[n-1]) \\
 &\text{from (8.46)}
 \end{aligned}$$

Now  $\sigma_n^2 = \sigma^2 = 1$  and  $\Sigma(n) = \text{var}(\hat{A}(n))$

$$\Rightarrow K(n) = \frac{\text{var}(\hat{A}(n-1)) r^n}{1 + r^{2n} \text{var}(\hat{A}(n-1))} \quad \text{from (8.47)}$$

$$\text{var}(\hat{A}(n)) = (1 - K(n)r^n) \text{var}(\hat{A}(n-1)) \quad \text{from (8.48)}$$

To find the variance explicitly let

$$N_n = \text{var}(\hat{A}(n))$$

$$\begin{aligned} N_n &= \left[ 1 - \frac{N_{n-1} r^{2n}}{1 + r^{2n} N_{n-1}} \right] N_{n-1} \\ &= \frac{N_{n-1}}{1 + r^{2n} N_{n-1}} \end{aligned}$$

Now let  $N_0 = 1$  as was given

$$\begin{aligned} N_1 &= \frac{1}{1+r^2} \quad N_2 = \frac{\frac{1}{1+r^2}}{1+r^4\left(\frac{1}{1+r^2}\right)} \\ &= \frac{1}{1+r^2+r^4} \end{aligned}$$

Or in general

$$N_n = \text{var}(\hat{A}(n)) = \frac{1}{\sum_{k=0}^n r^{2k}}$$

$$21) \quad K(N) = \frac{\text{var}(\hat{A}(N-1))}{\text{var}(\hat{A}(N-1)) + \sigma_N^2}$$

$$\text{var}(\hat{A}(N)) = (1 - K(N)) \text{var}(\hat{A}(N-1))$$

$$\text{Let } N_N = \text{var}(\hat{A}(N))$$

$$N_N = \left(1 - \frac{N_{N-1}}{N_{N-1} + \sigma_N^2}\right) N_{N-1}$$

$$= \frac{\sigma_N^2 N_{N-1}}{N_{N-1} + \sigma_N^2}$$

$$\frac{1}{N_N} = \frac{N_{N-1} + \sigma_N^2}{\sigma_N^2 N_{N-1}} = \frac{1}{\sigma_N^2} + \frac{1}{N_{N-1}}$$

$$= \frac{1}{r^N} + \frac{1}{N_{N-1}}$$

Since  $\frac{1}{N_0} = 1$ , we have

$$\frac{1}{N_N} = \sum_{n=0}^N \frac{1}{r^n} \quad \text{or} \quad N_N = \frac{1}{\sum_{n=0}^N \frac{1}{r^n}}$$

$$K(N) = \frac{\frac{1}{\sum_{n=0}^N \frac{1}{r^n}}}{\frac{1}{\sum_{n=0}^N \frac{1}{r^n} + r^N}} = \frac{1}{1 + r^N \sum_{n=0}^N \frac{1}{r^n}}$$

$$= \frac{1}{1 + \sum_{n=0}^N r^{N-n}} = \frac{1}{1 + \sum_{n=0}^N r^n}$$

If  $r = 1$ ,  $\text{var}(\hat{A}(N)) \rightarrow 0$  as  $N \rightarrow \infty$

$K(N) \rightarrow 0$  as  $N \rightarrow \infty$

If  $0 < r < 1$ ,  $\text{var}(\hat{A}(N)) \rightarrow 0$  as  $N \rightarrow \infty$   
 $K(N) \rightarrow \text{Constant}$  as  $N \rightarrow \infty$

If  $r > 1$ ,  $\text{var}(\hat{A}(N)) \rightarrow \frac{r-1}{r}$  as  $N \rightarrow \infty$

since  $\sum_{n=0}^{\infty} \frac{1}{r^n} = \frac{1}{1-\frac{1}{r}} = \frac{r}{r-1}$

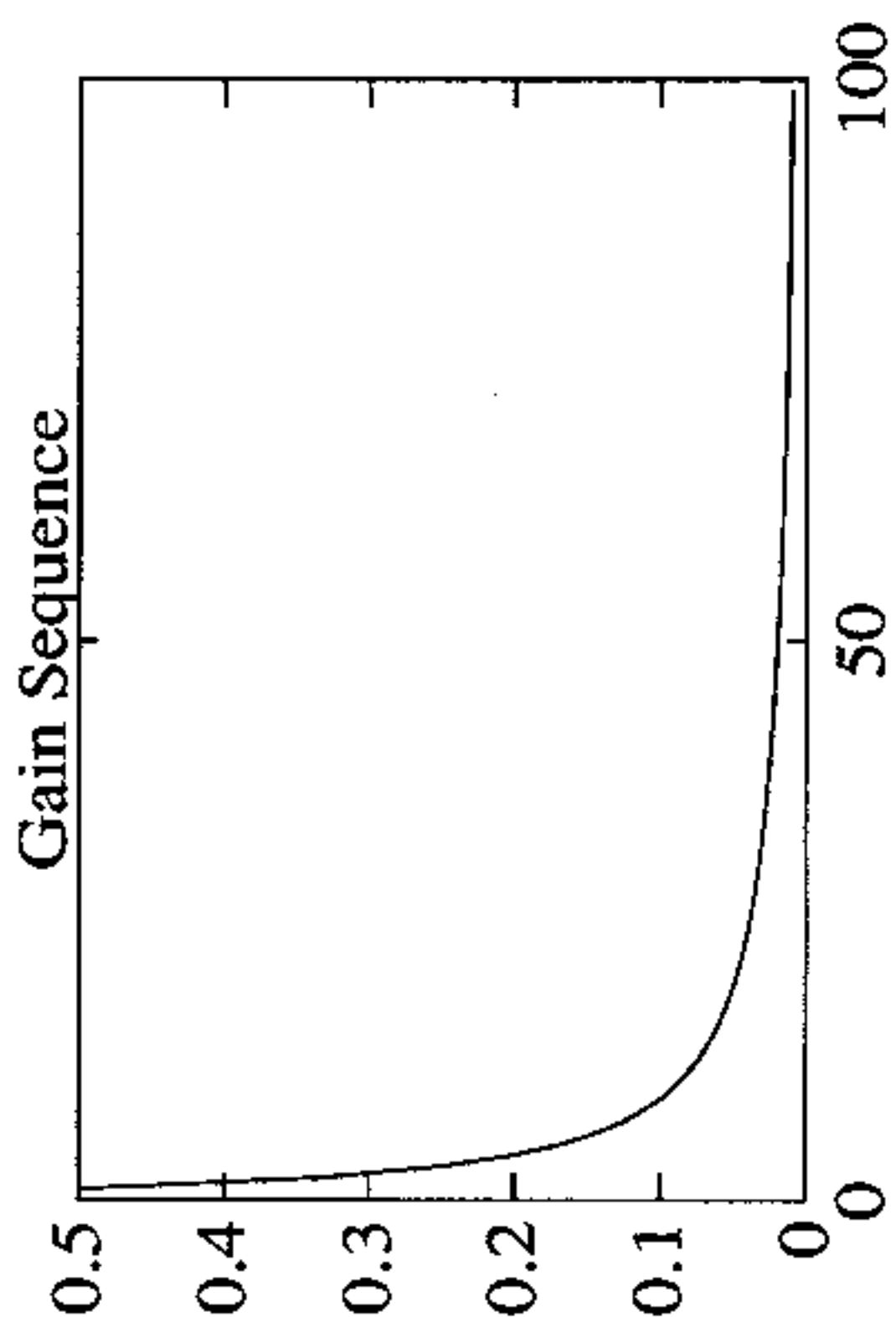
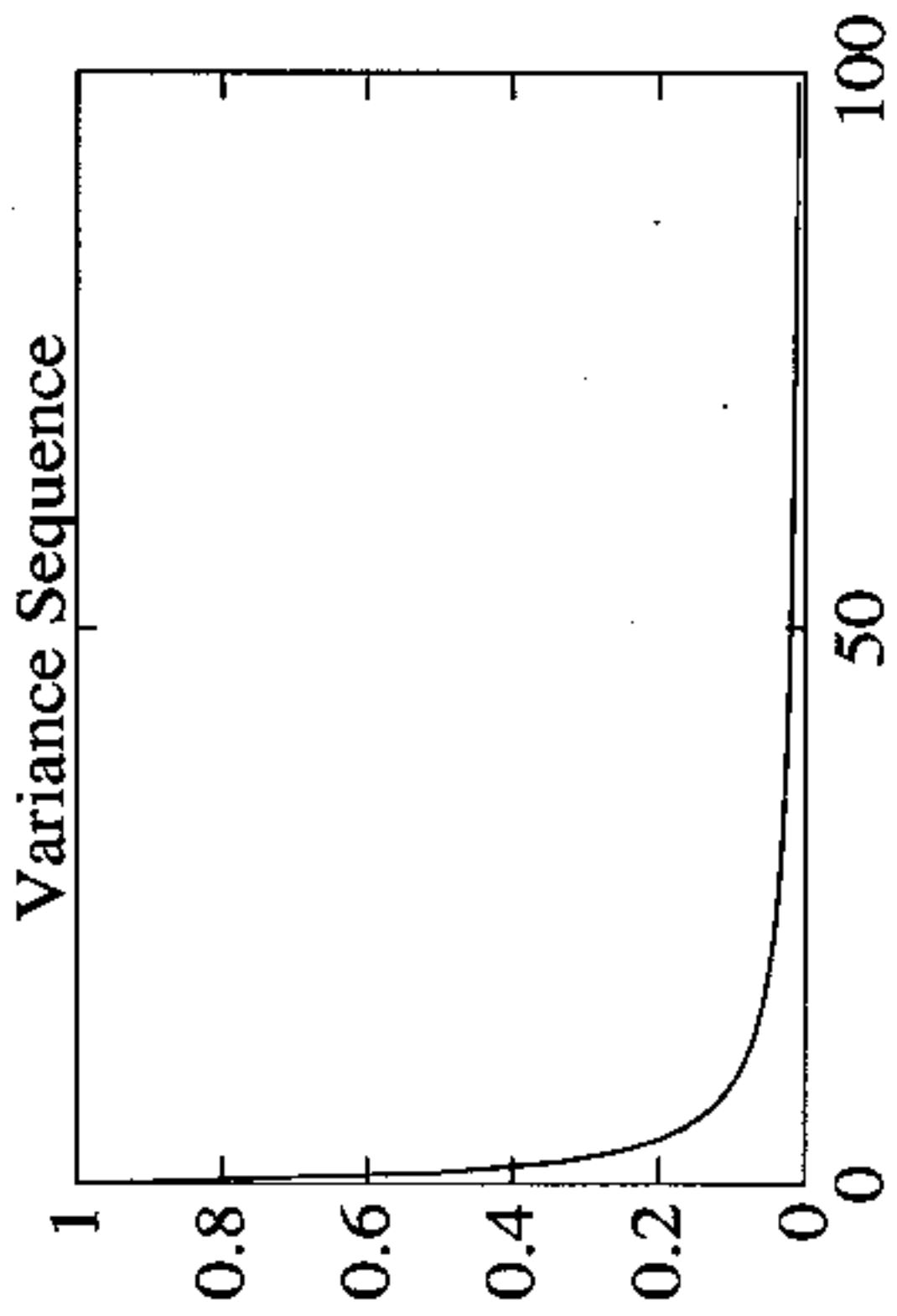
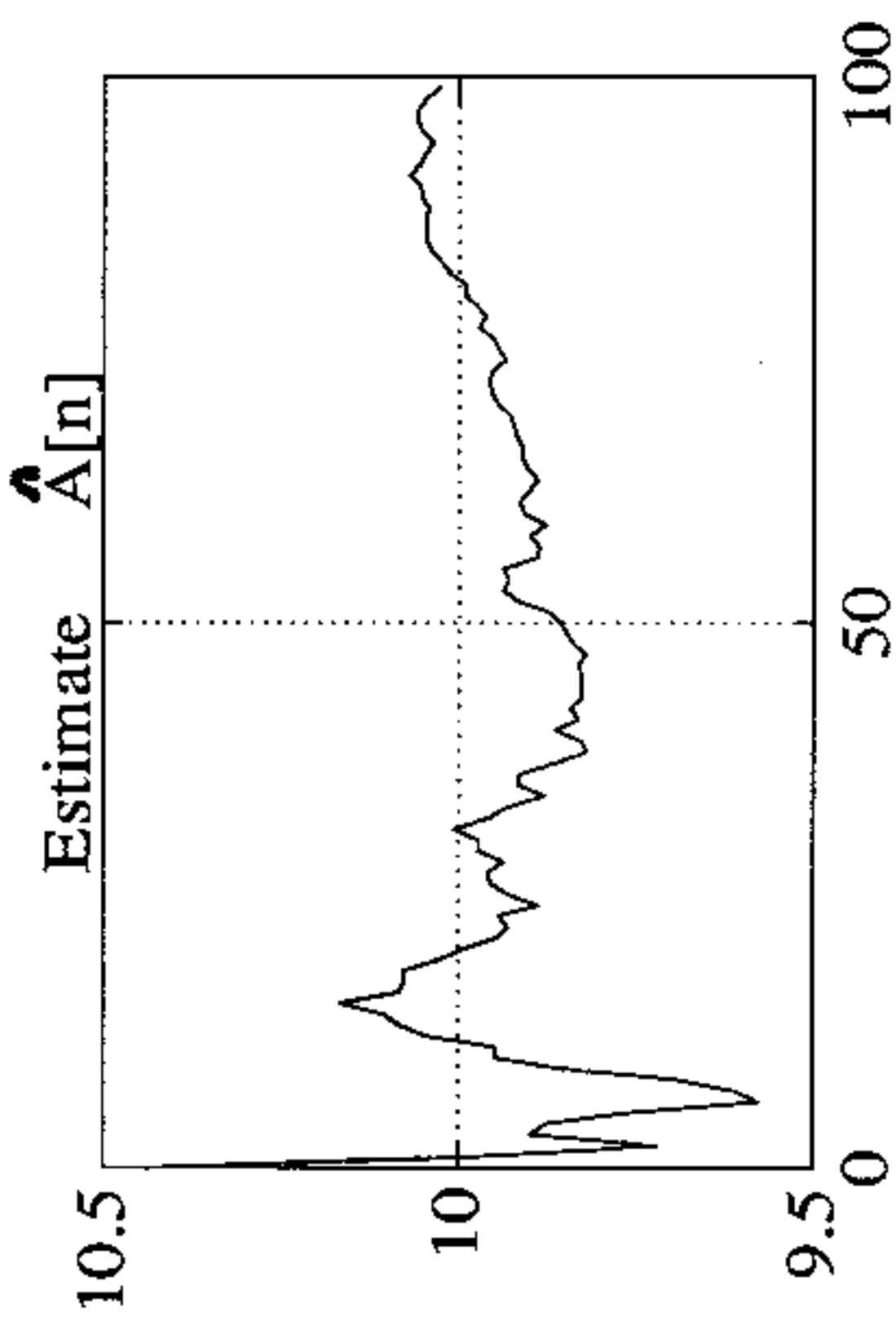
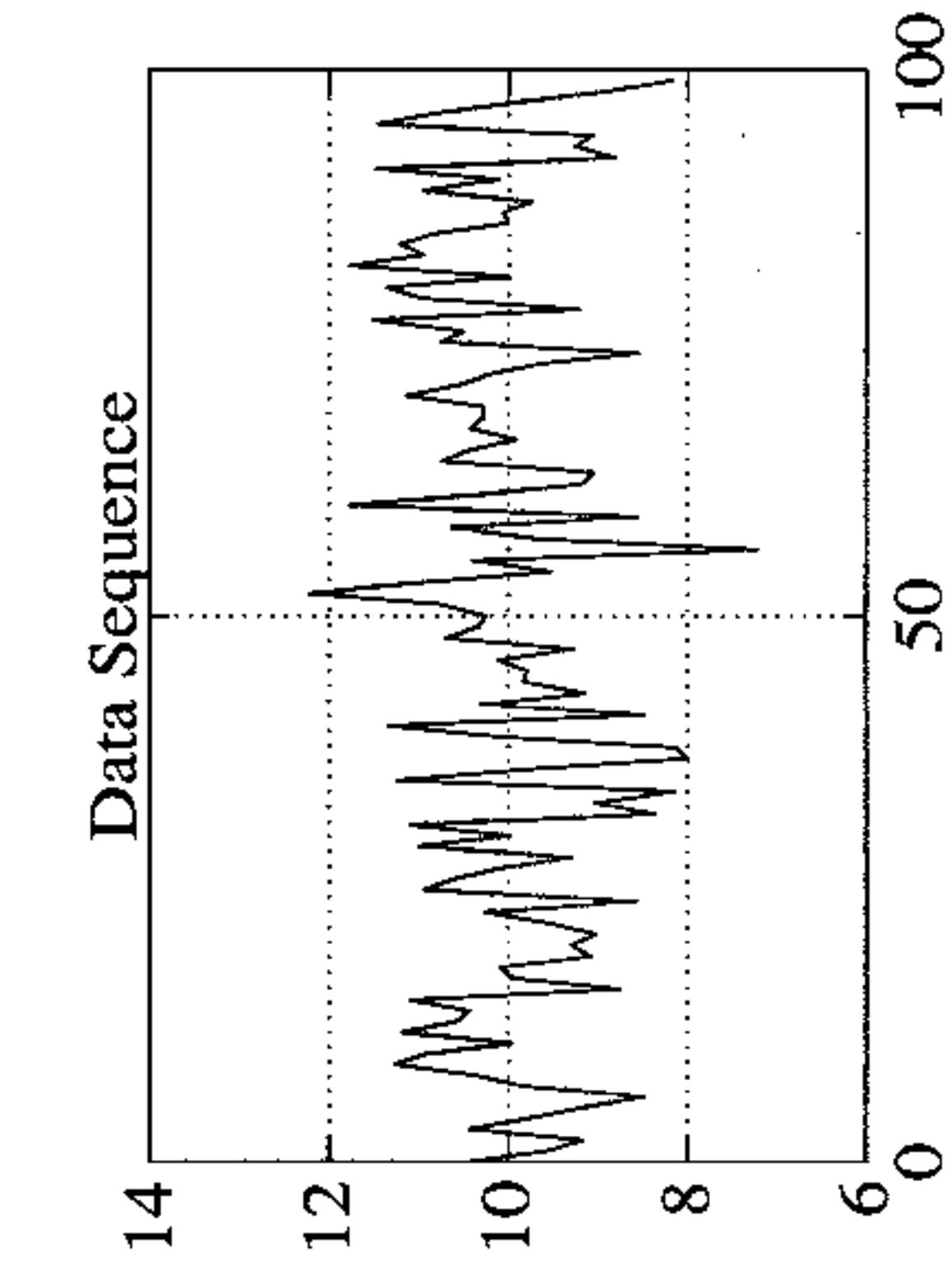
and  $K(N) \rightarrow 0$  as  $N \rightarrow \infty$ . In this case the data become so noisy that the gain  $\rightarrow 0$  (we do not use the data) and hence the variance does not go to zero.

22) See plots on next few pages. Compare these results to those of the previous problem.

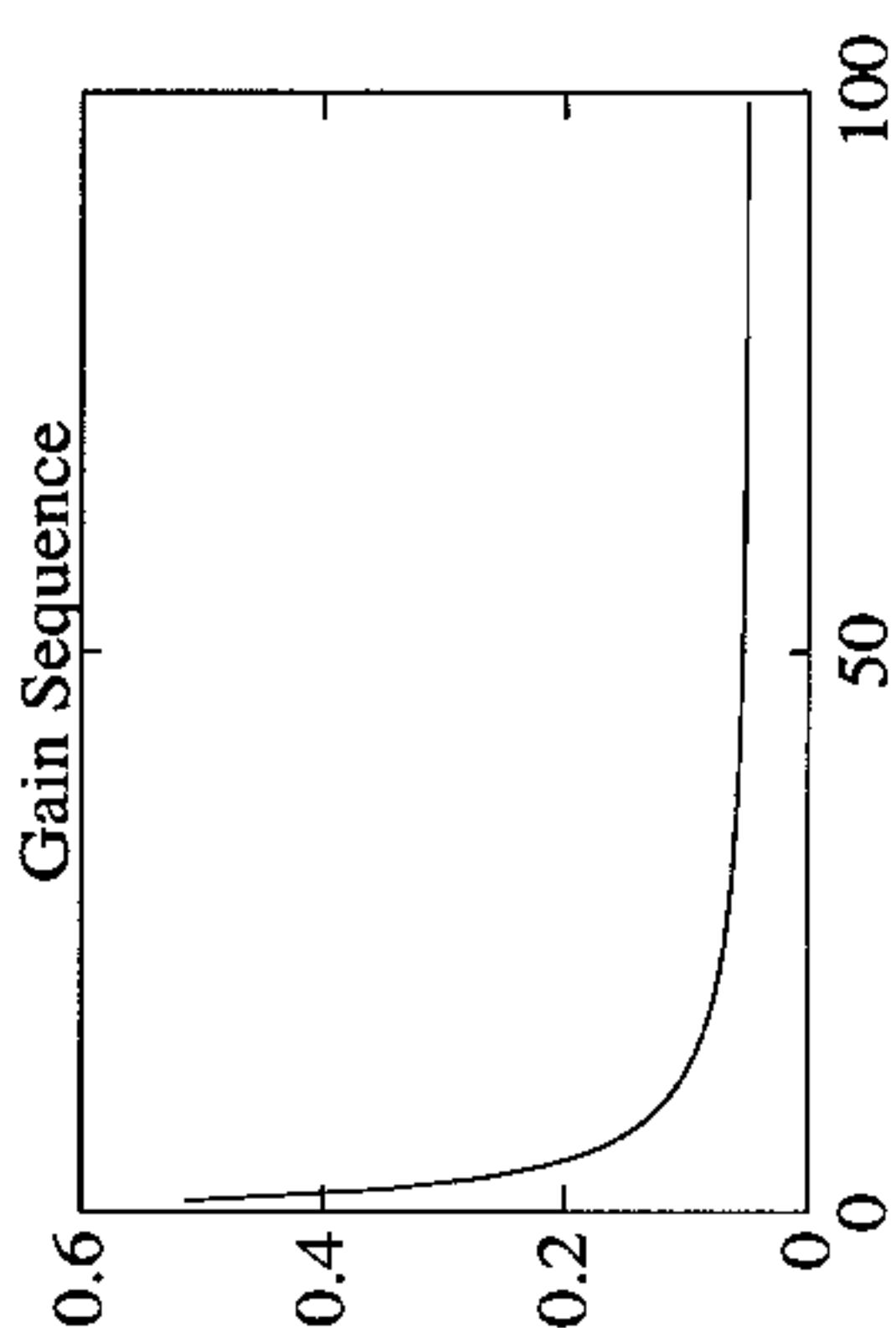
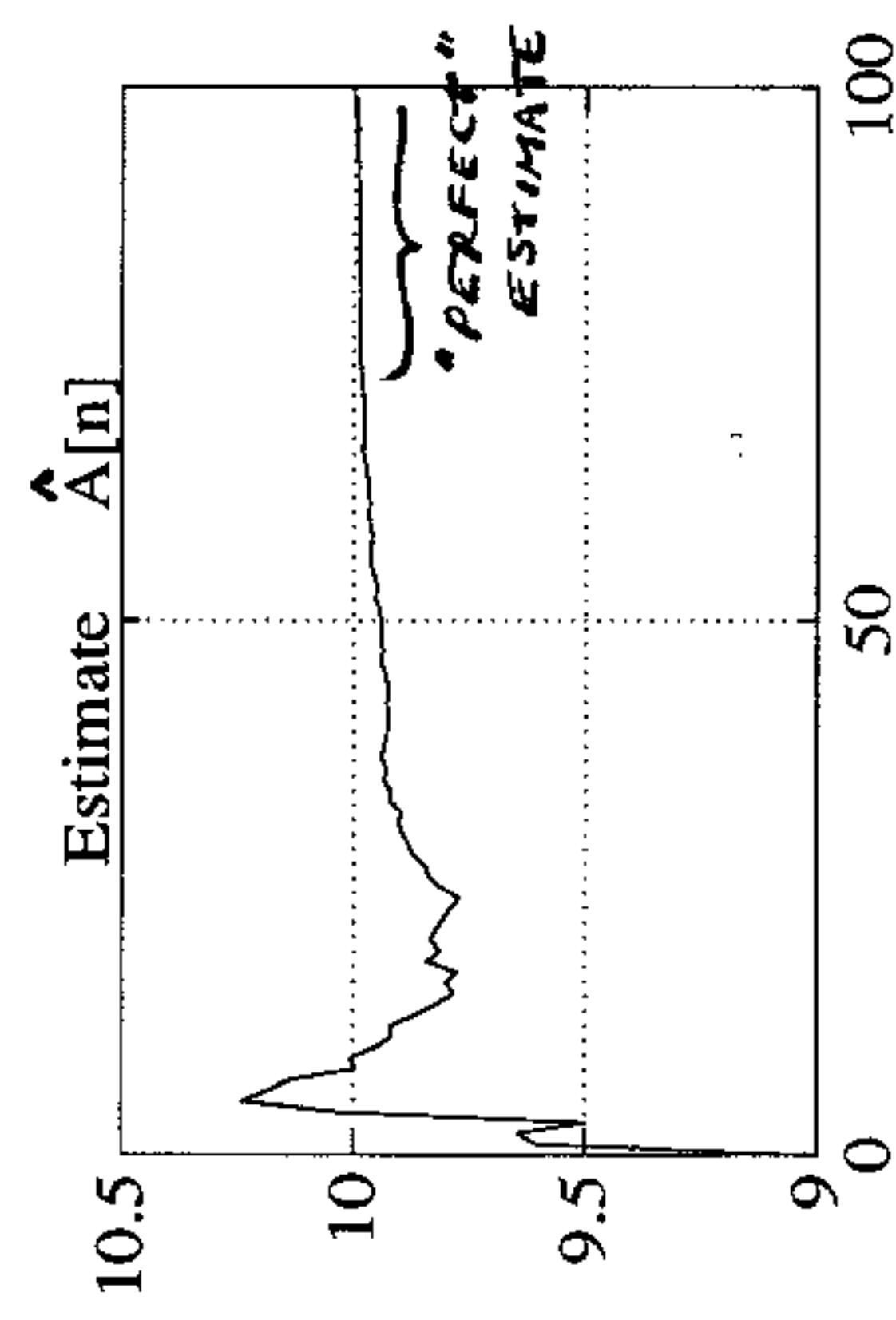
23)  $\hat{\Omega}_r(n) = \left( \sum_{k=-p}^{-1} \frac{1}{\sigma_k^2} h(n) h(k)^T + \sum_{k=0}^n \frac{1}{\sigma_k^2} h(k) h(k)^T \right)^{-1}$   
 $\cdot \left( \sum_{k=-p}^{-1} \frac{1}{\sigma_k^2} x(k) h^T(k) + \sum_{k=0}^n \frac{1}{\sigma_k^2} x(k) h^T(k) \right)$

But  $\hat{\Omega}(-1) = \left( \sum_{k=-p}^{-1} \frac{1}{\sigma_k^2} h(k) h^T(k) \right)^{-1}$   
 $\cdot \left( \sum_{k=-p}^{-1} \frac{1}{\sigma_k^2} x(k) h^T(k) \right)$

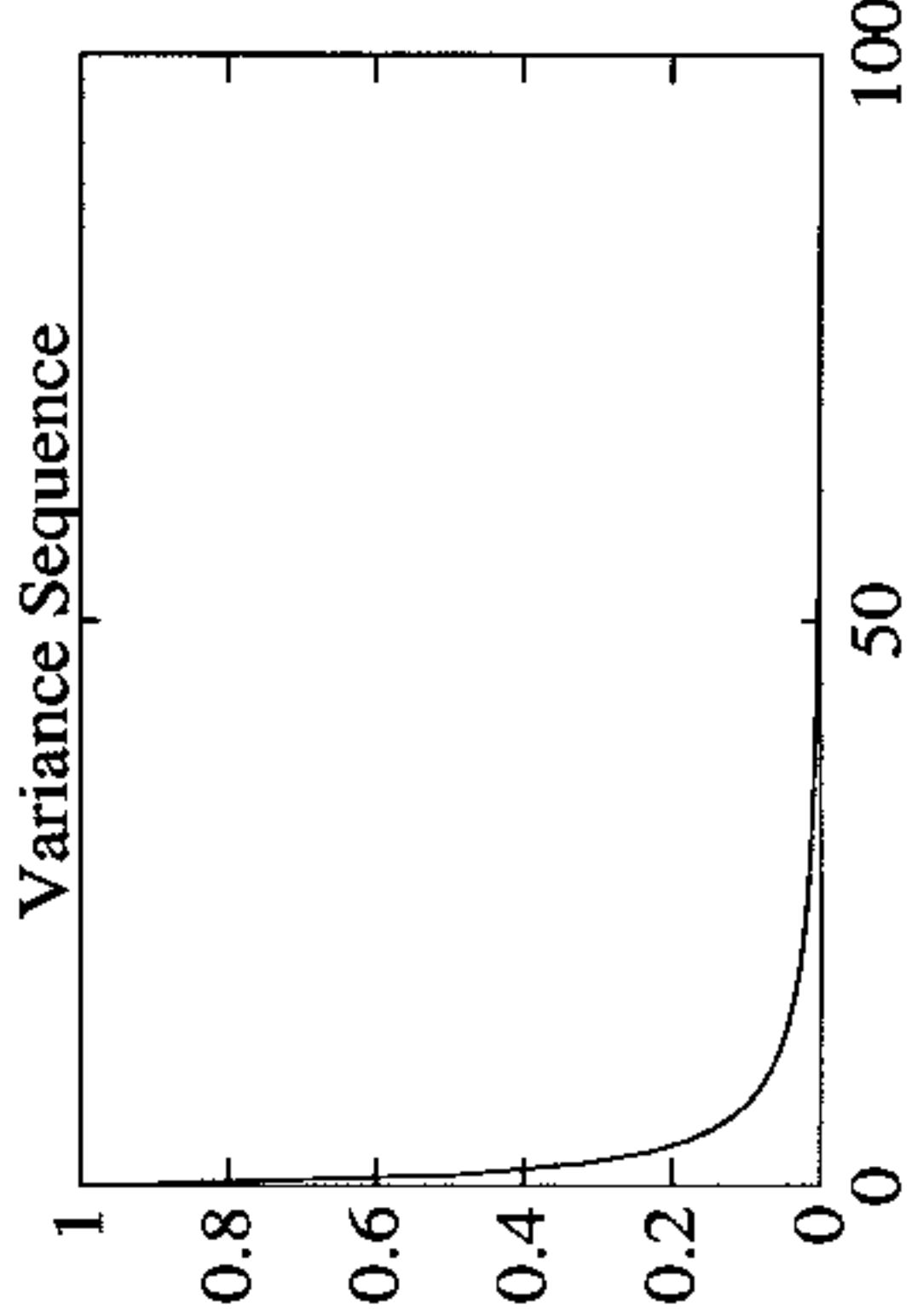
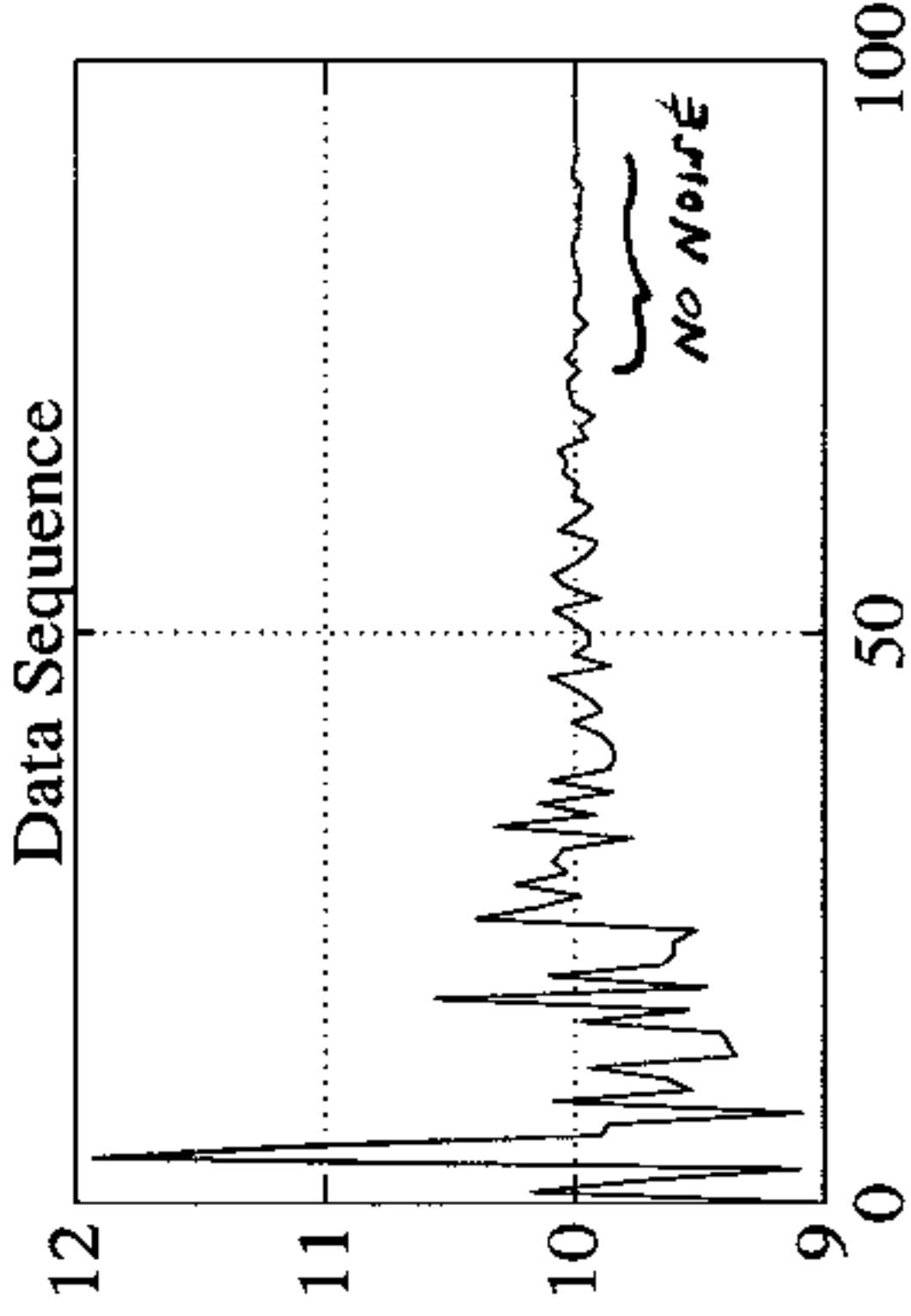
$\Sigma(-1) = \left( \sum_{k=-p}^{-1} \frac{1}{\sigma_k^2} h(k) h^T(k) \right)^{-1}$



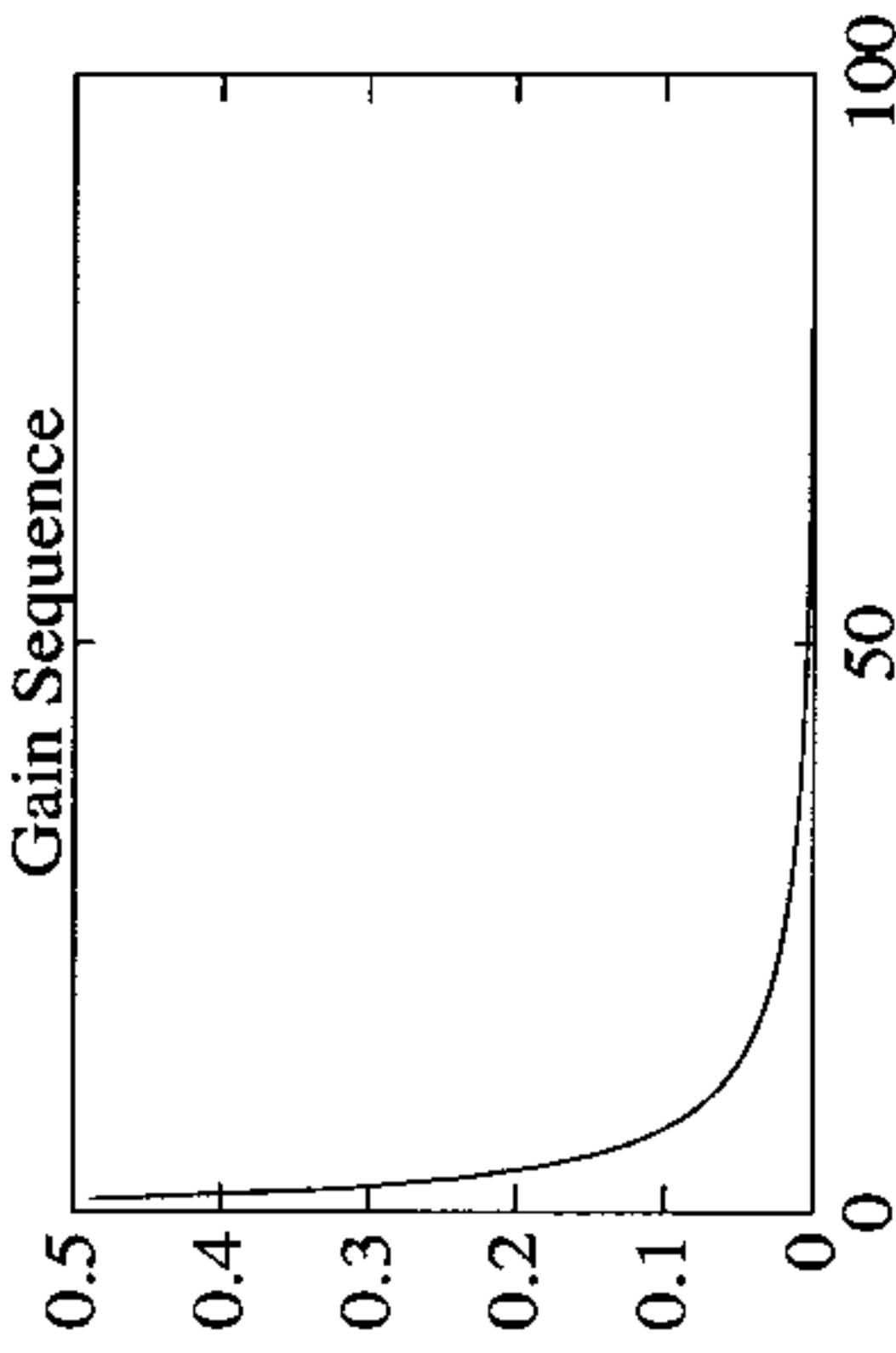
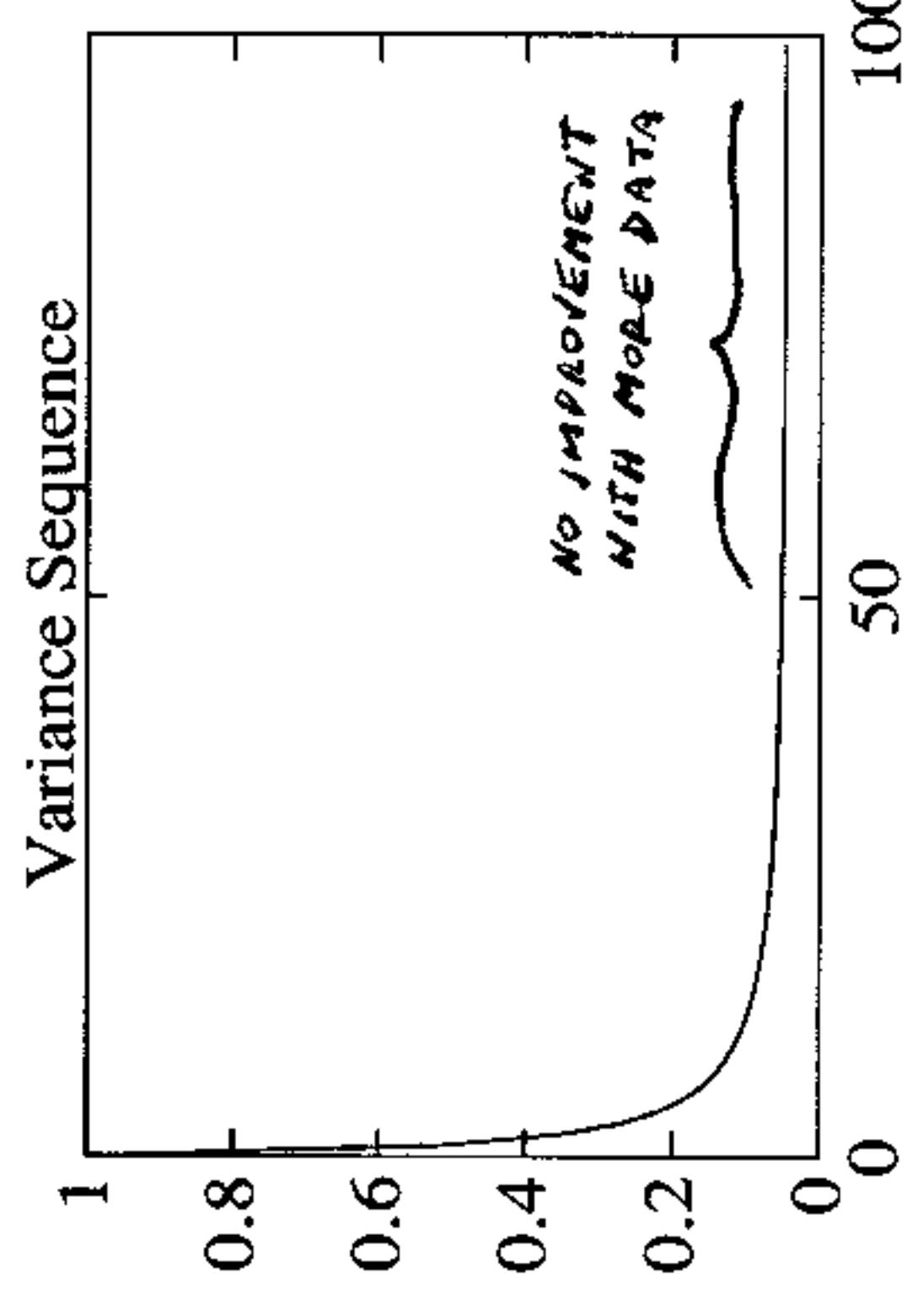
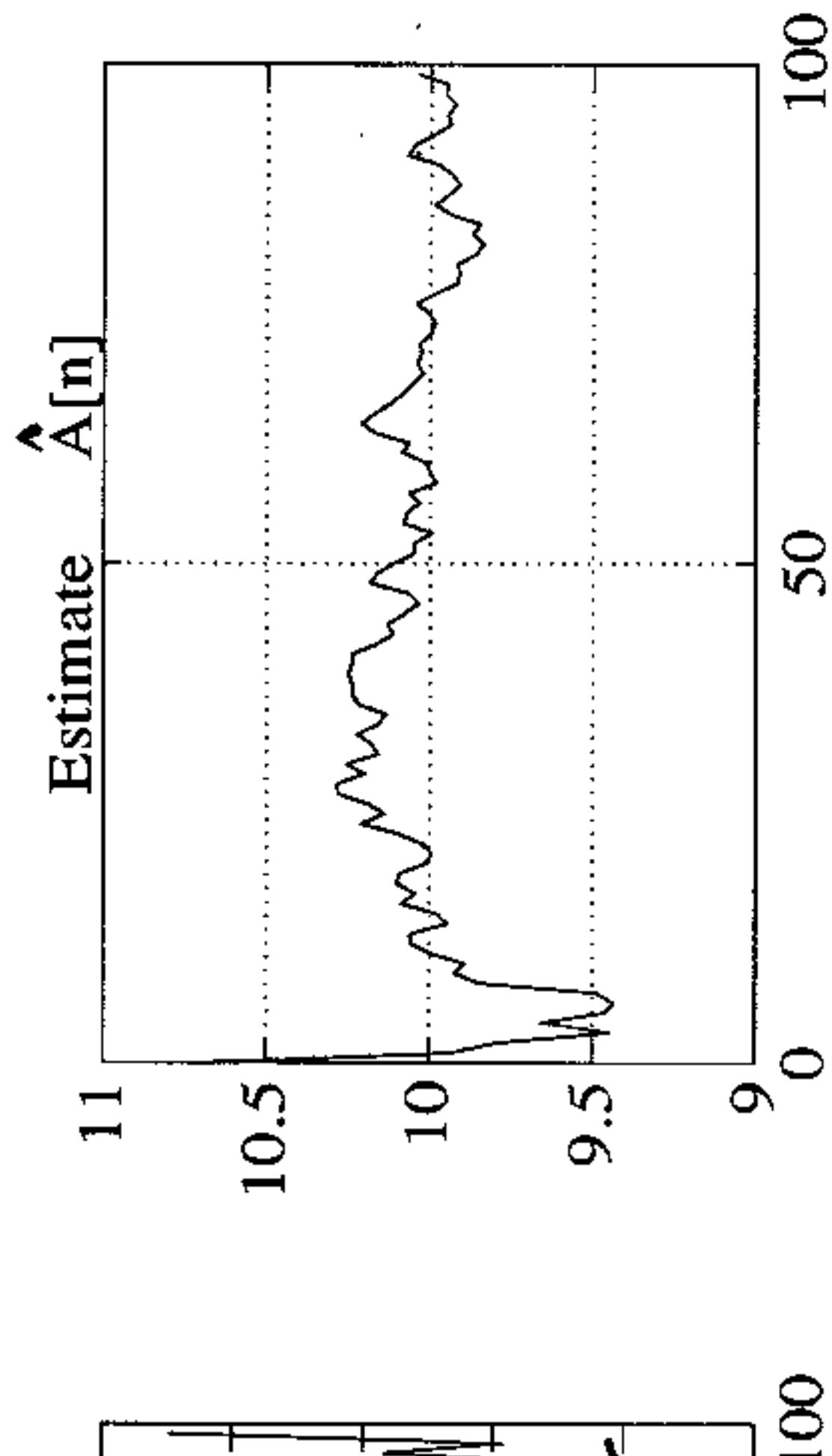
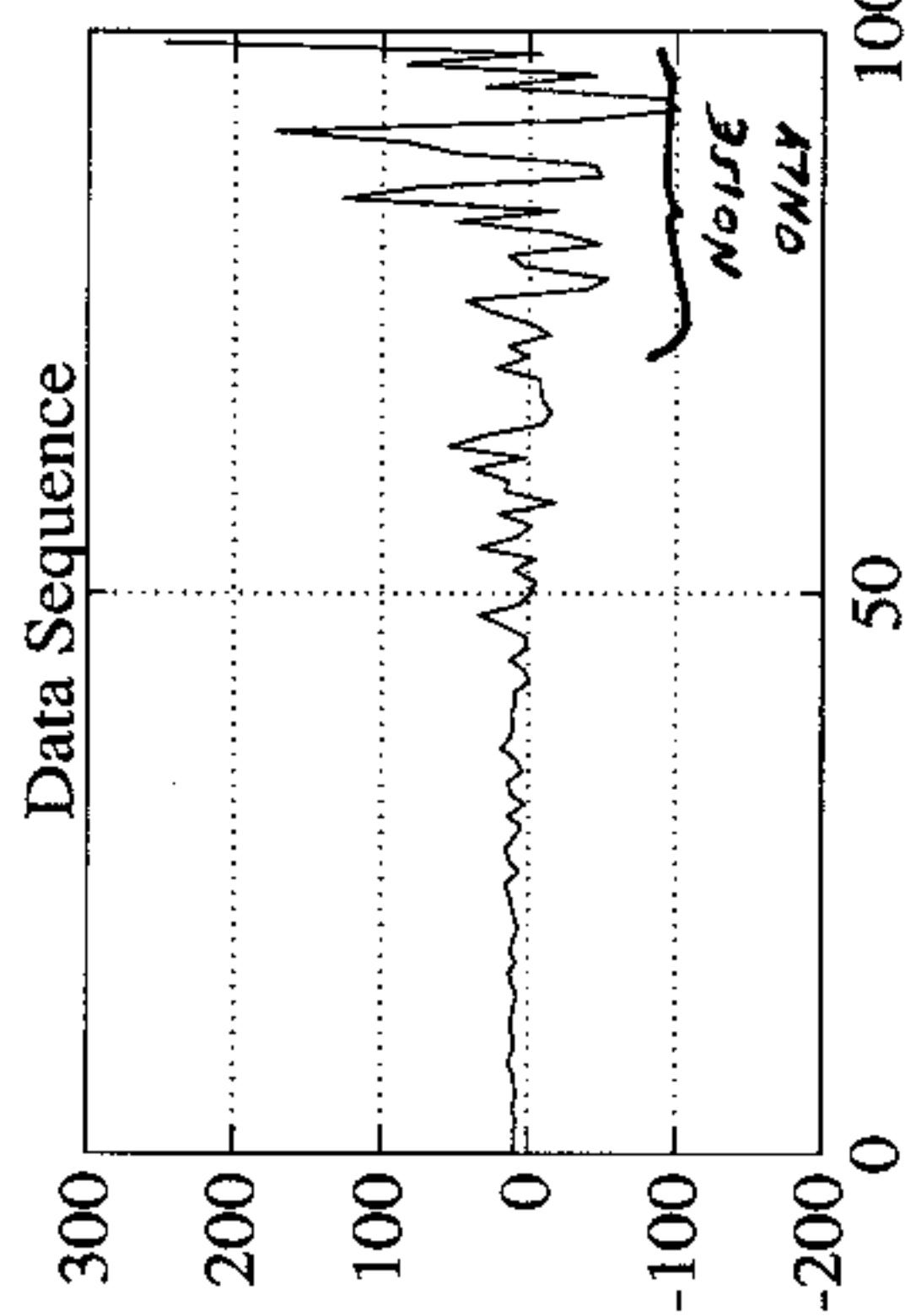
$$r = 1.00$$



$$r = 0.95$$



Prob. 8.22



$$Y = 1.05$$

$$\Rightarrow \hat{\theta}_r(n) = \left( \Sigma^{-1}(-1) + \sum_{h=0}^n \frac{1}{\sigma_h^2} \underline{h}(h) \underline{h}^T(h) \right)^{-1}$$

$$\cdot \left( \Sigma^{-1}(-1) \hat{\theta}_{L-1} + \sum_{h=0}^n \frac{1}{\sigma_h^2} x(h) \underline{h}^T(h) \right)$$

Now for  $n \geq p$ ,  $\sum_{h=0}^n \frac{1}{\sigma_h^2} \underline{h}(h) \underline{h}^T(h)$  will

be invertible so we can let  $\alpha \rightarrow \infty$ . Then,  
 $\Sigma^{-1}(-1) \rightarrow 0$  to yield

$$\hat{\theta}_r(n) = \left( \sum_{h=0}^n \frac{1}{\sigma_h^2} \underline{h}(h) \underline{h}^T(h) \right)^{-1}$$

$$\cdot \left( \sum_{h=0}^n \frac{1}{\sigma_h^2} x(h) \underline{h}^T(h) \right)$$

$$= \hat{\theta}_B(n)$$

Note that if  $\Sigma(-1) \rightarrow \infty$ , the choice of  
 $\hat{\theta}(-1)$  is immaterial.

24) Projecting  $\underline{x}$  onto subspace spanned by  
 $\underline{h}_1$  and  $\underline{h}_2$  produces  $\hat{\underline{x}} = H(H^T H)^{-1} H^T \underline{x}$   
where  $H = [\underline{h}_1 \ \underline{h}_2]$ . Now the constraint  
subspace is spanned by  $[1 \ 1 \ 0]^T$ . The  
projection onto this subspace is

$$\hat{\underline{x}}_c = \underline{e}_c (\underline{e}_c^T \underline{e}_c)^{-1} \underline{e}_c^T \hat{\underline{x}}$$

$$\underline{e}_c = [1 \ 1 \ 0]^T$$

$$= \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} x^{(0)} \\ x^{(1)} \end{bmatrix}$$

$$= \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{10} \\ x_{11} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}(x_{10} + x_{11}) \\ \frac{1}{2}(x_{10} - x_{11}) \\ 0 \end{bmatrix}$$

25)  $H = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow \text{columns of } H \text{ are orthogonal}$

$$\hat{\theta} = (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x} = \begin{bmatrix} N & 0 \\ 0 & N \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}^T \underline{x}$$

$$= \begin{bmatrix} \frac{1}{N} \sum_{n=0}^{N-1} x_{1n} \\ \frac{1}{N} \sum_{n=0}^{N-1} (-1)^n x_{1n} \end{bmatrix}$$

Now, if  $A = B$  we have  $(1-1)\underline{\theta} = 0$   
or  $A = [1-1]$ ,  $b = 0$  and thus from  
(8.52)

$$\hat{\theta}_c = \hat{\theta} - (\underline{H}^T \underline{H})^{-1} \underline{A}^T [A(\underline{H}^T \underline{H})^{-1} \underline{A}^T]^{-1} A \hat{\theta}$$

$$= \hat{\theta} - \frac{1}{N} \underline{A}^T (\frac{1}{N} \underline{A} \underline{A}^T)^{-1} \underline{A} \hat{\theta}$$

$$= (\underline{I} - \underline{A}^T (\underline{A} \underline{A}^T)^{-1} \underline{A}) \hat{\theta}$$

$$\underline{A}^T (\underline{A} \underline{A}^T)^{-1} \underline{A} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \left( [1-1] \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1} [1-1] \right)$$

$$= \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$\underline{I} - \underline{A}^T (\underline{A} \underline{A}^T)^{-1} \underline{A} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$\hat{\theta}_c = \begin{bmatrix} \frac{1}{2} (\bar{x} + \frac{1}{N} \sum_{n=0}^{N-1} (-1)^n x_{1n}) \\ \frac{1}{2} (\bar{x} + \frac{1}{N} \sum_{n=0}^{N-1} (-1)^n x_{2n}) \end{bmatrix}$$

$$\text{or } \hat{A}_c = \hat{B}_c = \frac{1}{2N} \sum_{\substack{n=0 \\ n \text{ even}}}^{N-1} x(n)$$

Makes sense, since if  $A = B$

$$S(n) = A \quad n \text{ even}$$

$$0 \quad n \text{ odd}$$

26) By assumption

$$h(g^{-1}(\hat{x})) \leq h(g^{-1}(x)) \text{ for all } x$$

$$\Rightarrow h(\theta_0) \leq h(\theta) \text{ for all } \theta$$

$$\text{where } \theta_0 = g^{-1}(\hat{x})$$

But then  $\theta_0$  minimizes  $h(\theta)$ , i.e.,  $\theta_0 = \hat{\theta}$

$$\Rightarrow \hat{\theta} = g^{-1}(\hat{x})$$

27) From (8.61)

$$\theta_{k+1} = \theta_k + \left( \underline{H}^T(\theta_k) \underline{H}(\theta_k) - \sum_{n=0}^{N-1} G_n(\theta_k) (x[n] - e^{\theta_k}) \right)^{-1} \cdot \underline{H}^T(\theta_k) (x - e^{\theta_k})$$

$$[\underline{H}(\theta)]_i = \frac{\partial S(i)}{\partial \theta} = e^\theta \quad i = 0, 1, \dots, N-1$$

$$[G_n(\theta)]_{ii} = \frac{\partial^2 S[i]}{\partial \theta^2} = e^\theta$$

$$\Rightarrow \underline{H}(\theta) = e^\theta I \quad G_n(\theta) = e^\theta$$

$$\theta_{k+1} = \theta_k + \left( N e^{2\theta_k} - \sum_{n=0}^{N-1} e^{\theta_k} (x[n] - e^{\theta_k}) \right)^{-1}$$

$$= e^{\theta_k} \underline{x}^T (\underline{x} - e^{\theta_k} \underline{1})$$

$$= \theta_k + \frac{e^{\theta_k} (N \bar{x} - N e^{\theta_k})}{N e^{2\theta_k} - N \bar{x} e^{\theta_k} + N e^{2\theta_k}}$$

$$= \theta_k + \frac{\bar{x} - e^{\theta_k}}{2 e^{\theta_k} - \bar{x}}$$

To find analytically let  $\alpha = e^\theta \Rightarrow \hat{\alpha} = \bar{x}$   
and from Prob 8.26  $\hat{\theta} = \ln \bar{x}$ .

$$28) \text{ Since } H(z) = \frac{B(z)}{A(z)} = B(z) \frac{1}{A(z)}$$

$$\begin{aligned} h[n] &= b[n] * g[n] \\ &= \sum_{k=0}^n b[k] g[n-k] \end{aligned}$$

since  $g[n]$  is causal. In matrix form  
for  $n = 0, 1, \dots, N-1$  we have  $\underline{s} = \underline{G} \underline{b}$ .

$$\begin{aligned} \mathcal{T}(\underline{a}, \underline{b}) &= (\underline{x} - \underline{s})^T (\underline{x} - \underline{s}) \\ &= (\underline{x} - \underline{G} \underline{b})^T (\underline{x} - \underline{G} \underline{b}) \\ \Rightarrow \hat{\underline{b}} &= (\underline{G}^T \underline{G})^{-1} \underline{G}^T \underline{x} \end{aligned}$$

and  $J(\underline{a}, \underline{b}) = \underline{x}^T (\underline{\underline{I}} - \underline{\underline{G}}(\underline{\underline{G}}^T \underline{\underline{G}})^{-1} \underline{\underline{G}}^T) \underline{x}$   
from (8.10)

$$\text{Now } G(z)A(z) = 1 \Rightarrow g[n] * a[n] = \delta[n]$$

$$\begin{aligned} [A^T \underline{\underline{G}}]_{ij} &= \sum_{k=1}^N (A^T)_{ik} (G)_{kj} & i = 1, 2, \dots, N-p \\ &= \sum_{k=1}^N a[p+i-k] g[k-j] & j = 1, 2, \dots, q+1 \\ &= \sum_{k=-\infty}^{\infty} a[p+i-k] g[k-j] & = 1, 2, \dots, p \end{aligned}$$

$$\begin{aligned} \text{where } a[n] &= 0 \quad \text{for } n < 0, n > p \\ g[n] &= 0 \quad \text{for } n < 0 \end{aligned}$$

$$\begin{aligned} &= \sum_{\ell=-\infty}^{\infty} g[\ell] a[p+i-j-\ell] \\ &= g[n] * a(n) |_{n=p+i-j} \\ &= \delta[p+i-j] \end{aligned}$$

Since  $1 \leq i \leq N-p, 1 \leq j \leq p,$   
 $p+i-j > 0 \quad \text{for all } i, j$

$$\Rightarrow A^T \underline{\underline{G}} = 0$$

$$\underline{L} (\underline{\zeta}^T \underline{L})^{-1} \underline{\zeta}^T = \underline{\zeta}$$

$$[\underline{A} \ \underline{G}] \left( \begin{bmatrix} \underline{A}^T \\ \underline{G}^T \end{bmatrix} [\underline{A} \ \underline{G}] \right)^{-1} \begin{bmatrix} \underline{A}^T \\ \underline{G}^T \end{bmatrix} = \underline{\zeta}$$

$$[\underline{A} \ \underline{G}] \begin{bmatrix} \underline{A}^T \underline{A} & \underline{0} \\ \underline{0} & \underline{G}^T \underline{G} \end{bmatrix}^{-1} \begin{bmatrix} \underline{A}^T \\ \underline{G}^T \end{bmatrix} = \underline{\zeta}$$

$$\Rightarrow \underline{A} (\underline{A}^T \underline{A})^{-1} \underline{A}^T + \underline{G} (\underline{G}^T \underline{G})^{-1} \underline{G}^T = \underline{\zeta}$$

29) If  $f_{0_{k+1}} = f_{0_k}$ ,  $\phi_{k+1} = \phi_k$ , we have

$$\sum_{n=-M}^M n \times \underline{x}(n) \sin(2\pi \hat{f}_0 n + \hat{\phi}) = 0$$

$$\sum_{n=-M}^M \underline{x}(n) \sin(2\pi \hat{f}_0 n + \hat{\phi}) = 0$$

At high SNR we have

$$\sum_n n \cos(2\pi f_0 n + \phi) \sin(2\pi \hat{f}_0 n + \hat{\phi}) = 0$$

$$\sum_n \cos(2\pi f_0 n + \phi) \sin(2\pi \hat{f}_0 n + \hat{\phi}) = 0$$

$\therefore$

$$\frac{1}{2} \sum_n n \left[ \sin(2\pi(f_0 + \hat{f}_0)n + \phi + \hat{\phi}) + \sin(2\pi(\hat{f}_0 - f_0)n + (\hat{\phi} - \phi)) \right] = 0$$

$$\frac{1}{2} \sum_n \left[ \sin(2\pi(f_0 + \hat{f}_0)n + \phi + \hat{\phi}) + \sin(2\pi(\hat{f}_0 - f_0)n + (\hat{\phi} - \phi)) \right] = 0$$

Neglecting the high frequency term we have

$$\sum_n n \sin(2\pi(\hat{f}_0 - f_0)n + \hat{\phi} - \phi) = 0$$

$$\sum_n \sin(2\pi(\hat{f}_0 - f_0)n + \hat{\phi} - \phi) = 0$$

for which the solution is  $\hat{f}_0 = f_0, \hat{\phi} = \phi.$

Chapter 9

$$1) E(x) = \int_0^\infty \frac{x^2}{\sigma^2} e^{-\frac{1}{2} \frac{x^2}{\sigma^2}} dx = \sqrt{\pi/2} \sigma$$

$$\Rightarrow \sigma^2 = \frac{2}{\pi} E^2(x)$$

$$\hat{\sigma}^2 = \frac{2}{\pi} \left( \frac{1}{N} \sum_{n=0}^{N-1} x[n] \right)^2 = \frac{2}{\pi} \bar{x}^2$$

$$2) E(x) = 0 \quad \text{since the PDF is even.}$$

Try  $E(x^2)$

$$\begin{aligned} E(x^2) &= \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{x^2}{\sigma^2}} dx \\ &= \frac{2}{\sqrt{2}\sigma} \int_0^\infty x^2 e^{-\frac{1}{2} \frac{x^2}{\sigma^2}} dx = \sigma^2 \end{aligned}$$

$$\Rightarrow \sigma = \sqrt{E(x^2)} \quad \text{and} \quad \hat{\sigma} = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} x^2[n]}$$

$$\begin{aligned} 3) \rho &= \text{cov}(u, v) \quad \text{where } \underline{x} = \begin{bmatrix} u \\ v \end{bmatrix} \\ &= E(uv) \end{aligned}$$

$$\Rightarrow \hat{\rho} = \frac{1}{N} \sum_{n=0}^{N-1} u_n v_n \quad \text{where } \underline{x}[n] = \begin{bmatrix} u_n \\ v_n \end{bmatrix}$$

The cubic equation was found in Prob. 7.11.  
 Clearly, the method of moments estimator is much simpler to find and implement.

4) Since  $\mu = E(x)$ ,  $\sigma^2 = \text{var}(x)$

$$\hat{\mu} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \hat{\mu})^2$$

5) Replace the theoretical ACF by  $\hat{r}_{xx}[n]$  where

$$\hat{r}_{xx}[n] = \frac{1}{N} \sum_{n=0}^{N-1-k_1} x[n] x[n+k_1]$$

$$\Rightarrow \hat{r}_{xx}[n] = - \sum_{k=1}^p a(k) \hat{r}_{xx}[n-k] \quad n > p$$

Choosing  $n = p+1, \dots, q+p$  yields the linear equations

$$\begin{bmatrix} \hat{r}_{xx}[p] & \dots & \hat{r}_{xx}[q-p+1] \\ \vdots & & \vdots \\ \hat{r}_{xx}[q+p-1] & \dots & \hat{r}_{xx}[q] \end{bmatrix} \begin{bmatrix} a(1) \\ \vdots \\ a(p) \end{bmatrix} = - \begin{bmatrix} \hat{r}_{xx}[q+1] \\ \vdots \\ \hat{r}_{xx}[q+p] \end{bmatrix}$$

which can be solved for the  $a[n]$ 's.

6) Let  $\theta = A^2$  so that  $\hat{\theta} = g(\bar{x})$  where  $g(x) = x^2$ .

$T = \bar{x}$  so from (9.15)

$$E(\hat{\theta}) = g(E(T)) = g(A) = A^2$$

From (9.16)

$$\text{var}(\hat{\theta}) = \left( \frac{\partial g}{\partial T} \Big|_{T=A} \right)^2 \text{var}(T)$$

$$= (2A)^2 \sigma^2/N = 4A^2 \sigma^2/N$$

$$= (2A)^2 \theta^2/N = 4A^2 \theta^2/N$$

$$7) \quad E(x(n)) = \cos \phi$$

$$\Rightarrow \phi = \arccos E(x(n))$$

or

$$\hat{\phi} = \arccos \left( \frac{1}{N} \sum_{n=0}^{N-1} x(n) \right)$$

$$\hat{\phi} = h(\underline{w}) \text{ where}$$

$$h(\underline{w}) = \arccos \left[ \frac{1}{N} \sum_{n=0}^{N-1} (\cos \phi + w(n)) \right]$$

From (9.18)

$$E(\hat{\phi}) = h(\underline{0}) = \arccos \left[ \frac{1}{N} \sum_{n=0}^{N-1} \cos \phi \right] = \phi$$

From (9.19)

$$\text{var}(\hat{\phi}) = \left. \frac{\partial h}{\partial w_i} \right|_{\underline{w}=\underline{0}} \sigma^2 \equiv \left. \frac{\partial h}{\partial w_i} \right|_{\underline{w}=\underline{0}}$$

$$\left. \frac{\partial h}{\partial w_i} \right|_{\underline{w}=\underline{0}} = - \frac{1}{\sqrt{1-u^2}} \left. \frac{\partial u}{\partial w_i} \right|_{\underline{w}=\underline{0}}$$

$$\text{where } u = \frac{1}{N} \sum_{n=0}^{N-1} (\cos \phi + w(n))$$

$$\left. \frac{\partial h}{\partial w_i} \right|_{\underline{w}=\underline{0}} = - \frac{1}{\sqrt{1-\cos^2 \phi}} \cdot \frac{1}{N} = - \frac{1}{N \sin \phi}$$

$$\text{var}(\hat{\phi}) = \sigma^2 \sum_{n=0}^{N-1} \left( \left. \frac{\partial h}{\partial w_i} \right|_{\underline{w}=\underline{0}} \right)^2$$

$$= \sigma^2 \sum_{n=0}^{N-1} \frac{1}{N^2 \sin^2 \phi} = \frac{\sigma^2}{N \sin^2 \phi}$$

$$8) \quad \hat{\theta} = g(\underline{x})$$

$$\approx g(\underline{\mu}) + \sum_{k=1}^r \left. \frac{\partial g}{\partial T_k} \right|_{T=\underline{\mu}} (T_k - \mu_k) \\ + \frac{1}{2} (\underline{T} - \underline{\mu})^T \left. \frac{\partial^2 g}{\partial T \partial T^T} \right|_{T=\underline{\mu}} (\underline{T} - \underline{\mu})$$

where  $\frac{\partial^2 g}{\partial T \partial T^T}$  is the Hessian

$$E(\hat{\theta}) = g(\underline{\mu}) + \frac{1}{2} E \left[ (\underline{T} - \underline{\mu})^T G(\underline{\mu}) (\underline{T} - \underline{\mu}) \right] \\ = g(\underline{\mu}) + \frac{1}{2} E \left[ \text{tr} ((\underline{T} - \underline{\mu})^T G(\underline{\mu}) (\underline{T} - \underline{\mu})) \right] \\ = g(\underline{\mu}) + \frac{1}{2} \text{tr} E [G(\underline{\mu}) (\underline{T} - \underline{\mu})(\underline{T} - \underline{\mu})^T] \\ = g(\underline{\mu}) + \frac{1}{2} \text{tr} (G(\underline{\mu}) \Sigma_T)$$

$$q) \quad \hat{\theta} \approx g(\underline{\mu}) + \left. \frac{\partial g}{\partial T} \right|_{T=\underline{\mu}} (\underline{T} - \underline{\mu}) \\ + \frac{1}{2} (\underline{T} - \underline{\mu}) G(\underline{\mu}) (\underline{T} - \underline{\mu})$$

$$\text{var}(\hat{\theta}) = E \left[ (\hat{\theta} - E(\hat{\theta}))^2 \right] \\ = E \left[ \left( \left. \frac{\partial g}{\partial T} \right|_{T=\underline{\mu}} (\underline{T} - \underline{\mu}) + \frac{1}{2} (\underline{T} - \underline{\mu})^T G(\underline{\mu}) (\underline{T} - \underline{\mu}) \right)^2 \right. \\ \left. - \frac{1}{2} \text{tr} (G(\underline{\mu}) \Sigma_T)^2 \right]$$

using the results from Prob 9.8

But  $\underline{T} - \underline{\mu} \sim N(0, \Sigma_T)$  so that all odd-order moments are zero. Let  $x = \underline{T} - \underline{\mu}$ ,  $b = \left. \frac{\partial g}{\partial T} \right|_{T=\underline{\mu}}$  and  $A = G(\underline{\mu})$ .

$$\begin{aligned}\text{var}(\hat{\theta}) &= E\left((\underline{b}^T \underline{x} + \frac{1}{2} \underline{x}^T \underline{A} \underline{x} - \frac{1}{2} \text{tr}(\underline{A} \underline{C}_T))^2\right) \\ &= \underline{b}^T \underline{C}_T \underline{b} + \frac{1}{4} E(\underline{x}^T \underline{A} \underline{x})^2 - \frac{1}{4} E(\underline{x}^T \underline{A} \underline{x}) \text{tr}(\underline{A} \underline{C}_T) \\ &\quad - \frac{1}{4} E(\underline{x}^T \underline{A} \underline{x}) \text{tr}(\underline{A} \underline{C}_T) + \frac{1}{4} \text{tr}^2(\underline{A} \underline{C}_T)\end{aligned}$$

But  $E(\underline{x}^T \underline{A} \underline{x}) = E[\text{tr}(\underline{A} \underline{x} \underline{x}^T)]$   
 $= \text{tr}(\underline{A} E(\underline{x} \underline{x}^T))$   
 $= \text{tr}(\underline{A} \underline{C}_T)$

$$\begin{aligned}E[(\underline{x}^T \underline{A} \underline{x})^2] &= \text{var}(\underline{x}^T \underline{A} \underline{x}) + E^2(\underline{x}^T \underline{A} \underline{x}) \\ &= 2 \text{tr}((\underline{A} \underline{C}_T)^2) + \text{tr}^2(\underline{A} \underline{C}_T)\end{aligned}$$

$$\begin{aligned}\text{var}(\hat{\theta}) &= \underline{b}^T \underline{C}_T \underline{b} + \frac{1}{2} \text{tr}((\underline{A} \underline{C}_T)^2) + \frac{1}{4} \text{tr}^2(\underline{A} \underline{C}_T) \\ &\quad - \frac{1}{2} \text{tr}^2(\underline{A} \underline{C}_T) + \frac{1}{4} \text{tr}^2(\underline{A} \underline{C}_T) \\ &= \underline{b}^T \underline{C}_T \underline{b} + \frac{1}{2} \text{tr}((\underline{A} \underline{C}_T)^2)\end{aligned}$$

(o)  $g(T_i) = 1/T_i$ , where  $T_i = \frac{1}{N} \sum_{n=0}^{N-1} x(n) = \bar{x}$   
 $\bar{x}$  will be approximately Gaussian due  
to the central limit theorem

$$E(\hat{\lambda}) = g(u) + \frac{1}{2} \text{tr}[G(u) C_T]$$

$$\text{But } u = E(T_i) = 1/\lambda$$

$$C_T = \text{var}(T_i) = \text{var}(x(n))/N = \frac{1}{N \lambda^2}$$

$$G(u) = \frac{\partial^2 g}{\partial T^2} \Big|_{T=u}$$

$$= \frac{2}{T^3} \Big|_{T=u} = \frac{2}{u^3} = 2 \lambda^3$$

$$E(\hat{\lambda}) = \lambda + \frac{1}{2} (2\lambda^3) \frac{1}{N\lambda^2} = \lambda + \lambda/N \\ = \lambda(1 + 1/N)$$

$$\text{var}(\hat{\lambda}) = \left( \frac{\partial g}{\partial \tau} \Big|_{T=\mu} \right)^2 \text{var}(\tau) \\ + \frac{1}{2} \text{tr}((G(\mu)(\tau))^2) \\ = \lambda^2/N + \frac{1}{2} (-2\lambda^2 1/N\lambda^2)^2 \\ = \lambda^2/N + 2\lambda^2/N^2 = \frac{\lambda^2}{N}(1 + 2/N)$$

The estimator displays a bias of  $\lambda/N$  and an additional variance of  $2\lambda^2/N^2$ .

Asymptotic MLE theory is valid to first-order.

If  $2/N \ll 1$ , the MLE asymptotics will hold.

$$(1) \quad \frac{1}{N-1} \sum_{n=0}^{N-2} A \cos(2\pi f_0 n + \phi) A \cos(2\pi f_0(n+1) + \phi) \\ = \frac{A^2}{2(N-1)} \sum_{n=0}^{N-2} [\cos(4\pi f_0 n + 2\pi f_0 + 2\phi) \\ + \cos(2\pi f_0)]$$

As  $N \rightarrow \infty$ , the double-frequency term  $\rightarrow 0$  for  $f_0$  not near 0 or  $1/2$ . Hence, the overall expression  $\rightarrow \frac{A^2}{2} \cos 2\pi f_0$ .

Method of moments is valid since ensemble mean equals temporal mean as  $N \rightarrow \infty$  (ergodic).

12) If  $A \neq \sqrt{2}$ , (9.20) can produce meaningless results. From Prob. 9.11, in the absence of noise, the argument of (9.20) will be approximately  $A^2/2 \cos 2\pi f_0 t$ . We need to normalize out the  $A^2/2$  factor.

Now at high SNR

$$\frac{1}{N-1} \sum_{n=0}^{N-2} x(n)x(n+1) \approx \frac{1}{N-1} \sum_{n=0}^{N-2} s(n)s(n+1)$$

$$\text{where } s(n) = A \cos(2\pi f_0 n + \phi)$$

and as  $N \rightarrow \infty$ , this becomes  $A^2/2 \cos 2\pi f_0 t$  from Prob. 9.11. Similarly,

$$\frac{1}{N} \sum_{n=0}^{N-1} x^2(n) \rightarrow A^2/2$$

so that  $\hat{f}_0 \rightarrow f_0$ . Also, from (9.18) we have  $E(\hat{f}_0) = f_0$  for large  $N$ . Note that

$$\hat{f}_0 = \frac{1}{2\pi} \arccos \left[ \frac{\hat{r}_{xx}(1)}{\hat{r}_{xx}(0)} \right] \text{ so that}$$

it is a method of moments estimator.

At lower SNR, as  $N \rightarrow \infty$  we have

$$\hat{r}_{xx}(1) \rightarrow r_{xx}(1) = A^2/2 \cos 2\pi f_0 t$$

$$\hat{r}_{xx}(0) \rightarrow r_{xx}(0) = \frac{A^2}{2} + \sigma^2$$

(assuming  $\phi$  is  $U[0, 2\pi]$ ). Thus, as  $N \rightarrow \infty$

$$\hat{f}_0 \rightarrow \frac{1}{2\pi} \arccos \left[ \frac{A^2 \cos 2\pi f_0 t}{A^2 + 2\sigma^2} \right] \neq f_0$$

$\hat{f}_0$  will be severely biased.

$$(13) \quad \text{var}(\hat{f}_0) = \frac{\sigma^2}{(2\pi)^2 (N-1)^2 \sin^2 2\pi f_0}$$

$$\cdot \left[ s^2(1) + 4 \cos^2 2\pi f_0 \sum_{n=1}^{N-2} s^2(n) + s^2(N-2) \right]$$

$$f_0 = 0.25 \Rightarrow \cos 2\pi f_0 = 0$$

$$s^2(1) = \sqrt{s^2} \cos 2\pi f_0 = 0$$

$$s^2(N-2) = T_2 \cos 2\pi \frac{1}{4}(N-2)$$

$$= T_2 \cos \frac{\pi}{2}(N-2) = 0$$

for  $N$  odd

First-order Taylor expansion is invalid since first-order derivatives are zero. We need a second-order expansion as in Prob 9.9. To see this consider

$$f(x) \approx f(x_0) + \left. \frac{df}{dx} \right|_{x=x_0} (x-x_0)$$

$$+ \frac{1}{2} \left. \frac{d^2 f}{dx^2} \right|_{x=x_0} (x-x_0)^2$$

Even if  $(x-x_0)^2 \ll (x-x_0)$ , the second-order term is not negligible if  $\left. \frac{df}{dx} \right|_{x=x_0} = 0$ .

## Chapter 10

$$\begin{aligned}
 1) \quad \hat{\theta} &= E(\theta | \underline{x}) = \int \theta p(\theta | \underline{x}) d\theta \\
 &= \int \theta \frac{p(\underline{x} | \theta) p(\theta)}{p(\underline{x})} d\theta \\
 &= \frac{\int \theta p(\underline{x} | \theta) p(\theta) d\theta}{\int p(\underline{x} | \theta) p(\theta) d\theta} = \frac{\int \theta p(\underline{x} | \theta) \delta(\theta - \theta_0) d\theta}{\int p(\underline{x} | \theta) \delta(\theta - \theta_0) d\theta} \\
 &= \theta_0 \frac{p(\underline{x} | \theta_0)}{p(\underline{x} | \theta_0)} = \theta_0
 \end{aligned}$$

The MME estimator is just the true value of  $\theta$  since our prior knowledge is perfect. Of course, this is not a valid estimator.

$$\begin{aligned}
 2) \quad p_{\underline{x}}(x^{(0)}, x^{(1)} | A) &= p_{\underline{w}}(x^{(0)} - A, x^{(1)} - A | A) \\
 \text{But } \underline{w} \text{ is independent of } A & \\
 \Rightarrow &= p_{\underline{w}}(x^{(0)} - A, x^{(1)} - A) \\
 \text{and } w^{(0)} \text{ is independent of } w^{(1)} & \\
 \Rightarrow &= p_w(x^{(0)} - A) p_w(x^{(1)} - A) \\
 &= p_{x^{(0)}}(x^{(0)} | A) p_{x^{(1)}}(x^{(1)} | A) \\
 \underline{x} = \begin{bmatrix} A + w^{(0)} \\ A + w^{(1)} \end{bmatrix} &= A \underline{1} + \underline{w} \sim N(0, \underline{I}^T \underline{I} + \underline{I}) \\
 \text{since } A, w^{(0)}, w^{(1)} \text{ are IID and } \sim N(0, I) &
 \end{aligned}$$

Now consider the exponent of  $p(\underline{x})$  or  $\underline{x}^T \underline{C}^{-1} \underline{x}$

$$\underline{C} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \Rightarrow \underline{C}^{-1} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\underline{x}^T \underline{C}^{-1} \underline{x} = \frac{1}{3} \underline{x}^T \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \underline{x} = \frac{1}{3} (2x^2_{(0)} + 2x^2_{(1)} - 2x_{(0)}x_{(1)})$$

which does not factor  $\Rightarrow x_{(0)}, x_{(1)}$  are not independent

$$3) p(\underline{x} | \theta) = e^{-\sum_n (x_{(n)} - \theta)} \quad \text{all } x_{(n)}'s > \theta$$

$$= e^{-n(\bar{x} - \theta)} \quad \bar{x} = \min x_{(n)} > \theta$$

$$p(\theta | \underline{x}) = \frac{p(\underline{x} | \theta) p(\theta)}{\int p(\underline{x} | \theta) p(\theta) d\theta}$$

$$= \frac{e^{-n(\bar{x} - \theta)} e^{-\theta}}{\int_0^{\bar{x}} e^{-n(\bar{x} - \theta)} e^{-\theta} d\theta}$$

$$= \frac{e^{\theta(n-1)}}{\frac{e^{\theta(n-1)}}{N-1} \Big|_0^{\bar{x}}} = \frac{(N-1)e^{\theta(n-1)}}{e^{(N-1)\bar{x}} - 1} \quad 0 < \theta < \bar{x}$$

otherwise

$$\hat{\theta} = E(\theta | \underline{x}) = \frac{\int_0^{\bar{x}} \theta(n-1) e^{\theta(n-1)} d\theta}{e^{(N-1)\bar{x}} - 1}$$

$$= \frac{N-1}{e^{(N-1)\bar{x}} - 1} \left[ \frac{\theta(n-1)-1}{(N-1)^2} e^{\theta(n-1)} \Big|_0^{\bar{x}} \right]$$

$$\begin{aligned}
 &= \frac{1}{(N-1)(e^{(N-1)J} - 1)} \left[ (3(N-1) - 1) e^{(N-1)J} + 1 \right] \\
 &= \frac{1}{(N-1)(e^{(N-1)J} - 1)} \left[ 1 - e^{(N-1)J} + (N-1)J e^{(N-1)J} \right] \\
 &= \frac{3 e^{(N-1)J}}{e^{(N-1)J} - 1} - \frac{1}{N-1}
 \end{aligned}$$

$$\hat{\theta} = \frac{\min x(n)}{1 - e^{-(N-1)\min x(n)}} - \frac{1}{N-1}$$

(4)  $p(x \leq 1\theta) = \frac{1}{\theta^n}$  all  $x(n) \leq \theta$  or  $\max x(n) \leq \theta$

$$p(\theta) = 1/\beta \quad 0 \leq \theta \leq \beta$$

$p(\theta|x)$  will be nonzero for  $\max x(n) \leq \theta \leq \beta$   
or  $3 \leq \theta \leq \beta$

$$p(\theta|x) = \frac{p(x \leq 1\theta) p(\theta)}{\int p(x \leq 1\theta) p(\theta) d\theta}$$

$$\begin{aligned}
 &= \frac{\frac{1}{\theta^n} \frac{1}{\beta}}{\int_3^\beta \frac{1}{\theta^n} \frac{1}{\beta} d\theta} = \frac{\frac{1}{\theta^n}}{-\frac{\theta^{n-1}}{n-1} \Big|_3^\beta} \\
 &= \frac{-(n-1) 1/\theta^n}{\beta^{n-1} - 3^{n-1}}
 \end{aligned}$$

$$= \frac{1/\theta^n}{(N-1)(3^{-(N-n)} - \beta^{-(N-n)})}$$

$$\begin{aligned}
 \hat{\theta} &= E(\theta | \underline{x}) = \int \theta p(\theta | \underline{x}) d\theta \\
 &= C \int_3^{\beta} \theta \frac{d\theta}{\theta^n} \quad C = \frac{1}{\frac{1}{N-1} (\beta^{-(N-1)} - \underline{x}^{-(N-1)})} \\
 &= C \left. \frac{\theta^{-(N-2)}}{-(N-2)} \right|_3^{\beta} = -\frac{C}{N-2} (\beta^{-(N-2)} - \underline{x}^{-(N-2)}) \\
 &= \frac{(\max x_{(n)})^{-(N-2)} - \beta^{-(N-2)}}{(\max x_{(n)})^{-(N-1)} - \beta^{-(N-1)}} \cdot \frac{\frac{N-1}{N-2}}{\frac{N-1}{N-2}}
 \end{aligned}$$

Note that for  $\beta$  large (no prior knowledge) and  $N$  large

$$\hat{\theta} \approx \frac{N-1}{N-2} \max x_{(1)} \approx \max x_{(n)}$$

which agrees with the MLE.

$$\begin{aligned}
 5) \text{ Bias}^2(\hat{\theta}) &= E_{x,\theta} \left\{ [(\theta - E(\theta | \underline{x})) + (E(\theta | \underline{x}) - \hat{\theta})]^2 \right\} \\
 &= E_x \left[ E_{\theta | \underline{x}} \left\{ [(\theta - E(\theta | \underline{x})) + (E(\theta | \underline{x}) - \hat{\theta})]^2 \right\} \right] \\
 &= E_x \left\{ E_{\theta | \underline{x}} [(\theta - E(\theta | \underline{x}))^2] \right. \\
 &\quad \left. + 2 E_{\theta | \underline{x}} [(\theta - E(\theta | \underline{x})) (E(\theta | \underline{x}) - \hat{\theta})] \right. \\
 &\quad \left. + E_{\theta | \underline{x}} [(E(\theta | \underline{x}) - \hat{\theta})^2] \right\}
 \end{aligned}$$

The middle term is zero since conditioned on  $\underline{x}$ ,  $\hat{\theta}$  is a constant so that

$$\begin{aligned}
 & E_{\theta|X} \left[ (\theta - E(\theta|X))(E(\theta|X) - \hat{\theta}) \right] \\
 &= E_{\theta|X} \left[ \theta - E(\theta|X) \right] [E(\theta|X) - \hat{\theta}] \\
 &= \underbrace{[E_{\theta|X}(\theta) - E(\theta|X)]}_{E(\theta|X)} [E(\theta|X) - \hat{\theta}] = 0
 \end{aligned}$$

$$\text{Bmse}(\hat{\theta}) = E_X \left\{ E_{\theta|X} [(\theta - E(\theta|X))^2] + E_{\theta|X} [(E(\theta|X) - \hat{\theta})^2] \right\}$$

Clearly, to minimize  $\text{Bmse}(\hat{\theta})$  we choose  $\hat{\theta} = E(\theta|X)$  so that the last term (which is nonnegative) is zero.

b) Now  $A$  and  $w[n]$  are not independent.

$$p(w[n]|A) = \frac{1}{\sqrt{2\pi\sigma_w^2}} e^{-\frac{1}{2\sigma_w^2} w^2[n]} \quad A \geq 0$$

$$\frac{1}{\sqrt{2\pi\sigma_w^2}} e^{-\frac{1}{2\sigma_w^2} w^2[n]} \quad A < 0$$

$$\Rightarrow p(x[n]|A) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{1}{2\sigma_x^2} (x[n]-A)^2} \quad A \geq 0$$

$$\frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{1}{2\sigma_x^2} (x[n]-A)^2} \quad A < 0$$

or

$$p(x|A) = \frac{1}{(2\pi\sigma_x^2)^{N/2}} e^{-\frac{1}{2\sigma_x^2} \sum_n (x[n]-A)^2} \quad A \geq 0$$

$$\frac{1}{(2\pi\sigma_x^2)^{N/2}} e^{-\frac{1}{2\sigma_x^2} \sum_n (x[n]-A)^2} \quad A < 0$$

Since conditioned on  $A$  the  $w[n]$ 's and hence the  $x[n]$ 's are independent. Clearly,

for  $\sigma_x^2 \neq \sigma_z^2$ ,  $p(x|A) \neq p(z; A)$ .

For  $\sigma_x^2 = \sigma_z^2$  they are identical.

$$\hat{A} = \frac{\int_{-A_0}^{A_0} A \frac{1}{\sqrt{2\pi\sigma_z^2}} e^{-\frac{1}{2\sigma_z^2}(A-z)^2} dA}{\int_{-A_0}^{A_0} \frac{1}{\sqrt{2\pi\sigma_z^2}} e^{-\frac{1}{2\sigma_z^2}(A-z)^2} dA}$$

$$I_1 = \int_{-A_0}^{A_0} A e^{-a(A-z)^2} dA \quad a = \frac{1}{2\sigma_z^2}$$

$$= \int_{-A_0}^{A_0} (A-z) e^{-a(A-z)^2} dA + z \int_{-A_0}^{A_0} e^{-a(A-z)^2} dA$$

$$= \underbrace{\int_{-A_0-z}^{A_0-z} y e^{-ay^2} dy}_{I_2} + z \int_{-A_0}^{A_0} e^{-a(A-z)^2} dA$$

$$I_2 = \left. \frac{e^{-ay^2}}{-2a} \right|_{-A_0-z}^{A_0-z} = -\frac{1}{2a} (e^{-a(A_0-z)^2} - e^{-a(A_0+z)^2})$$

The numerator becomes

$$-\frac{1}{\sqrt{2\pi\sigma_z^2}} \frac{1}{2} \frac{1}{\frac{1}{2\sigma_z^2}} (e^{-\frac{1}{2\sigma_z^2}(A_0-z)^2} - e^{-\frac{1}{2\sigma_z^2}(A_0+z)^2})$$

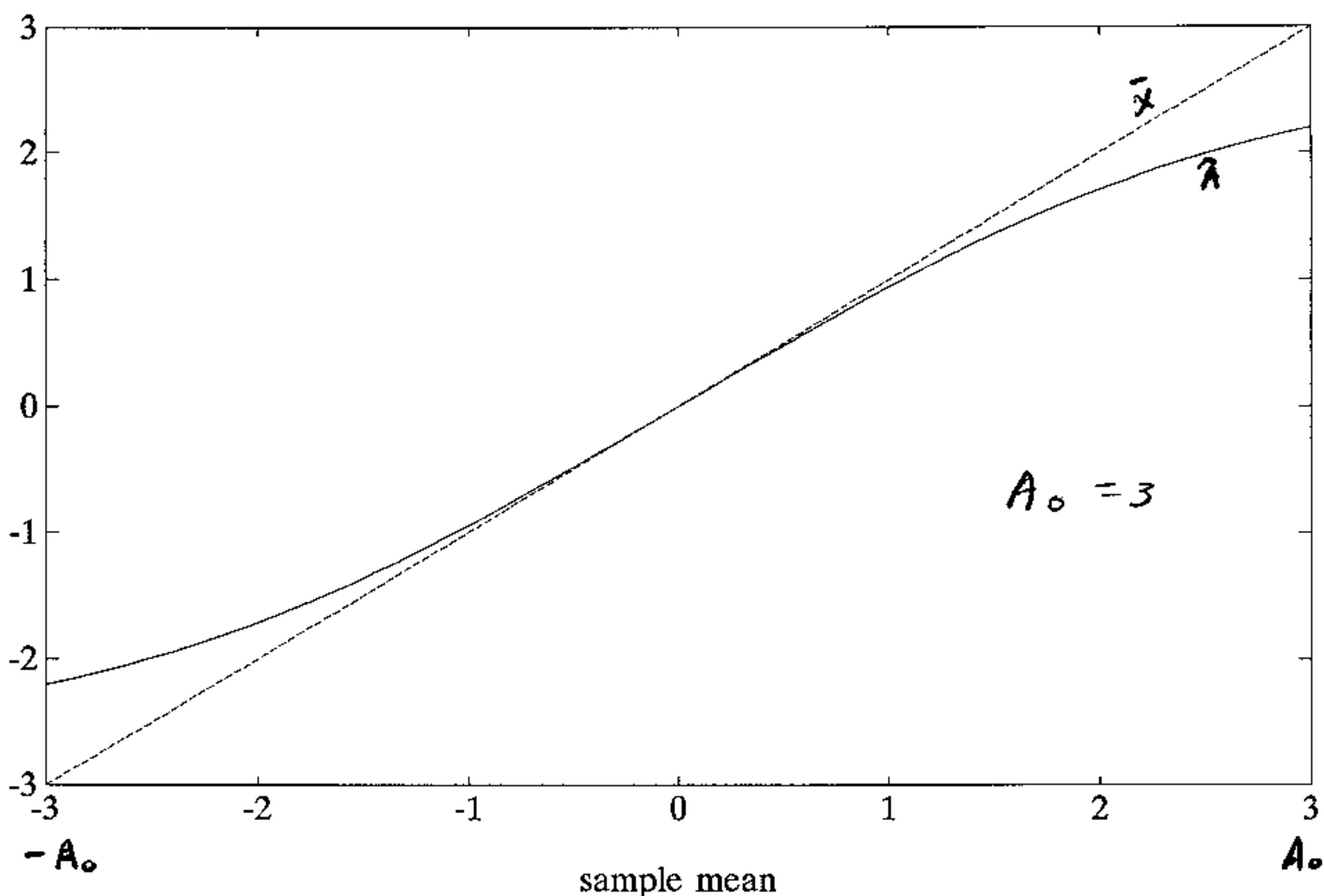
$$+ \frac{z}{\sqrt{2\pi\sigma_z^2}} \int_{-A_0}^{A_0} e^{-\frac{1}{2\sigma_z^2}(A-z)^2} dA$$

or letting  $\Phi(x)$  be the CDF for a  $N(0, 1)$  random variable

$$\hat{A} = \bar{x} + \sqrt{\frac{\sigma^2 N}{2\pi}} \left[ e^{-\frac{1}{2\sigma^2 N}(A_0+\bar{x})^2} - e^{-\frac{1}{2\sigma^2 N}(A_0-\bar{x})^2} \right] / \Phi\left(\frac{A_0-\bar{x}}{\sqrt{\sigma^2 N}}\right) - \Phi\left(\frac{-A_0-\bar{x}}{\sqrt{\sigma^2 N}}\right)$$

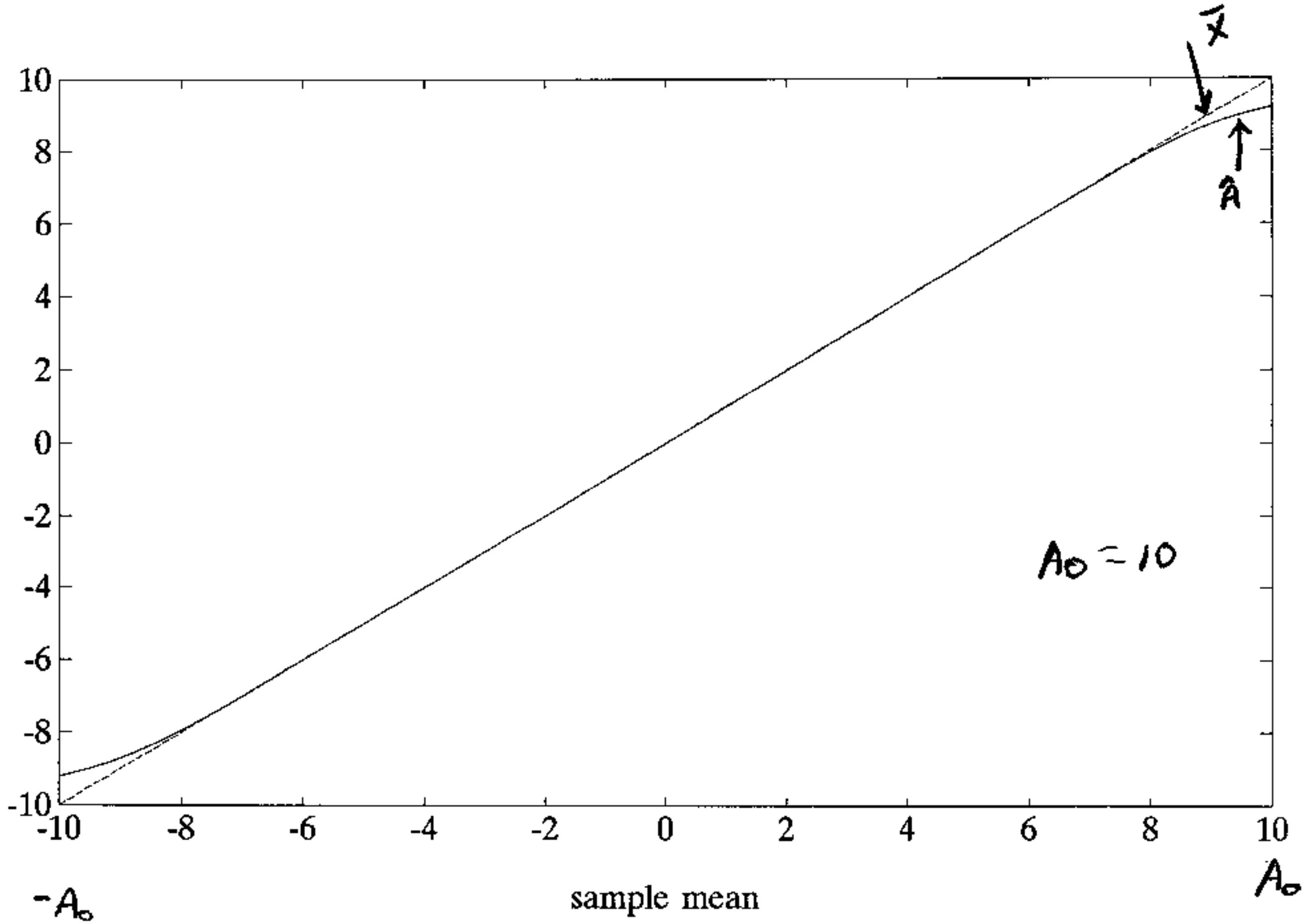
For  $\sqrt{\sigma^2 N} = 1$ ,  $A_0 = 3$  we have

$$\hat{A} = \bar{x} + \frac{e^{-\frac{1}{2}(3+\bar{x})^2} - e^{-\frac{1}{2}(3-\bar{x})^2}}{\sqrt{2\pi} [\Phi(\bar{x}-3) - \Phi(-\bar{x}-3)]}$$



and for  $A_0 = 10$  (see the plot on next page). The curves are nearly identical or  $\hat{A} \approx \bar{x}$  for  $|\bar{x}| \leq A_0$ . Also, as  $\bar{x} \rightarrow \infty$ ,

we will always have  $\hat{A} \rightarrow A_0$ .



$$\begin{aligned}
 8) \quad & \text{Let } J = E[(\theta - \hat{\theta})^2] \\
 & \sigma = E[((\theta - E(\theta)) + (E(\theta) - \hat{\theta}))^2] \\
 & = E[(\theta - E(\theta))^2] + 2E[(\theta - E(\theta))(E(\theta) - \hat{\theta})] \\
 & \quad + E[(E(\theta) - \hat{\theta})^2]
 \end{aligned}$$

Since  $\hat{\theta}$  is a constant, the middle term is zero,

$$\begin{aligned}
 E[(\theta - E(\theta))(E(\theta) - \hat{\theta})] &= E[(\theta - E(\theta))](E(\theta) - \hat{\theta}) \\
 &= (E(\theta) - E(\theta))(E(\theta) - \hat{\theta}) = 0
 \end{aligned}$$

$$\begin{aligned}
 J &= E((\theta - E(\theta))^2) + (E(\theta) - \hat{\theta})^2 \\
 &\geq E((\theta - E(\theta))^2) \\
 \Rightarrow \hat{\theta} &= E(\theta). \text{ The minimum MRE is} \\
 &\text{just } E((\theta - E(\theta))^2) = \text{var}(\theta).
 \end{aligned}$$

From Example 10.1 with no data

$$\text{Bmre}(\hat{A}) = \text{var}(\hat{A}) = \sigma_A^2$$

and with data

$$\text{Bmre}(\hat{A}) = \sigma_A^2 \frac{\sigma^2/N}{\sigma_A^2 + \sigma^2/N} < \sigma_A^2.$$

9) Prior PDF:  $N(\mu_0, \sigma_R^2)$

Data model:  $x[n] = R + w[n] \quad n = 0, 1, \dots, N-1$

where  $w[n] \sim N(0, \frac{1}{\sigma^2})$  and  $w(n)'s$

are independent

This is just Example 10.1.

$$\Rightarrow \text{Bmre}(\hat{R}) = \frac{\sigma_R^2 \sigma^2/N}{\sigma_R^2 + \sigma^2/N}$$

For the error to be 0.1 on the average

$$\Rightarrow \text{Bmre}(\hat{R}) = 0.01$$

$$0.01 = \frac{0.011/N}{0.011 + 1/N} \Rightarrow N = 9.09$$

or  $N = 10$

Without prior knowledge or as  $\sigma_R^2 \rightarrow \infty$

$$\text{Bmre}(\hat{R}) \rightarrow \sigma^2/N$$

$0.01 = \gamma_N \Rightarrow \text{require } N = 100.$

$$10) p(\theta | x) = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta)p(\theta)d\theta}$$

$$= \frac{\theta^N e^{-N\bar{x}} \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\lambda\theta}}{\frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty \theta^{N+\alpha-1} e^{-(N\bar{x}+\lambda)\theta} d\theta}$$

$$\text{But } \int_0^\infty p(\theta)d\theta = 1 \Rightarrow \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\lambda\theta} d\theta = 1$$

$$\int_0^\infty \theta^{\underbrace{N+\alpha-1}_{\alpha'}} e^{-\underbrace{(N\bar{x}+\lambda)\theta}_{\lambda'}} d\theta = \frac{\Gamma(\alpha')}{\lambda'^{\alpha'}}$$

$$= \frac{\Gamma(N+\alpha)}{(N\bar{x}+\lambda)^{N+\alpha}}$$

$$p(\theta | x) = \frac{(N\bar{x}+\lambda)^{N+\alpha}}{\Gamma(N+\alpha)} \theta^{N+\alpha-1} e^{-(N\bar{x}+\lambda)\theta}$$

0

 $\theta > 0$ 

0

This PDF is also a Gamma PDF with parameters  $\alpha' = N+\alpha$ ,  $\lambda' = N\bar{x}+\lambda$ .

Only the PDF parameters change from the prior to the posterior PDF. Note that the trick here is to have  $p(x|\theta)$  and  $p(\theta)$  of the same form and to retain it after multiplication. The denominator is just a

scaling constant.

- 11) The scatter diagram in Figure 10.9a indicates that the PDF of the random vector  $\begin{bmatrix} h \\ w \end{bmatrix}$  is concentrated within an ellipse. This could be a bivariate Gaussian PDF. Hence, assuming  $\begin{bmatrix} h \\ w \end{bmatrix}$  is Gaussian, we estimate  $w$  based on  $h$  using a MMSE estimator or from (10.16)

$$\hat{w} = E(w) + \frac{\text{cov}(h, w)}{\text{var}(h)} (h - E(h))$$

From Fig 10.9b there does not appear to be any correlation between height and weight  
 $\Rightarrow \text{Cov}(h, w) = 0$  or  $\hat{w} = E(w) \approx 150$   
for any height.

$$12) p(x, y) = \frac{1}{2\pi \det \Sigma(\Sigma)} e^{-\frac{1}{2} x^T \Sigma^{-1} x}$$

$$g(y) = p(x_0, y) = \frac{1}{2\pi \det \Sigma(\Sigma)} e^{-\frac{1}{2} h(y)}$$

$$\begin{aligned} \text{where } h(y) &= \begin{bmatrix} x_0 \\ y \end{bmatrix}^T \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{-1} \begin{bmatrix} x_0 \\ y \end{bmatrix} \\ &= \frac{\begin{bmatrix} x_0 \\ y \end{bmatrix}^T \begin{bmatrix} 1-\rho^2 & \rho \\ -\rho & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y \end{bmatrix}}{1-\rho^2} \\ &= (x_0^2 + y^2 - 2\rho x_0 y) / (1-\rho^2) \end{aligned}$$

$g(y)$  is maximized when  $h(y)$  is minimized or

$$\frac{dh}{dy} = \frac{2y - 2\rho x_0}{1-\rho^2} = 0 \Rightarrow y = \rho x_0$$

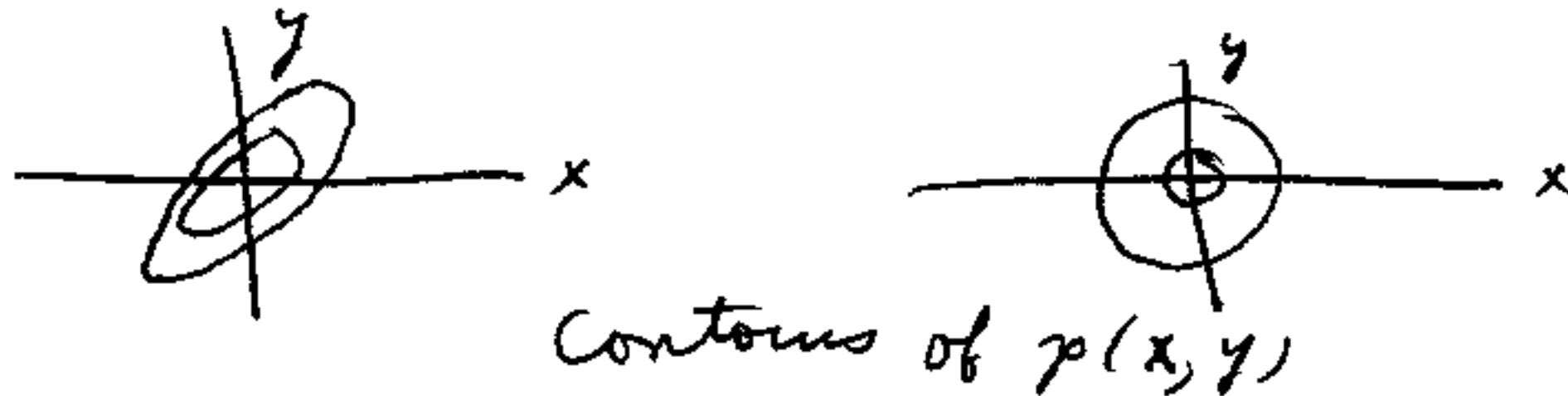
Now from (10.16) with  $E(x) = E(y) = 0$ ,  
 $\text{Cov}(x, y) = \rho$ ,  $\text{var}(x) = 1$   
 $\Rightarrow E(y|x) = \rho x$

The joint PDF  $p(x_0, y)$  has the identical form as  $p(y|x_0)$ , with the only difference being the normalization since

$$p(y|x_0) = \frac{p(x_0, y)}{\int p(x_0, y) dy}$$

Hence, for a given  $x_0$  the maximum of  $p(y|x_0)$  is the same as that for  $p(x_0, y)$ . Also, the posterior PDF is Gaussian so that the mode (maximizing value of  $y$ ) is identical to the mean.

$$\text{If } \rho = 0, E(y|x_0) = 0 = E(y)$$



13) Since

$$(\underline{A} + \underline{B}\underline{\zeta}\underline{D})^{-1} = \underline{A}^{-1} - \underline{A}^{-1}\underline{B}(\underline{D}\underline{A}^{-1}\underline{B} + \underline{\zeta}^{-1})^{-1}\underline{D}\underline{A}^{-1}$$

$$(\underline{\zeta}_\theta^{-1} + \underline{H}^T \underline{C}_w^{-1} \underline{H})^{-1} = \underline{\zeta}_\theta - \underline{\zeta}_\theta \underline{H}^T (\underline{H} \underline{\zeta}_\theta \underline{H}^T + \underline{C}_w)^{-1} \underline{H} \underline{\zeta}_\theta$$

$$\Rightarrow (10.33)$$

Now, using the inversion lemma again

$$(\underline{\zeta}_\theta^{-1} + \underline{H}^T \underline{C}_w^{-1} \underline{A})^{-1} \underline{H}^T \underline{C}_w^{-1} = \underline{\zeta}_\theta \underline{H}^T \underline{C}_w^{-1}$$

$$= \underline{\zeta}_\theta \underline{H}^T (\underline{H} \underline{\zeta}_\theta \underline{H}^T + \underline{C}_w)^{-1} \underline{H} \underline{\zeta}_\theta \underline{H}^T \underline{C}_w^{-1}$$

$$= \underline{\zeta}_\theta \underline{H}^T \underbrace{[\underline{C}_w^{-1} - (\underline{H} \underline{\zeta}_\theta \underline{H}^T + \underline{C}_w)^{-1} \underline{H} \underline{\zeta}_\theta \underline{H}^T \underline{C}_w^{-1}]}_{E = \underline{A}^{-1} - (\underline{Q} + \underline{A})^{-1} \underline{B} \underline{A}^{-1}}$$

$$\underline{\zeta} = \underline{A}^{-1} - \underline{A}^{-1} \underline{B} (\underline{B}^{-1} \underline{A}) (\underline{Q} + \underline{A})^{-1} \underline{B} \underline{A}^{-1}$$

$$= \underline{A}^{-1} - \underline{A}^{-1} \underline{B} [\dots \underline{B}^{-1} (\underline{Q} + \underline{A}) (\underline{B}^{-1} \underline{A})^{-1}]^{-1} \underline{A}^{-1}$$

$$= \underline{A}^{-1} - \underline{A}^{-1} \underline{B} [\dots \underline{B}^{-1} (\underline{Q} + \underline{A}) (\underline{A}^{-1} \underline{B})]^{-1} \underline{A}^{-1}$$

$$= \underline{A}^{-1} - \underline{A}^{-1} \underline{B} [\dots \underline{I} + \underline{A}^{-1} \underline{B}]^{-1} \underline{A}^{-1}$$

$$= (\underline{A} + \underline{B})^{-1} \quad \text{by inversion lemma}$$

$$\Rightarrow E = (\underline{H} \underline{\zeta}_\theta \underline{H}^T + \underline{C}_w)^{-1}$$

14) This is the Bayesian linear model with  
 $\underline{H} = [1, r, \dots, r^{n-1}]^T$

Use (10.32), (10.33) since  $\underline{H}$  is a column vector and thus  $\underline{H}^T \underline{C}_w^{-1} \underline{H}$  is a scalar.

No matrix inversion required as in (10.28)

$$\begin{aligned}
 \hat{A} &= \theta + \left( \frac{1}{\sigma_A^2} + \frac{\underline{H}^T \underline{H}}{\sigma^2} \right)^{-1} \underline{H}^T \frac{1}{\sigma^2} (\underline{x} - \underline{\theta}) \\
 &= \frac{\frac{1}{\sigma^2} \sum_{n=0}^{N-1} x(n) r^n}{\frac{1}{\sigma_A^2} + \frac{\sum_{n=0}^{N-1} r^{2n}}{\sigma^2}} \\
 &= \frac{\frac{\sigma_A^2}{\sigma_A^2 + \sum_{n=0}^{N-1} r^{2n} + \sigma^2} \sum_{n=0}^{N-1} x(n) r^n}{\sigma_A^2 + \sum_{n=0}^{N-1} r^{2n} + \sigma^2}
 \end{aligned}$$

From (10.13)

$$B_{MSE}(\hat{A}) = \int \text{var}(A|\underline{x}) p(\underline{x}) d\underline{x}$$

$$\text{But } \text{var}(\hat{A}) = \frac{1}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n}} \quad \text{from (10.33)}$$

which does not depend on  $\underline{x}$

$$\Rightarrow B_{MSE}(\hat{A}) = \frac{1}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n}}$$

$$(15) \quad \ln p(\theta) = \ln \frac{1}{\sqrt{2\pi\sigma_\theta^2}} - \frac{1}{2\sigma_\theta^2} (\theta - \mu_\theta)^2$$

$$\begin{aligned}
 E[\ln p(\theta)] &= -\frac{1}{2} \ln 2\pi\sigma_\theta^2 - \frac{1}{2\sigma_\theta^2} \underbrace{E[(\theta - \mu_\theta)^2]}_{\sigma_\theta^2} \\
 &= -\frac{1}{2} [1 + \ln 2\pi\sigma_\theta^2]
 \end{aligned}$$

$$H(\theta) = \frac{1}{2} (1 + \ln 2\pi\sigma_\theta^2)$$

The more concentrated the PDF or for  $\sigma_\theta^2$  small,

the smaller will be  $H(\theta)$ . Higher entropy  
 $\Rightarrow$  larger  $\sigma_\theta^2$  or a more random  $\theta$ .

$$\begin{aligned}
 I &= H(\theta) - H(\theta | x) \\
 &= -\int p(\theta) \ln p(\theta) d\theta + \iint p(x, \theta) \ln p(\theta | x) dx d\theta \\
 &= -\iint p(x, \theta) dx \ln p(\theta) d\theta + " \\
 &= -\iint p(x, \theta) \ln p(\theta) dx d\theta \\
 &\quad + \iint p(x, \theta) \ln p(\theta | x) dx d\theta \\
 &= \iint p(x, \theta) \ln \frac{p(\theta | x)}{p(\theta)} dx d\theta \\
 &= \underbrace{\iint p(\theta | x) \ln \frac{p(\theta | x)}{p(\theta)} d\theta}_{\geq 0 \text{ by given inequality}} p(x) dx \\
 &\geq 0 \quad \text{since } p(x) \geq 0. \\
 &= 0 \quad \text{if and only if } p(\theta | x) = p(\theta)
 \end{aligned}$$

Makes sense since if posterior PDF is the same as prior PDF  $\Rightarrow$  no information.

$$\begin{aligned}
 16) \quad H(\theta) &= \frac{1}{2} (1 + \ln 2\pi \sigma_\theta^2) \text{ from Prob 10.15.} \\
 &\text{Similarly, since } p(\theta | x) \text{ is Gaussian,} \\
 &H(\theta | x) = \frac{1}{2} (1 + \ln 2\pi \sigma_{\theta|x}^2) \\
 \Rightarrow I &= H(\theta) - H(\theta | x) = \frac{1}{2} \ln \sigma_\theta^2 / \sigma_{\theta|x}^2
 \end{aligned}$$

$$\text{But } \sigma_{\theta}^2 = \sigma_A^2, \quad \sigma_{\theta|x}^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/N}$$

$$\begin{aligned} I &= \frac{1}{2} \ln \frac{\sigma_A^2 + \sigma^2/N}{\sigma^2/N} \\ &= \frac{1}{2} \ln \left( 1 + \frac{\sigma_A^2}{\sigma^2/N} \right) \end{aligned}$$

- (7) From Prob 10.16 we want to maximize  $I$ . We do so by letting  $\sigma_A^2 \rightarrow \infty$ . It doesn't matter what we choose for  $\mu_A$ . This choice swamps out the prior as we have already observed.

## Chapter 11

- 1) This is just the DC level in WGN or Example 10.1.  
 From (10.11), the MMSE estimator is

$$\hat{\mu} = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2/N} \bar{x} + \frac{\sigma^2/N}{\sigma_0^2 + \sigma^2/N} \mu_0$$

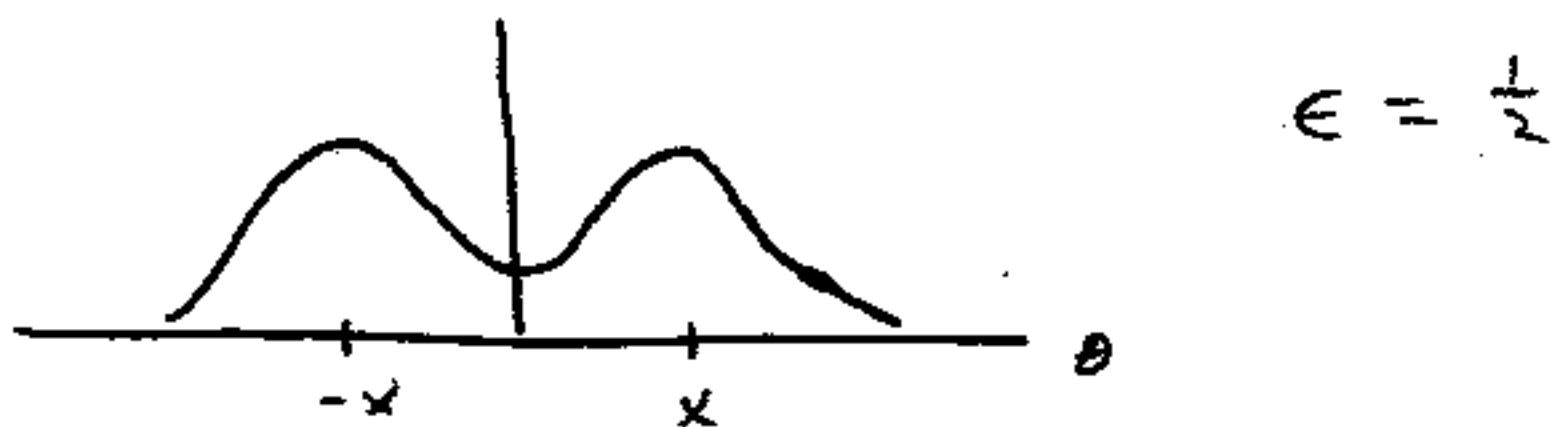
Also, since the posterior PDF is Gaussian

$\Rightarrow$  MMSE estimator = MAP estimator

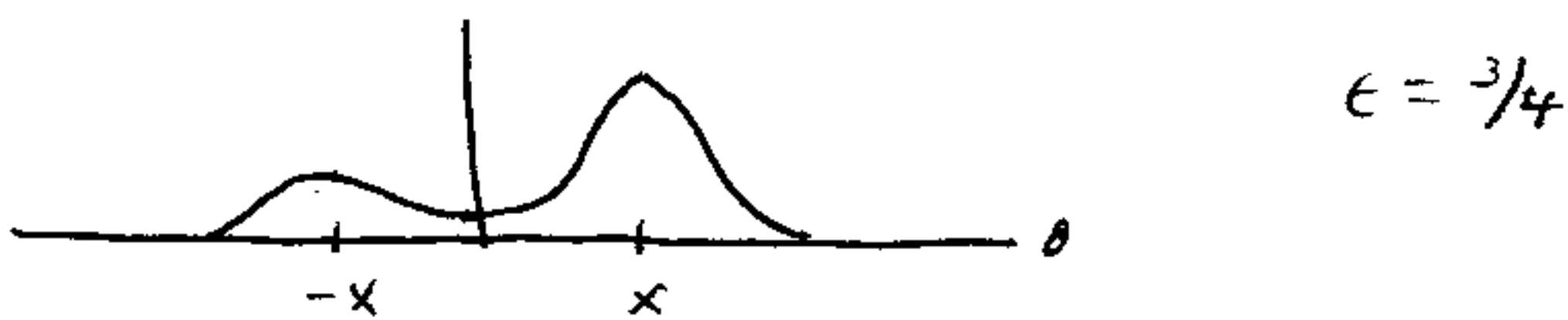
As  $\sigma_0^2 \rightarrow 0$ ,  $\hat{\mu} \rightarrow \mu_0$  prior knowledge dominates

As  $\sigma_0^2 \rightarrow \infty$ ,  $\hat{\mu} \rightarrow \bar{x}$  data " "

2)



MMSE estimator is  $E(\theta|x) = 0$  due to even symmetry of PDF. MAP estimator is mode or  $\pm x$  (not unique)



To find MMSE estimator

$$\begin{aligned}\hat{\theta} &= E(\theta|x) = \int_{-\infty}^{\infty} \theta \frac{\epsilon}{\sqrt{2\pi}} e^{-\frac{1}{2}(\theta-x)^2} d\theta \\ &\quad + \int_{-\infty}^{\infty} \theta \frac{1-\epsilon}{\sqrt{2\pi}} e^{-\frac{1}{2}(\theta+x)^2} d\theta\end{aligned}$$

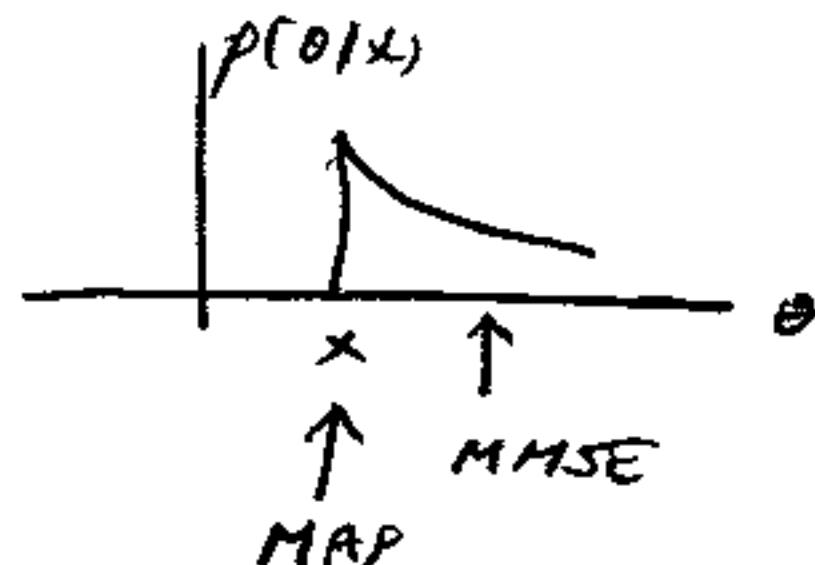
$$= \epsilon x + (1-\epsilon)(1-x) = x(2\epsilon - 1) = x/2$$

The MAP estimator is  $\hat{\theta} = x$ . Note that  
 $MAP \neq MMSE$ .

3) MMSE :

$$\begin{aligned}\hat{\theta} &= \int_x^{\infty} \theta e^{-(\theta-x)} d\theta = e^x [-\theta e^{-\theta} - e^{-\theta}] \Big|_x^{\infty} \\ &= e^x (x e^{-x} + e^{-x}) = x + 1\end{aligned}$$

MAP est. is just  $\hat{\theta} = x$



$$\begin{aligned}4) \quad g(A) &= p(x|A)p(A) \\ &= \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} \sum_n (x(n)-A)^2} A e^{-\lambda A} \quad A > 0\end{aligned}$$

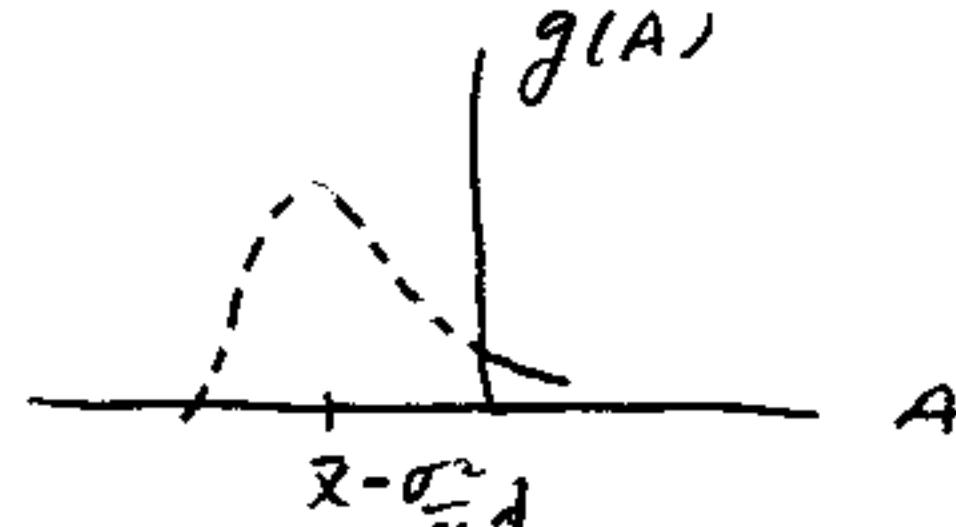
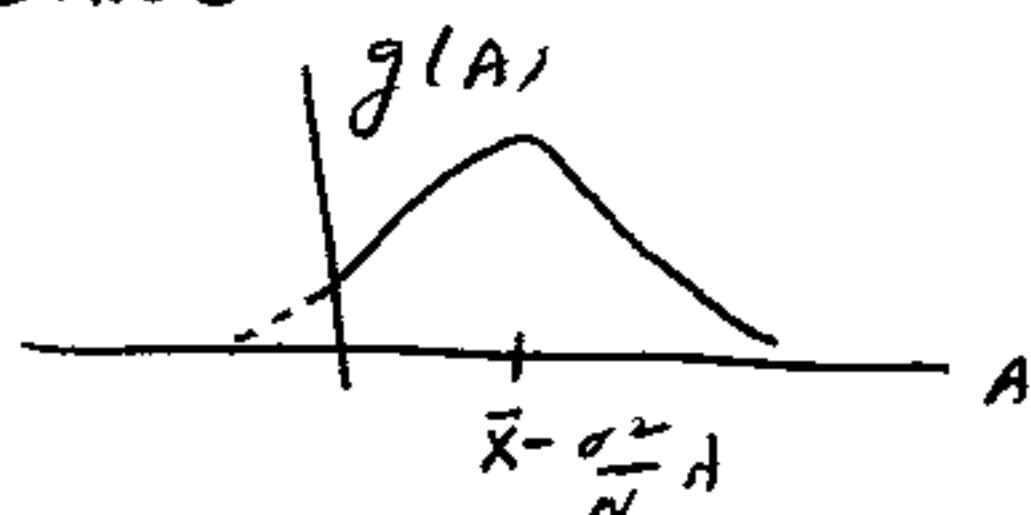
0

$A < 0$

$$\frac{\partial \ln g}{\partial A} = \frac{1}{\sigma^2} \sum_n (x(n)-A) - \lambda = 0$$

$$\Rightarrow \hat{A} = \bar{x} - \frac{\sigma^2}{N} \lambda$$

Note that if  $\hat{A}$  is negative we use  $\hat{A} = 0$   
 since



$$\Rightarrow \hat{A} = \max(0, \bar{x} - \frac{\sigma^2}{N} \lambda)$$

5) From (11.17)

$$\hat{\theta} = E[\underline{\theta} | \underline{x}]$$

$$= E[\underline{\theta}] + \underline{\Sigma}_{\theta x} \underline{\Sigma}_{xx'} (\underline{x} - E(\underline{x}))$$

From (11.27)

$$\text{Bnse}(\hat{\theta}_{ii}) = (\underline{M}\hat{\theta})_{ii}$$

$$= [(\underline{\Sigma}_{\theta\theta} - \underline{\Sigma}_{\theta x} \underline{\Sigma}_{xx'} \underline{\Sigma}_{x\theta})_{ii}]$$

$$\text{If } \underline{\Sigma}_{\theta x} = 0, \quad \hat{\theta} = E[\underline{\theta}]$$

$$\text{Bnse}(\hat{\theta}_{ii}) = (\underline{\Sigma}_{\theta\theta})_{ii} = \text{var}(\theta_i)$$

Data is irrelevant to estimator since  $\underline{\theta}$  and  $\underline{x}$  are independent (due to Gaussian assumption)

6) The Jacobian is

$$\underline{J} = \frac{\partial [\underline{A} \underline{\phi}]^T}{\partial (a, b)^T} = \begin{bmatrix} \frac{\partial A}{\partial a} & \frac{\partial A}{\partial b} \\ \frac{\partial \phi}{\partial a} & \frac{\partial \phi}{\partial b} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{a}{\sqrt{a^2+b^2}} & \frac{b}{\sqrt{a^2+b^2}} \\ \frac{b/a^2}{1+(b/a)^2} & -\frac{1/a}{1+(b/a)^2} \end{bmatrix}$$

$$= \begin{bmatrix} a/A & b/A \\ b/A^2 & -a/A^2 \end{bmatrix}$$

$$|\det \underline{J}| = \left| -a^2/A^3 - b^2/A^3 \right| = 1/A$$

$$p(A, \phi) = \frac{p(a, b)}{|\det \underline{J}|} \quad \text{Note: transformation is 1-1}$$

$$= \frac{1}{2\pi\sigma_0^2} e^{-\frac{1}{2\sigma_0^2}(a^2+b^2)} \frac{1}{A}$$

$$= \frac{A}{\sigma_0^2} e^{-\frac{1}{2\sigma_0^2} A^2} \cdot \frac{1}{2\pi} \quad A > 0 \\ 0 \leq \phi \leq 2\pi$$

$$p(A) = \int_0^{2\pi} p(A, \phi) d\phi = \frac{A}{\sigma_0^2} e^{-\frac{1}{2\sigma_0^2} A^2}$$

$$p(\phi) = \int_0^\infty p(A, \phi) dA = \frac{1}{2\pi}$$

Since  $p(A, \phi)$  factors,  $A$  and  $\phi$  are independent.

- 7) For Bayesian linear model, MMSE estimator  
 = MAP estimator since  $p(\underline{\theta}|x)$  is Gaussian.  
 But MAP estimator maximizes  $p(x|\underline{\theta})p(\underline{\theta})$ .  
 With no prior information this is equivalent  
 to maximizing  $p(x|\underline{\theta})$ . In the Bayesian  
 model  $p(x|\underline{\theta}) = p(x; \underline{\theta})$ . Thus,  
 maximizing  $p(x; \underline{\theta})$ , which yields the  
 MLE or MVU estimator, also yields the  
 MMSE estimator.

8)  $\hat{\underline{\theta}}(n) = E(\underline{\theta}(n)|x) = E(A\underline{\theta}(n-1)|x)$   
 $= A E(\underline{\theta}(n-1)|x) = A \hat{\underline{\theta}}(n-1)$

due to linearity of the expectation

Let  $n=1 \Rightarrow \hat{\underline{\theta}}(1) = A \hat{\underline{\theta}}(0)$

$n=2 \Rightarrow \hat{\underline{\theta}}(2) = A \hat{\underline{\theta}}(1) = A^2 \hat{\underline{\theta}}(0)$

etc.

$$9) \quad \underline{\theta}(n) = \begin{bmatrix} x(n) \\ y(n) \\ nx \\ ny \end{bmatrix} = \begin{bmatrix} x(n-1) + nx \\ y(n-1) + ny \\ nx \\ ny \end{bmatrix}$$

Since  $x(n) - x(n-1) = x(0) + nx - n$   
 $= x(0) - nx(n-1)$   
 $= nx$

and similarly for the  $y$  component.

$$\underline{\theta}(n) = \underbrace{\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_A \underbrace{\begin{bmatrix} x(n-1) \\ y(n-1) \\ nx \\ ny \end{bmatrix}}_{\underline{\theta}(n-1)}$$

$$\Rightarrow \text{from Prob. 11.8 } \hat{\underline{\theta}}(n) = \underline{A}^n \hat{\underline{\theta}}(0)$$

10)  $\hat{\theta} = \frac{1}{\bar{x} + A_N}$  As  $N \rightarrow \infty$ ,  $\hat{\theta} \rightarrow 1/\bar{x}$  and thus  
as  $N \rightarrow \infty$  for a given realization of  $\theta$ ,  $\bar{x} \rightarrow E(x) = 1/\theta \Rightarrow \hat{\theta} \rightarrow \theta$ . In general, as  $N \rightarrow \infty$   
the MAP estimator is just the value that  
maximizes  $p(\underline{x}|\theta)$  or the Bayesian MLE.  
For a given realization of  $\theta$ , if we assume  
 $p(\underline{x}|\theta) = p(\underline{x}; \theta)$ , then we can treat  $\hat{\theta}$   
as an MLE. (The family of PDFs is now  
characterized by  $p(\underline{x}|\theta)$ ). If the MLE is  
consistent, then so will be the MAP estimator.

$$\begin{aligned}
 11) \quad R &= E[c(\underline{\theta})] \\
 &= \iint c(\underline{\theta}) p(\underline{x}, \underline{\theta}) d\underline{x} d\underline{\theta} \\
 &= \underbrace{\int \int c(\underline{\theta}) p(\underline{\theta}|\underline{x}) d\underline{\theta}}_{\text{minimize over } \hat{\underline{\theta}}} p(\underline{x}) d\underline{x}
 \end{aligned}$$

$$\int c(\underline{\theta}) p(\underline{\theta}|\underline{x}) d\underline{\theta} = \int p(\underline{\theta}|\underline{x}) d\underline{\theta} \quad \{\underline{\theta}: \| \underline{\theta} - \hat{\underline{\theta}} \| > \delta\}$$

$$= \int p(\underline{\theta}|\underline{x}) d\underline{\theta} = \int p(\underline{\theta}|\underline{x}) d\underline{\theta} \quad \{\underline{\theta}: \| \underline{\theta} - \hat{\underline{\theta}} \| \leq \delta\}^c$$

$c$  denotes  
Complement

$$= 1 - \int p(\underline{\theta}|\underline{x}) d\underline{\theta} \quad \{\underline{\theta}: \| \underline{\theta} - \hat{\underline{\theta}} \| \leq \delta\}$$

As  $\delta \rightarrow 0$ , this is minimized if the integral is maximized or if we choose  $\hat{\underline{\theta}} = \arg \max_{\underline{\theta}} p(\underline{\theta}|\underline{x})$

12) MAP estimator of  $\underline{\theta}$  maximizes

$$p(\underline{x}|\underline{\theta}) p(\underline{\theta}) = p(\underline{x}, \underline{\theta})$$

$$\text{If } \underline{\alpha} = \underline{A}\underline{\theta}, \quad \frac{\partial \underline{\alpha}}{\partial \underline{\theta}} = \underline{A}$$

$$p(\underline{x}, \underline{\alpha}) = \frac{p(\underline{x}, \underline{\theta})}{|\det \frac{\partial \underline{\alpha}}{\partial \underline{\theta}}|} = \frac{p(\underline{x}, \underline{\theta})}{|\det \underline{A}|}$$

But  $\underline{A}$  does not depend on  $\underline{\alpha}$  and  $\underline{\theta} = \underline{A}^{-1}\underline{\alpha}$

$$\text{so that } p(\underline{x}, \underline{\alpha}) = \frac{p_{x,\theta}(\underline{x}, \underline{A}^{-1}\underline{\alpha})}{|\det \underline{A}|}$$

The MAP estimator of  $\underline{\alpha}$  maximizes  $P_{x,\theta}(\underline{x}, \underline{A}^{-1}\underline{\alpha})$  or because  $\underline{\Omega} = \underline{A}^{-1}\underline{\alpha}$  is invertible we can maximize  $p(\underline{x}, \underline{\theta}) \Rightarrow$  maximizing value is  $\hat{\underline{\theta}}$  and since  $\underline{\alpha} = \underline{A}\underline{\theta} \Rightarrow \hat{\underline{\alpha}} = \underline{A}\hat{\underline{\theta}}$ .

$$(3) \quad \underline{x}^T \underline{C}^{-1} \underline{x} = \underline{x}^T \underline{D}^T \underline{D} \underline{x} = \underline{y}^T \underline{y} \quad \text{where } \underline{y} = \underline{D} \underline{x}$$

$$\begin{aligned} \text{But } \underline{C}_{\underline{y}} &= E(\underline{y} \underline{y}^T) = E(\underline{D} \underline{x} \underline{x}^T \underline{D}^T) = \underline{D} \underline{C} \underline{D}^T \\ &= \underline{D} (\underline{D}^T \underline{D})^{-1} \underline{D}^T = \underline{I} \end{aligned}$$

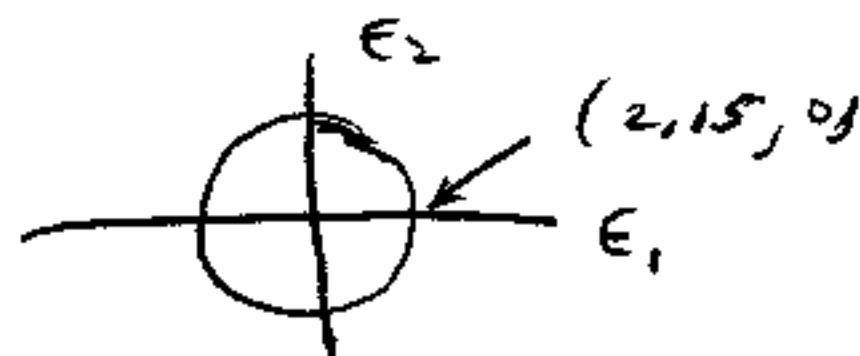
$$\Rightarrow \underline{y} \sim N(\underline{0}, \underline{I}) \text{ and}$$

$$\underline{y}^T \underline{y} = y_1^2 + y_2^2 \sim \chi^2_2$$

since  $y_1 \sim N(0, 1)$  } independent  
 $y_2 \sim N(0, 1)$

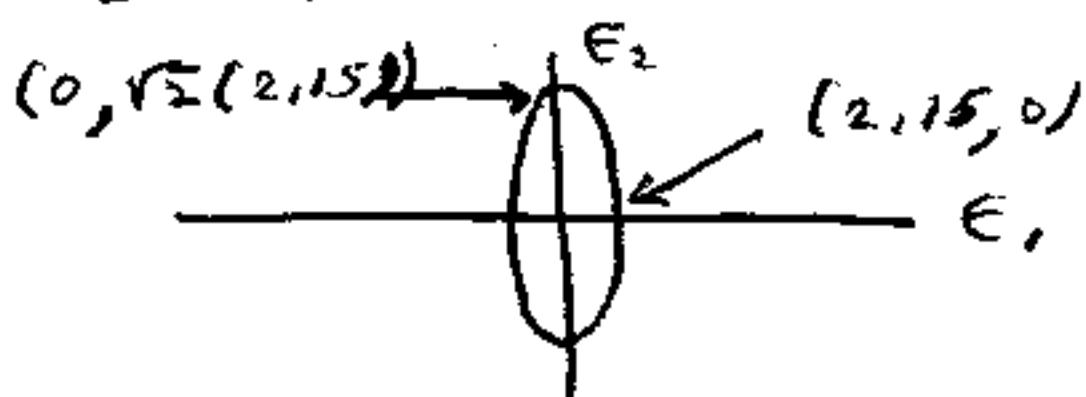
$$(4) \quad \underline{\epsilon}^T \underline{M}_{\hat{\alpha}}^{-1} \underline{\epsilon} = 2 \ln \frac{1}{1-p} = (2, 15)^2$$

$$\text{a)} \quad \underline{\epsilon}^T \underline{\epsilon} = (2, 15)^2$$



$$\text{b)} \quad \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix}$$

$$\underline{\epsilon}^T \underline{M}_{\hat{\alpha}}^{-1} \underline{\epsilon} = \epsilon_1^2 + \frac{1}{2} \epsilon_2^2 = (2, 15)^2$$



$$\text{c) } \underline{M}_{\theta}^{-1} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$$(\epsilon_1, \epsilon_2) \stackrel{\perp}{=} \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix} = (2, 15)^2$$

$$2\epsilon_1^2 + 2\epsilon_2^2 - 2\epsilon_1\epsilon_2 = 3(2, 15)^2$$

$$\epsilon_1^2 - \epsilon_1\epsilon_2 + \epsilon_2^2 = \frac{3}{2}(2, 15)^2$$

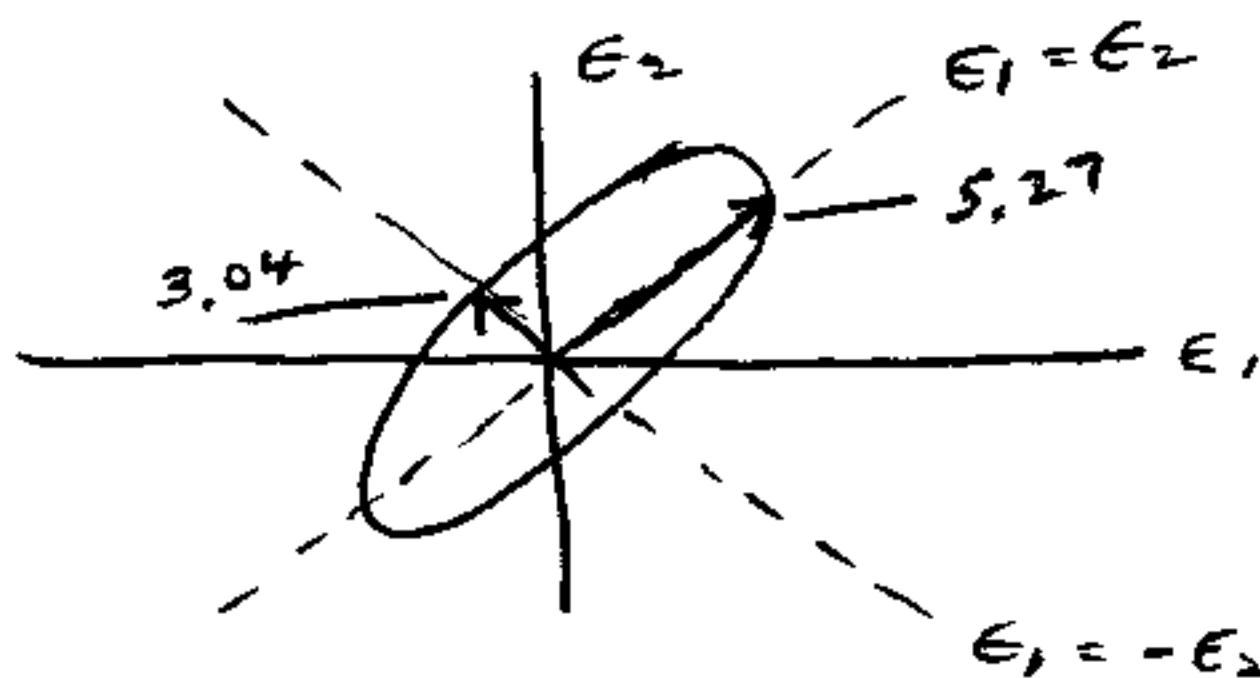
This is a rotated ellipse. It can be rewritten in standard form as

$$\frac{(\epsilon_1 + \epsilon_2)^2}{4} + \frac{(\epsilon_1 - \epsilon_2)^2}{4/3} = \frac{3}{2}(2, 15)^2$$

$$\text{or } \frac{(\epsilon_1 + \epsilon_2)^2}{a^2} + \frac{(\epsilon_1 - \epsilon_2)^2}{b^2} = 1,$$

$$\text{where } a = 5.27$$

$$b = 3.04$$



$$15) \quad \underline{M}_{\theta} = (\underline{C}_{\theta}^{-1} + \underline{H}^T \underline{C}_W^{-1} \underline{H})^{-1} = (\underline{C}_{\theta}^{-1} + \underline{C}_W^{-1})^{-1}$$

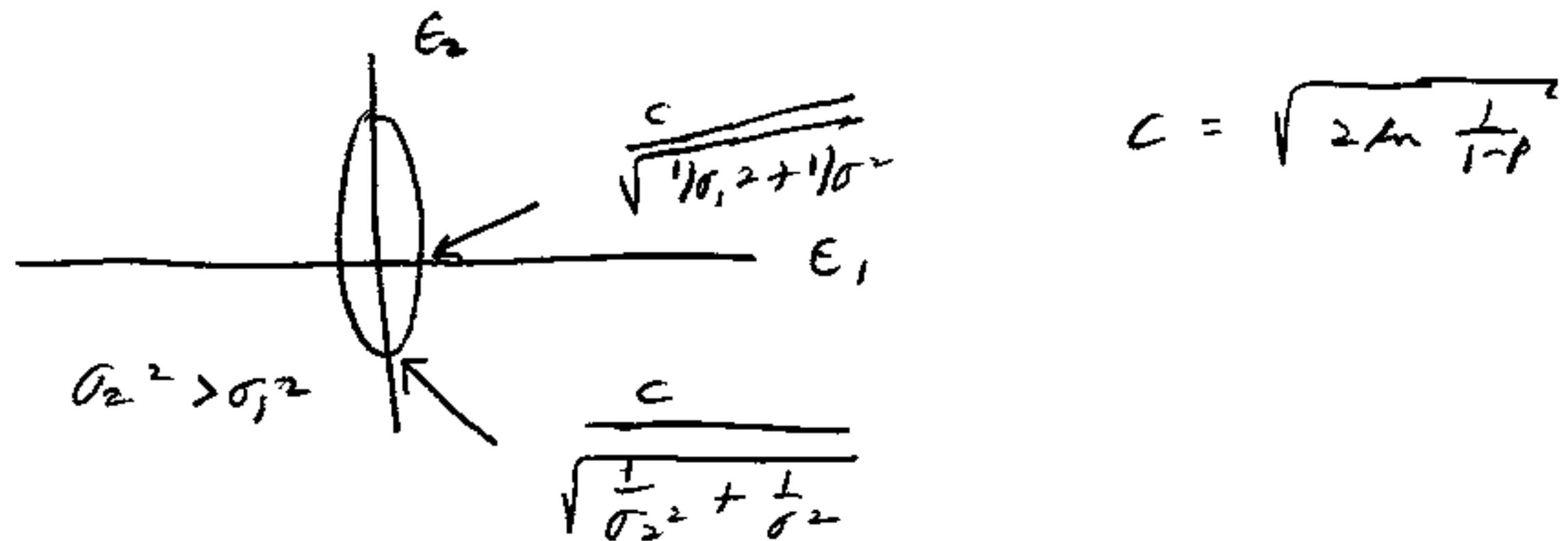
$$\underline{M}_{\theta}^{-1} = \underline{C}_{\theta}^{-1} + \underline{C}_W^{-1} = \begin{bmatrix} 1/\sigma_1^2 + 1/\sigma_2^2 & 0 \\ 0 & 1/\sigma_2^2 + 1/\sigma_1^2 \end{bmatrix}$$

$$\underline{\epsilon}^T \underline{M}_{\theta}^{-1} \underline{\epsilon} = 2 \ln \frac{1}{1-p}$$

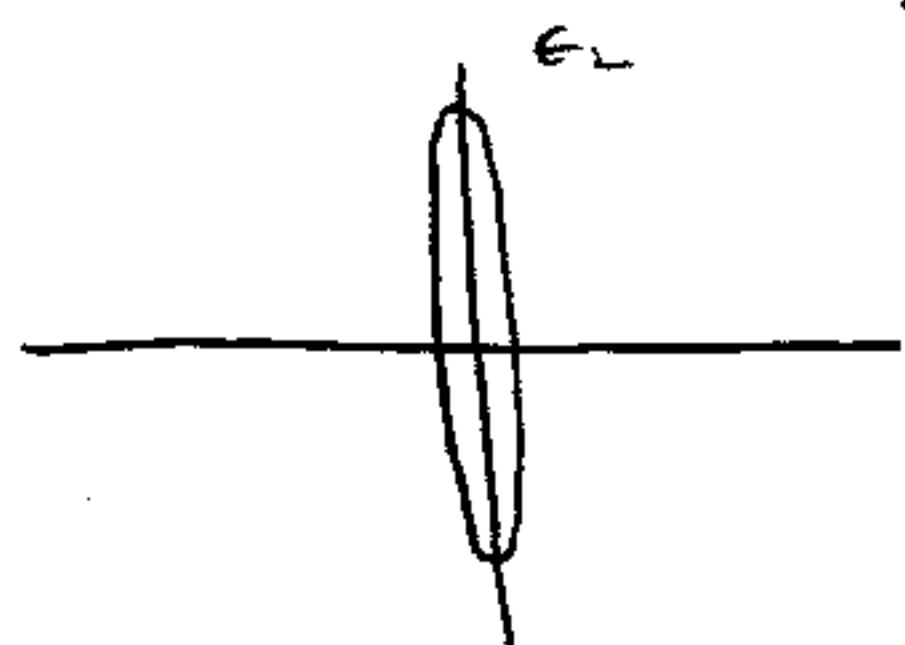
$$\epsilon_1^2 \left( \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right) + \epsilon_2^2 \left( \frac{1}{\sigma_2^2} + \frac{1}{\sigma_1^2} \right) = 2 \ln \frac{1}{1-p}$$

If  $\sigma_1^2 = \sigma_2^2$ , we have a circle.

If  $\sigma_2^2 > \sigma_1^2$ , we have an ellipse



For  $\sigma_2^2 \gg \sigma_1^2$  the ellipse appears as



Most of uncertainty  $\epsilon_1$  is along  $E_2$  direction, as expected.

- 16) This is the Bayesian linear model.  
From (11.38)

$$\hat{\theta} = \underline{m}_0 + (\underline{C}_0^{-1} + \underline{H}^T \underline{C}_W^{-1} \underline{H})^{-1} \underline{H}^T \underline{C}_W^{-1} (\underline{x} - \underline{H} \underline{m}_0)$$

$$\text{where } \underline{m}_0 = [A_0 \ B_0]^T$$

$$\underline{C}_0 = \begin{pmatrix} \sigma_A^{-2} & 0 \\ 0 & \sigma_B^{-2} \end{pmatrix}$$

$$\underline{H} = \begin{bmatrix} 1 & -M \\ 1 & -M+1 \\ \vdots & \vdots \\ 1 & M \end{bmatrix}$$

$$\underline{C}_W = \sigma^2 \underline{I}$$

$$\Rightarrow \underline{H}^T \underline{C}_W^{-1} \underline{H} = \frac{1}{\sigma^2} \underline{H}^T \underline{H} = \frac{1}{\sigma^2} \begin{bmatrix} N & 0 \\ 0 & \sum n^2 \end{bmatrix} \quad \text{where } N = 2M+1$$

$$(\underline{C}_0^{-1} + \underline{H}^T \underline{C}_W^{-1} \underline{H})^{-1} = \begin{bmatrix} \frac{1}{\sigma_A^{-2}} + \frac{N}{\sigma^2} & 0 \\ 0 & \frac{1}{\sigma_B^{-2}} + \frac{\sum n^2}{\sigma^2} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \frac{1}{\sigma_A^2 + \frac{N}{\sigma^2}} & 0 \\ 0 & \frac{1}{\sigma_B^2 + \frac{\sum n^2}{\sigma^2}} \end{bmatrix}$$

$$\underline{H}^T \underline{C}_{\underline{W}}^{-1} (\underline{x} - \underline{H} \underline{A}_0) = \frac{1}{\sigma^2} (\underline{H}^T \underline{x} - \underline{H}^T \underline{H} \underline{A}_0)$$

$$\underline{H}^T \underline{H} = \begin{bmatrix} N & 0 \\ 0 & \sum n^2 \end{bmatrix}$$

$$\underline{H}^T \underline{C}_{\underline{W}}^{-1} (\underline{x} - \underline{H} \underline{A}_0) = \frac{1}{\sigma^2} \begin{bmatrix} \sum x(n) - NA_0 \\ \sum n x(n) - \sum n^2 B_0 \end{bmatrix}$$

$$\hat{A} = A_0 + \frac{\frac{1}{\sigma^2} [\sum x(n) - NA_0]}{\frac{1}{\sigma_A^2} + \frac{N}{\sigma^2}}$$

$$= A_0 + \frac{N\sigma^2}{\frac{1}{\sigma_A^2} + \frac{N}{\sigma^2}} (\bar{x} - A_0)$$

$$\hat{B} = B_0 + \frac{\frac{1}{\sigma^2} [\sum n x(n) - \sum n^2 B_0]}{\frac{1}{\sigma_B^2} + \frac{\sum n^2}{\sigma^2}}$$

$$= B_0 + \frac{\frac{\sum n^2}{n=-M} / \sigma^2}{\frac{1}{\sigma_B^2} + \frac{\sum n^2}{\sigma^2}} \left[ \frac{\frac{\sum n x(n)}{n=-M}}{\frac{\sum n^2}{n=-M}} - B_0 \right]$$

$$B_{mse}(\hat{A}) = [(\underline{C}_{\underline{W}}^{-1} + \underline{H}^T \underline{C}_{\underline{W}}^{-1} \underline{H})^{-1}]_{11}$$

$$= \left( \frac{1}{\sigma_A^2} + \frac{N}{\sigma^2} \right)^{-1}$$

$$\text{Bnse}(\hat{\beta}) = ((\underline{C}_0^{-1} + \underline{H}^T \underline{C}_W^{-1} \underline{H})^{-1})_{22}$$

$$= \left( \frac{1}{\sigma_B^2} + \frac{\sum n^2 / \sigma^2}{\sum n^2 / \sigma^2} \right)^{-1}$$

The intercept ( $\hat{A}$ ) will benefit most from prior knowledge since the reduction in Bnse due to the data is much larger for the slope ( $\sum n^2 / \sigma^2$ ) than for the intercept ( $n / \sigma^2$ ).

$$(17) \quad \underline{\xi} = \begin{bmatrix} 1 & -M \\ 1 & -M+1 \\ \vdots & \vdots \\ 1 & M \end{bmatrix} \underline{\theta} = \underline{H}\underline{\theta}$$

$$\hat{\underline{\xi}} = E(\underline{\xi} | \underline{x}) = E(\underline{H}\underline{\theta} | \underline{x}) = \underline{H}\hat{\underline{\theta}}$$

where  $\hat{\underline{\theta}}$  is given in prob 10.16.

$$\underline{\epsilon} = \underline{\xi} - \hat{\underline{\xi}} = \underline{\xi} - \underline{H}\hat{\underline{\theta}} = \underline{H}(\underline{\theta} - \hat{\underline{\theta}})$$

$$E(\underline{\epsilon}) = \underline{H}(E(\underline{\theta}) - E(\hat{\underline{\theta}}))$$

$$\text{But } E(\hat{\underline{\theta}}) = \underline{\mu}_{\underline{\theta}} \Rightarrow E(\underline{\epsilon}) = \underline{H}(\underline{\mu}_{\underline{\theta}} - \underline{\mu}_{\underline{\theta}}) = \underline{0}$$

The covariance is

$$E(\underline{\epsilon}\underline{\epsilon}^T) = \underline{H} E[(\underline{\theta} - \hat{\underline{\theta}})(\underline{\theta} - \hat{\underline{\theta}})^T] \underline{H}^T$$

$$= \underline{H} \underline{M}\hat{\underline{\theta}} \underline{H}^T$$

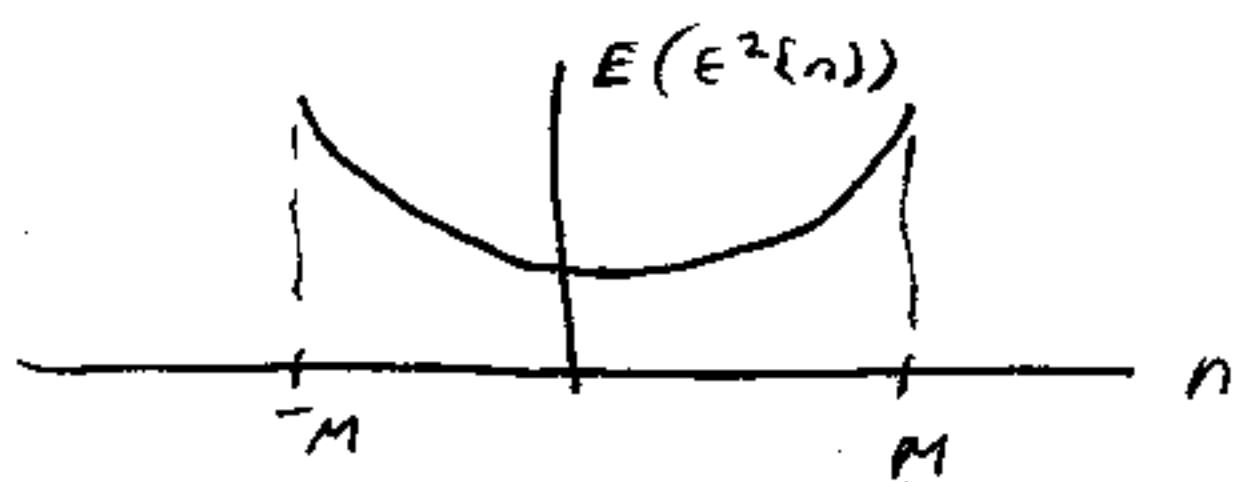
$$= \begin{bmatrix} 1 & -M \\ 1 & -M+1 \\ \vdots & \vdots \\ 1 & M \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_A^2 + n / \sigma^2} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_B^2 + \sum n^2 / \sigma^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sigma_B^2 + \sum n^2 / \sigma^2} \end{bmatrix} \begin{bmatrix} 1 & \dots & 1 \\ -M & \dots & M \end{bmatrix}$$

$$= \frac{1}{\sigma_A^2 + \frac{n}{\sigma^2}} \underline{1}\underline{1}^T + \frac{1}{\sigma_B^2 + \frac{\sum n^2}{\sigma^2}} \underline{m}\underline{m}^T$$

where  $\underline{m} = [-m \dots M]^T$   
 Since  $\underline{x}, \underline{\theta}$  are jointly Gaussian,  $\underline{\epsilon}$  is also Gaussian. Note, however, that  $\underline{\Sigma}_{\epsilon}$  is singular, being of rank 2.

The mean squared error is for  $\hat{x}(n)$ :

$$(\underline{\Sigma}_{\epsilon})_{n+m+1, n+m+1} = \frac{1}{\frac{1}{\sigma_A^2} + \frac{n}{\sigma^2}} + \frac{1}{\frac{1}{\sigma_B^2} + \frac{\sum n^2}{\sigma^2}} n^2$$



As we depart from  $n=0$  the signal estimation error increases. This is because any errors in  $\underline{\theta}$  are magnified by  $n$  due to the signal dependence being  $Bn$ .

18) From (11.40)

$$\begin{aligned}\hat{x} &= \underline{\Sigma}_S (\underline{\Sigma}_S + \sigma^2 \underline{I})^{-1} \underline{x} \\ &= [(\underline{\Sigma}_S + \sigma^2 \underline{I}) \underline{\Sigma}_S^{-1}]^{-1} \underline{x} \\ &= (\underline{I} + \sigma^2 \underline{\Sigma}_S^{-1})^{-1} \underline{x}\end{aligned}$$

$$(\underline{I} + \sigma^2 \underline{\Sigma}_S^{-1}) \hat{x} = \underline{x}$$

$$(\underline{\Sigma}_S + \sigma^2 \underline{I}) \hat{x} = \underline{\Sigma}_S \underline{x}$$

Let  $\underline{x} = (x[-M] \dots x[M])^T$   
 $\underline{s} = (s[-M] \dots s[M])^T$  so that  
 $(\underline{s}\underline{x})_{ij} = E(s[i]x[j]) \quad i, j = -M, \dots, 0, \dots, M$   
 $= r_{ss}(i-j)$

 $\Rightarrow \sum_{j=-M}^M r_{xx}(i-j) \hat{s}(j) = \sum_{j=-M}^M r_{ss}(i-j) x(j)$

for  $i = -M, \dots, M$

As  $M \rightarrow \infty \Rightarrow$

$$\sum_{j=-\infty}^{\infty} r_{xx}(i-j) \hat{s}(j) = \sum_{j=-\infty}^{\infty} r_{ss}(i-j) x(j)$$

$-\infty < i < \infty$

or  $r_{xx}(n) \star \hat{s}(n) = r_{ss}(n) \star x(n)$

Taking Fourier transforms

$$P_{xx}(f) \hat{s}(f) = P_{ss}(f) X(f)$$

$$\Rightarrow H(f) = \frac{P_{ss}(f)}{P_{xx}(f)} = \frac{P_{ss}(f)}{P_{ss}(f) + \sigma^2}$$

Wiener filter emphasizes high SNR regions  
 and attenuates low SNR regions since

$$H(f) = \frac{\eta(f)}{\eta(f) + 1} \quad \text{where } \eta(f) = \frac{P_{ss}(f)}{\sigma^2}$$

If  $\eta(f) \gg 1$  (high SNR),  $H(f) \approx 1$

If  $\eta(f) \ll 1$  (low SNR),  $H(f) \approx 0$ .

Chapter 12

$$1) \text{Bmse}(\hat{\theta}) = E[(\theta - a x^2(0) - b x(0) - c)^2]$$

$$\frac{\partial \text{Bmse}}{\partial a} = -2 E[(\theta - a x^2(0) - b x(0) - c) x^2(0)] = 0$$

$$\frac{\partial \text{Bmse}}{\partial b} = -2 E[(\theta - a x^2(0) - b x(0) - c) x(0)] = 0$$

$$\frac{\partial \text{Bmse}}{\partial c} = -2 E[(\theta - a x^2(0) - b x(0) - c)] = 0$$

$$E(\theta x^2) = a E(x^4) + b E(x^3) + c E(x^2)$$

$$E(\theta x) = a E(x^3) + b E(x^2) + c E(x)$$

$$E(\theta) = a E(x^2) + b E(x) + c$$

$$\begin{bmatrix} E(x^4) & E(x^3) & E(x^2) \\ E(x^3) & E(x^2) & E(x) \\ E(x^2) & E(x) & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} E(\theta x^2) \\ E(\theta x) \\ E(\theta) \end{bmatrix}$$

The minimum MSE is

$$\text{Bmse}(\hat{\theta}) = E[\theta (\theta - \hat{a} x^2(0) - \hat{b} x(0) - \hat{c})^2]$$

$$= E(\theta^2) - \hat{a} E(\theta x^2) - \hat{b} E(\theta x) - \hat{c} E(\theta)$$

Now if  $x(0) \sim U[-1/2, 1/2]$  and  $\theta = \cos 2\pi x(0)$

$$E(x) = 0, E(\theta) = 0$$

$$E(\theta x) = \int_{-\frac{1}{2}}^{\frac{1}{2}} x \cos 2\pi x dx = 0$$

$$E(\theta x^2) = \int_{-\frac{1}{2}}^{\frac{1}{2}} x^2 \cos 2\pi x dx = -\frac{1}{2\pi^2}$$

$$E(x^2) = \int_{-1/2}^{1/2} x^2 dx = \frac{1}{12}$$

$$E(x^3) = 0$$

$$E(x^4) = \int_{-\frac{1}{2}}^{\frac{1}{2}} x^4 dx = \frac{1}{80}$$

$$\begin{bmatrix} \frac{1}{80} & 0 & \frac{1}{12} \\ 0 & \frac{1}{12} & 0 \\ \frac{1}{12} & 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} -\frac{1}{2\pi^2} \\ 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \hat{a} = -90/\pi^2$$

$$\hat{b} = 0$$

$$\hat{c} = 15/2\pi^2$$

$$Bmse(\hat{\theta}) = \frac{1}{2} - (-90/\pi^2)(-1/2\pi^2) - 0 - 0 = 0,04$$

Using a linear estimator we have

$$\hat{\theta} = b x(0) + c$$

$$\begin{bmatrix} E(x^2) & E(x) \\ E(x) & 1 \end{bmatrix} \begin{bmatrix} b \\ c \end{bmatrix} = \begin{bmatrix} E(0x) \\ E(0) \end{bmatrix}$$

$$\text{But } E(0x) = 0, E(0) = 0 \Rightarrow \hat{b} = \hat{c} = 0$$

and  $\hat{\theta} = 0$ . The minimum MSE is  
 $E(\theta^2) = 1/2$ .

2) From (12.27)

$$\hat{A} = u_A + \left( 1/\sigma_A^2 + \frac{\underline{h}^T \underline{h}}{\sigma^2} \right)^{-1} \frac{\underline{h}^T}{\sigma^2} (x - \underline{h} u_A)$$

$$\text{where } \underline{h} = [1, r, \dots, r^{N-1}]^T$$

$$= \mu_A + \frac{\sum_{n=0}^{N-1} r^n (x(n) - r^n \mu_A)}{\frac{1}{\sigma_A^2} + \sum_{n=0}^{N-1} r^{2n}}$$

$$Bmse(\hat{A}) = \frac{1}{\frac{1}{\sigma_A^2} + \frac{1}{r^2} \sum_{n=0}^{N-1} r^{2n}}$$

from (12.29), (12.30).

3)  $\hat{x}_2 = a_1 x_1 + b$

Let  $\theta = x_2$ . Then from (12.6)

$$\begin{aligned}\hat{\theta} &= E(\theta) + C_{\theta x_1} C_{x_1 x_1}^{-1} (x_1 - E(x_1)) \\ &= E(x_2) + \frac{E(x_1 x_2)}{E(x_1^2)} (x_1 - E(x_1))\end{aligned}$$

$$\begin{aligned}Bmse(\hat{\theta}) &= C_{\theta \theta} - C_{\theta x_1} C_{x_1 x_1}^{-1} C_{x_1 \theta} \quad \text{from (12.8)} \\ &= E(x_2^2) - \frac{E^2(x_1 x_2)}{E(x_1^2)}\end{aligned}$$

$$\text{But } C_{x_1 x_2} = \begin{bmatrix} E(x_1^2) & E(x_1 x_2) \\ E(x_2 x_1) & E(x_2^2) \end{bmatrix}$$

This is singular if and only if

$$E(x_1^2) E(x_2^2) - E^2(x_1 x_2) = 0$$

and is equivalent to  $Bmse(\hat{\theta}) = 0$ .

In general, if

$$\hat{x}_1 = \sum_{i=2}^n a_i x_i$$

$$Bmse(\hat{x}_1) = E((x_1 - \hat{x}_1)^2)$$

$$\begin{aligned}
 &= E \left[ (x_i - \sum_{i=2}^N a_i x_i)^2 \right] \\
 &= E[(\underline{b}^T \underline{x})^2] \quad \text{where } b_1 = 1 \\
 &\qquad\qquad\qquad b_i = -a_i \\
 &= \underline{b}^T C_{xx} \underline{b} \quad i = 2, \dots, N
 \end{aligned}$$

But  $C_{xx}$  is positive semidefinite and for the Bmse to be zero we require  $\underline{b}^T C_{xx} \underline{b} = 0$  for some  $\underline{b} \neq \underline{0}$ . Hence,  $C_{xx}$  must be singular.

4)  $(x, x) = E(x^2) \geq 0$  and  $= 0$  if and only if  $x = 0$

$$(x, y) = (y, x) \text{ obvious}$$

$$\begin{aligned}
 (c_1 x_1 + c_2 x_2, y) &= E((c_1 x_1 + c_2 x_2) y) \\
 &= E(c_1 x_1 y + c_2 x_2 y) \\
 &= c_1 E(x_1 y) + c_2 E(x_2 y) \\
 &= c_1 (x_1, y) + c_2 (x_2, y)
 \end{aligned}$$

5)  $(x, x) = 0 \Rightarrow \text{cov}(x, x) = 0 \Rightarrow \text{var}(x) = 0$

$\not\Rightarrow x = 0$  but that  $x$  is a constant

6) Using (12.20)

$$\hat{\underline{s}} = (\underline{s} \underline{x} \underline{C}_{xx}^{-1} \underline{x})$$

$$C_{xx} = E(\underline{x} \underline{x}^T) = R_{ss} + R_{ww} = (\sigma_s^2 + \sigma^2) \underline{I}$$

$$\underline{s} \underline{x} = E(\underline{s} \underline{x}^T) = E(\underline{s} (\underline{s} + \underline{w})^T)$$

$$= E(\underline{s} \underline{s}^T) = \sigma_s^2 \underline{I}$$

$$\hat{\underline{s}} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma^2} \underline{x} \quad \text{or} \quad \hat{s}^{(n)} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma^2} \times L^n$$

From (12.21)

$$\begin{aligned}\underline{M}\hat{\theta} &= \underline{C}_{\theta\theta} - \underline{C}_{\theta x} \underline{C}_{xx}^{-1} \underline{C}_{x\theta} \\ &= \sigma^2 \underline{I} - \frac{(\sigma^2)^2}{\sigma^2 + \sigma^2} \underline{I} \\ &= \sigma^2 \left[ 1 - \frac{\sigma^2}{\sigma^2 + \sigma^2} \right] \underline{I} \\ &= \frac{\sigma^2 \sigma^2}{\sigma^2 + \sigma^2} \underline{I}\end{aligned}$$

$$\begin{aligned}7) \quad \hat{\theta} &= E(\underline{\theta}) + \underline{C}_{\theta x} \underline{C}_{xx}^{-1} (\underline{x} - E(\underline{x})) \\ \underline{M}\hat{\theta} &= E((\underline{\theta} - \hat{\theta})(\underline{\theta} - \hat{\theta})^T) \\ &= E \left[ (\underline{\theta} - E(\underline{\theta}) - \underline{C}_{\theta x} \underline{C}_{xx}^{-1} (\underline{x} - E(\underline{x}))) \cdot \right. \\ &\quad \left. (\underline{\theta} - E(\underline{\theta}) - \underline{C}_{\theta x} \underline{C}_{xx}^{-1} (\underline{x} - E(\underline{x})))^T \right] \\ &= \underline{C}_{\theta\theta} - E \left[ (\underline{\theta} - E(\underline{\theta})) (\underline{x} - E(\underline{x}))^T \right] \underline{C}_{xx}^{-1} \underline{C}_{x\theta} \\ &\quad - \underline{C}_{\theta x} \underline{C}_{xx}^{-1} E \left[ (\underline{x} - E(\underline{x})) (\underline{\theta} - E(\underline{\theta}))^T \right] \\ &\quad + \underline{C}_{\theta x} \underline{C}_{xx}^{-1} \underline{C}_{xx} \underline{C}_{xx}^{-1} \underline{C}_{x\theta} \\ &= \underline{C}_{\theta\theta} - \underline{C}_{\theta x} \underline{C}_{xx}^{-1} \underline{C}_{x\theta} - \underline{C}_{\theta x} \underline{C}_{xx}^{-1} \underline{C}_{x\theta} \\ &\quad + \underline{C}_{\theta x} \underline{C}_{xx}^{-1} \underline{C}_{x\theta} \\ &= \underline{C}_{\theta\theta} - \underline{C}_{\theta x} \underline{C}_{xx}^{-1} \underline{C}_{x\theta}\end{aligned}$$

$$B_{MSL}(\hat{\theta}_i) = E[(\theta_i - \hat{\theta}_i)^2]$$

where the expectation is with respect  
to  $p(\underline{x}, \theta_i)$

$$\begin{aligned}\text{But } B_{MSL}(\hat{\theta}_i) &= \int (\theta_i - \hat{\theta}_i)^2 p(\underline{x}, \theta_i) d\theta_i d\underline{x} \\ &= \int \dots \int (\theta_i - \hat{\theta}_i)^2 p(\underline{x}, \theta) d\theta d\underline{x} \\ \text{since } \hat{\theta}_i \text{ depends on } \underline{x} \text{ only}\end{aligned}$$

$$= \int \cdots \int [(\underline{\theta} - \hat{\theta})(\underline{\theta} - \hat{\theta})^T]_{ii} p(\underline{x}, \underline{\theta}) d\underline{\theta} d\underline{x}$$

$$= \left[ E[(\underline{\theta} - \hat{\theta})(\underline{\theta} - \hat{\theta})^T] \right]_{ii} = [\underline{M}_{\theta}]_{ii}$$

8)  $\hat{\underline{\alpha}} = E(\underline{\alpha}) + C_{\alpha x} (\underline{x}\underline{x}') (\underline{x} - E(\underline{x}))$

$$\text{But } E(\underline{\alpha}) = \underline{A}E(\underline{\theta}) + \underline{b}$$

$$\begin{aligned} C_{\alpha x} &= E[(\underline{\alpha} - E(\underline{\alpha})) (\underline{x} - E(\underline{x}))^T] \\ &= E[\underline{A}(\underline{\theta} - E(\underline{\theta})) (\underline{x} - E(\underline{x}))^T] \\ &= \underline{A} C_{\theta x} \end{aligned}$$

$$\Rightarrow \hat{\underline{\alpha}} = \underline{A}E(\underline{\theta}) + \underline{b} + \underline{A}C_{\theta x} \underline{C}_{xx}' (\underline{x} - E(\underline{x})) \\ = \underline{A}\hat{\underline{\theta}} + \underline{b}$$

$$\hat{\underline{\alpha}} = E(\underline{\alpha}) + C_{\alpha x} (\underline{x}\underline{x}') (\underline{x} - E(\underline{x}))$$

$$\text{If } \underline{\alpha} = \underline{\theta}_1 + \underline{\theta}_2,$$

$$E(\underline{\alpha}) = E(\underline{\theta}_1) + E(\underline{\theta}_2)$$

$$\begin{aligned} C_{\alpha x} &= E[(\underline{\theta}_1 + \underline{\theta}_2 - E(\underline{\theta}_1) - E(\underline{\theta}_2)) (\underline{x} - E(\underline{x}))^T] \\ &= E[(\underline{\theta}_1 - E(\underline{\theta}_1)) (\underline{x} - E(\underline{x}))^T] \\ &\quad + E[(\underline{\theta}_2 - E(\underline{\theta}_2)) (\underline{x} - E(\underline{x}))^T] \\ &= C_{\theta_1 x} + C_{\theta_2 x} \end{aligned}$$

$$\Rightarrow \hat{\underline{\alpha}} = E(\underline{\theta}_1) + E(\underline{\theta}_2) + (C_{\theta_1 x} + C_{\theta_2 x}) \underline{C}_{xx}' (\underline{x} - E(\underline{x})) \\ = \hat{\underline{\theta}}_1 + \hat{\underline{\theta}}_2$$

9)  $\hat{A}(N-1) = \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/N} \bar{x} + \frac{\sigma^2/N}{\sigma_A^2 + \sigma^2/N} MA$

$$\hat{A}(N) = \frac{N\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} \frac{1}{N} \left( \sum_{n=1}^{N-1} x(n) + x(N) \right)$$

$$\begin{aligned}
 & + \frac{\sigma^2}{(N+1)\sigma_A^2 + \sigma^2} u_A \\
 = & \frac{N\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} \frac{\sigma_A^2 + \sigma^2/N}{\sigma_A^2} \underbrace{\frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/N}}_{\frac{1}{N} \sum_0^{N-1} x(n)} \hat{A}(N-1) \\
 & + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} x(N) + \frac{N\sigma_A^2 + \sigma^2}{(N+1)\sigma_A^2 + \sigma^2} \underbrace{\frac{\sigma^2}{N\sigma_A^2 + \sigma^2} u_A}_{\hat{A}(N-1)} \\
 = & \frac{N\sigma_A^2 + \sigma^2}{(N+1)\sigma_A^2 + \sigma^2} \hat{A}(N) + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} x(N) \\
 = & \hat{A}(N-1) + \left( \frac{N\sigma_A^2 + \sigma^2}{(N+1)\sigma_A^2 + \sigma^2} - 1 \right) \hat{A}(N-1) \\
 & + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} x(N) \\
 = & \hat{A}(N-1) - \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} \hat{A}(N-1) \\
 & + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} x(N) \\
 = & \hat{A}(N-1) + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} (x(N) - \hat{A}(N-1))
 \end{aligned}$$

Since  $B_{MS} = (\hat{A}(N-1))$  for the case  $u_A \neq 0$  is also given by (12.31), the rest of the derivation is identical. Thus, we have the identical set of equations.

- 10) This procedure is illustrated in Figure 12.5.

$$\underline{e}_1 = \frac{\begin{bmatrix} 1 \\ 2 \end{bmatrix}}{\sqrt{1+1+4}} = \begin{bmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}$$

$$\begin{aligned} \underline{z}_2 &= \underline{x}_2 - (\underline{x}_2, \underline{e}_1) \underline{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - (1, 0, 1) \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{3}{\sqrt{6}} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} \end{aligned}$$

$$\underline{e}_2 = \frac{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}}{\sqrt{1/4 + 1/4}} = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{bmatrix}$$

$$\begin{aligned} \underline{z}_3 &= \underline{x}_3 - (\underline{x}_3, \underline{e}_1) \underline{e}_1 - (\underline{x}_3, \underline{e}_2) \underline{e}_2 \\ &= \begin{bmatrix} 1/3 \\ 1/3 \\ -1/3 \end{bmatrix} \end{aligned}$$

$$\underline{e}_3 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ -1/\sqrt{3} \end{bmatrix}$$

$$(1) \quad y_1 = \frac{x_1}{\sqrt{\text{var}(x_1)}} = x_1$$

$$\begin{aligned} z_2 &= x_2 - (x_2, y_1) y_1 = x_2 - E(x_2 y_1) y_1 \\ &= x_2 - E(x_1 x_2) x_1 = x_2 - \rho x_1 \end{aligned}$$

$$y_2 = \frac{x_2 - \rho x_1}{\sqrt{E((x_2 - \rho x_1)^2)}} = \frac{x_2 - \rho x_1}{\sqrt{1 - \rho^2 + \rho^2}} = \frac{x_2 - \rho x_1}{\sqrt{1 - \rho^2}}$$

$$z_3 = x_3 - (x_3, y_1) y_1 - (x_3, y_2) y_2$$

$$\begin{aligned}
 &= x_3 - E(x_3 x_1) x_1 - E\left(x_3 \frac{x_2 - \rho x_1}{\sqrt{1-\rho^2}}\right) \frac{x_2 - \rho x_1}{\sqrt{1-\rho^2}} \\
 &= x_3 - \rho^2 x_1 - \frac{\rho - \rho^3}{\sqrt{1-\rho^2}} \frac{x_2 - \rho x_1}{\sqrt{1-\rho^2}} \\
 &= x_3 - \rho^2 x_1 - \rho(x_2 - \rho x_1) = x_3 - \rho \cdot x_2
 \end{aligned}$$

$$y_3 = \frac{x_3 - \rho x_2}{\sqrt{E((x_3 - \rho x_2)^2)}} = \frac{x_3 - \rho x_2}{\sqrt{1-\rho^2}}$$

$$\underline{y} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ -\frac{\rho}{\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} & 0 \\ 0 & -\frac{\rho}{\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} \end{bmatrix}}_{A} \underline{x}$$

$\underline{A}$  is lower triangular and will be in general.

$$\text{Since } \underline{\Sigma}_{yy} = \underline{\Sigma} \Rightarrow \underline{\Sigma}_{yy} = \underline{A} \underline{\Sigma}_{xx} \underline{A}^T$$

$$\underline{\Sigma}_{xx} = \underline{A}^{-1} \underline{A}^{T-1} \Rightarrow \underline{\Sigma}_{xx^{-1}} = \underline{A}^T \underline{A}$$

Thus,  $\underline{A}$  is a whitening transformation and  $\underline{A}^T \underline{A}$  provides a Cholesky decomposition of  $\underline{\Sigma}_{xx^{-1}}$ .

- 12) From (12.47) for a scalar parameter so that  $a[n-1]$  and  $b[n]$  are scalars, as  $\sigma_n^2 \rightarrow 0$

$$K[n] \rightarrow \frac{M[n-1] h[n]}{M[n-1] h^2[n]} = \frac{1}{h[n]}$$

$$\Rightarrow \hat{\theta}[n] = \hat{\theta}[n-1] + \frac{1}{h[n]} (x[n] - h[n] \hat{\theta}[n-1]) \\ = x[n] / h[n]$$

We discard all previous data since  $x[n]$  is a perfect measurement (no noise). Also,

$$M[n] = (1 - K[n] h[n]) M[n-1] = 0$$

In vector case we do not obtain analogous result since we require  $p$  noiseless measurements to determine  $\underline{\theta}$ .

13) Here  $h[n] = 1$ ,  $\sigma_n^2 = \sigma^2$

$$\Rightarrow \hat{A}[n] = \hat{A}[n-1] + K[n] (x[n] - \hat{A}[n-1])$$

$$K[n] = \frac{M[n-1]}{\sigma^2 + M[n-1]}$$

$$M[n] = (1 - K[n]) M[n-1]$$

This is the same as the introductory example of last 12.6 except for the initialization since

$$\hat{A}[-1] = E(A) = 0$$

$$M[-1] = E[(A - \hat{A}[-1])^2] = E(A^2) = A_0^2 / 3$$

Solving for  $K[n]$ :

$$M[n] = \left(1 - \frac{M[n-1]}{\sigma^2 + M[n-1]}\right) M[n-1]$$

$$= \frac{\sigma^2 M[n-1]}{\sigma^2 + M[n-1]} = K[n] \sigma^2$$

$$\Rightarrow K[n] = \frac{KL[n-1] \sigma^2}{\sigma^2 + KL[n-1] \sigma^2} = \frac{K[n-1]}{1 + KL[n-1]}$$

$$\frac{i}{KL[n]} = \frac{i}{KL[n-1]} + 1 \quad n \geq 1$$

$$K[0] = \frac{M[-1]}{\sigma^2 + M[-1]} = \frac{A_0^2/3}{\sigma^2 + A_0^2/3}$$

$$\frac{i}{KL[0]} = i + \frac{\sigma^2}{A_0^2/3}$$

$$\frac{i}{KL[n]} = \frac{i}{KL[0]} + n = (n+1) + \frac{\sigma^2}{A_0^2/3}$$

$$K[n] = \frac{A_0^2/3}{(n+1)A_0^2/3 + \sigma^2}$$

$$1 - K[n] = \frac{nA_0^2/3 + \sigma^2}{(n+1)A_0^2/3 + \sigma^2}$$

$$\begin{aligned}\hat{A}[n] &= (1 - K[n]) \hat{A}[n-1] + K[n] x[n] \\ &= \frac{nA_0^2/3 + \sigma^2}{(n+1)A_0^2/3 + \sigma^2} \hat{A}[n-1] \\ &\quad + \frac{A_0^2/3}{(n+1)A_0^2/3 + \sigma^2} x[n]\end{aligned}$$

where  $\hat{A}[-1] = 0$

$$\hat{A}[0] = \frac{A_0^2/3}{A_0^2/3 + \sigma^2} x[0]$$

$$\hat{A}[1] = \frac{A_0^2/3 + \sigma^2}{2A_0^2/3 + \sigma^2} - \frac{A_0^2/3}{A_0^2/3 + \sigma^2} x[0]$$

$$\begin{aligned}
 & + \frac{A_0^2/3}{2 A_0^2/3 + \sigma^2} x(1) \\
 = & \frac{A_0^2/3}{2 A_0^2/3 + \sigma^2} (x(0) + x(1)) \\
 = & \frac{A_0^2/3}{A_0^2/3 + \sigma^2/2} \underbrace{\frac{1}{2} (x(0) + x(1))}_{\bar{x}}
 \end{aligned}$$

etc.

- (4) To minimize  $E[(x(n) - \hat{x}(n))^2]$  we use the orthogonality principle or

$$E[(x(n) - \hat{x}(n)) x(n-l)] = 0 \quad l = -M, \dots, M$$

$$l \neq 0$$

$$\begin{aligned}
 r_{xx}(l) &= E \left[ \sum_k a_k x(n-k) x(n-l) \right] \\
 &= \sum_k a_k r_{xx}(l-k)
 \end{aligned}$$

To show that  $a_{-k} = a_k$ :

Let  $k' = -k$

$$\begin{aligned}
 r_{xx}(l) &= \sum_{\substack{k'= -M \\ k' \neq 0}}^M a_{-k'} r_{xx}(l+k') \\
 &\quad l = -M, \dots, M \\
 &\quad l \neq 0
 \end{aligned}$$

Let  $k' = -l$

$$\begin{aligned}
 r_{xx}(l-k') &= \sum_{\substack{k'= -M \\ k' \neq 0}}^M a_{-k'} r_{xx}(-k'+k') \\
 &\quad l' = -M, \dots, M \\
 &\quad l' \neq 0
 \end{aligned}$$

$$\begin{aligned}
 r_{xx}(l-k') &= \sum_{\substack{k'= -M \\ k' \neq 0}}^M a_{-k'} r_{xx}[l'-k']
 \end{aligned}$$

Since  $r_{xx}(-k) = r_{xx}(k)$

But these are the same set of equations for which there is a unique solution. Hence,  $a_{-k} = a_k$ . This must be true since the correlation of  $x(n)$  with  $x(n+k)$  is the same as that with  $x(n-k)$ , due to the even symmetry of the ACF.

$$15) \quad W = \frac{r_{ss}(0)}{r_{ss}(0) + r_{ww}(0)} = \eta_{n+1}$$

$$M\hat{s} = \left(1 - \frac{1}{\eta_{n+1}}\right) r_{ss}(0)$$

$$\begin{aligned} \text{But } \rho &= \frac{E(s(0)(s(0) + W(0)))}{\sqrt{r_{ss}(0)(r_{ss}(0) + r_{ww}(0))}} \\ &= \frac{r_{ss}(0)}{\sqrt{r_{ss}(0)(r_{ss}(0) + r_{ww}(0))}} \\ &= \frac{1}{\sqrt{1 + \eta_{n+1}}} = \sqrt{\eta_{n+1}} \end{aligned}$$

$$\Rightarrow W = \rho^2 \quad \text{or} \quad \hat{s}(0) = \rho^2 s(0)$$

$$M\hat{s} = (1 - \rho^2) r_{ss}(0)$$

The higher the SNR,  $\eta$ , the larger is  $\rho$  and hence the better is the estimator.

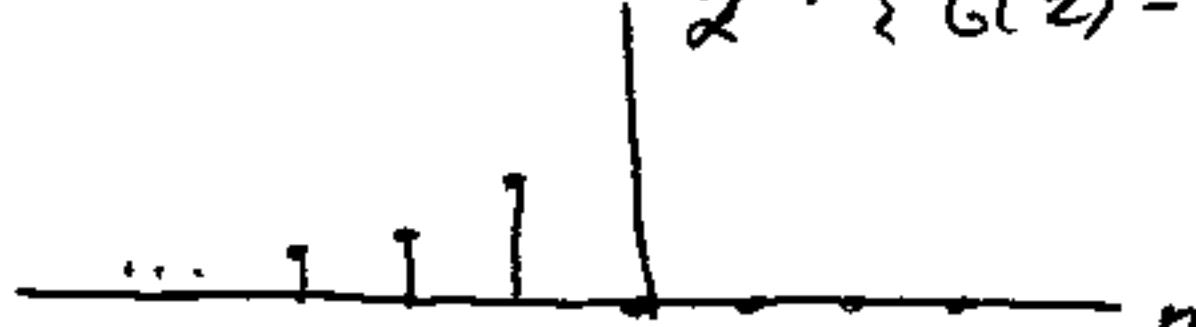
$$16) \quad \left[ B(z^{-1}) \left( G(z) + \frac{P_{SS}(z)}{B(z^{-1})} \right) \right]_+ = 0$$

But  $B(z^{-1})$  is the z-transform of an antisignal sequence and thus if

$G(z) - \frac{P_{55}(z)}{B(z^{-1})}$  is also antisignal

with a sequence value of zero at  $n = 0$  or

$$2^{-1} \{ G(z) - \rho_{SV}(z) / \beta(z^{-1}) \}$$



the convolution will be zero for  $n \geq 0$  as required. Now, since  $\text{Psr}(z)/B(z^{-1})$  is a two-sided sequence we let

$$\mathcal{Z}^{-1}\{G(z)\} = \mathcal{Z}^{-1}\left\{\frac{P_{S^n}(z)}{B(z^{-1})}\right\} \text{ for } n \geq 0$$

$$\text{or } G(z) = \left[ \frac{\rho_{ss}(z)}{B(z^{-1})} \right]_+$$

$$\text{or } H(z) = \frac{1}{B(z)} \left[ \rho_{xx}(z)/B(z^{-1}) \right]_+$$

$$17) \quad E \left\{ (s[n] - \hat{s}[n]) x[n-\ell] \right\} = 0 \quad -\infty < \ell < \infty$$

$$E((S_{L,n} - \sum_k h(k) \times (n-k)) \times (n-\ell)) = 0$$

$$E(s(n) \times (n-\ell)) = \sum_k h(k) E(x(n-k) \times (n-\ell))$$

$$r_{ss}(\ell) = \sum_{k=-\infty}^{\infty} h(k) r_{xx}(\ell-k)$$

$$\begin{aligned}
 18) M_{\hat{s}} &= E \{ (s[n] - \hat{s}[n])^2 \} \\
 &= E \{ (s[n] - \hat{s}[n])(s[n] - \sum_k h(k) \times (n-k)) \} \\
 &= E \{ (s[n] - \hat{s}[n]) s[n] \} \\
 &\quad - \underbrace{E \{ (s[n] - \hat{s}[n]) \sum_k h(k) \times (n-k) \}}_{= 0 \text{ by orthogonality principle}} \\
 &= P_{ss}(0) - \sum_k h(k) \underbrace{E \{ x[n-k] s[n] \}}_{r_{ss}(k)}
 \end{aligned}$$

Now by Parseval's theorem

$$\sum_k h(k) r_{ss}(k) = \int P_{ss}(f) H(f) df$$

$$\begin{aligned}
 \Rightarrow M_{\hat{s}} &= \int P_{ss}(f) df - \int P_{ss}(f) H(f) df \\
 &= \int_{-\frac{1}{2}}^{\frac{1}{2}} P_{ss}(f) (1 - H(f)) df
 \end{aligned}$$

For the example given

$$H(f) = \begin{cases} \frac{P_0}{P_0 + \sigma^2} & |f| \leq \frac{1}{4} \\ 0 & \frac{1}{4} < |f| \leq \frac{1}{2} \end{cases}$$

$$M_{\hat{s}} = \int_{-\frac{1}{4}}^{\frac{1}{4}} P_0 \left( 1 - \frac{P_0}{P_0 + \sigma^2} \right) df = \frac{1}{2} \frac{P_0 \sigma^2}{P_0 + \sigma^2}$$

Since there is no signal power above  $f = \frac{1}{4}$ ,  $H(f) = 0$ . For  $f \leq \frac{1}{4}$  we weight all frequencies (since signal and noise have flat PSDs) by an SNR weighting or

$$H(f) = \frac{P_0}{P_0 + \sigma^2} = \frac{\eta}{\eta + 1}$$

19)  $\hat{x}(n) = \sum_{k=1}^N h(k)x(n-k)$

$$E[(x(n) - \hat{x}(n)) x(n-l)] = 0 \quad l = 1, 2, \dots, N$$

$$r_{xx}[l] = \sum_{k=1}^N h(k) E(x(n-k)x(n-l))$$

$$= \sum_{k=1}^N h(k) r_{xx}[l-k]$$

The equations are independent of  $n$  (in deriving (12.65) we assumed  $n = N$  was the index of the sample to be predicted) since the ACF does not depend on  $n$ .

$$\begin{aligned} M_2 &= E[(x(n) - \hat{x}(n)) x(n)] \\ &\quad - \underbrace{E[(x(n) - \hat{x}(n)) \hat{x}(n)]}_{= 0 \text{ by orthogonality principle}} \end{aligned}$$

$$= E[x^2(n)] - \sum_{k=1}^N h(k) E[(n-k)x(n)]$$

$$= r_{xx}(0) - \sum_{k=1}^N h(k) r_{xx}(k)$$

20) From the previous problem we have

$$r_{xx}[l] = \sum_{k=1}^N h(k) r_{xx}[l-k] \quad l = 1, 2, \dots, N$$

must be solved for the optimal one-step predictor. But for an AR( $N$ ) process we know that (see Appendix 1)

$$r_{xx}(k) = - \sum_{h=1}^N a(h) r_{xx}(k-h) \quad k \geq 1$$

which are the Yule-Walker equations. Since the solution for the  $h(k)$ 's is unique,

$$h(k) = -a(k)$$

so that  $\hat{x}[n] = -\sum_{h=1}^N a(h) \times [n-h]$  and the MMSE is

$$M_x^2 = r_{xx}(0) - \sum_{h=1}^N h(h) r_{xx}(h)$$

$$= r_{xx}(0) + \sum_{h=1}^N a(h) r_{xx}(h)$$

$$= \sigma_u^2 \quad (\text{see Appendix 1})$$

### Chapter 13

$$1) \quad s[n] = a^{n+1} s[-1] + \sum_{k=0}^n a^k u[n-k]$$

$$\text{Let } \underline{s} = [s[n_1] \ s[n_2] \ \dots \ s[n_k]]^T$$

Assume  $n_k > n_{k-1} > \dots > n_1$

$$\underline{s} = \begin{bmatrix} a^{n_1+1} & a^{n_1} & \dots & 1 & 0 & 0 & \dots & 0 \\ a^{n_2+1} & a^{n_2} & \dots & & 1 & 0 & \dots & 0 \\ \vdots & & & & & & & \\ a^{n_k+1} & a^{n_k} & \dots & & & & & \end{bmatrix} \begin{bmatrix} s[-1] \\ u[0] \\ \vdots \\ u[n_k] \end{bmatrix}$$

Since  $s[-1], u[0], \dots, u[n_k]$  are independent, they are jointly Gaussian and thus  $\underline{s}$  has a multivariate Gaussian PDF, being a linear transformation.

$$2) \quad \text{If } u_r = 0, \text{ from (13.4), } E(s[n]) = 0$$

$$\begin{aligned} \{s[m, n]\} &= a^{m+n+2} \sigma_s^2 + \sigma_u^2 a^{m-n} \\ &\quad \cdot \sum_{k=0}^n a^{2k} \\ &= \frac{a^{m+n+2} \sigma_u^2}{1-a^2} + \frac{\sigma_u^2 a^{m-n} (1-a^{2(n+1)})}{1-a^2} \end{aligned}$$

$$= \frac{\sigma_u^2}{1-a^2} a^{m-n}$$

$\Rightarrow$  Now. By setting the initial condition as given the process is in steady-state for  $n = -1$ .

$$3) E(s(n)) = a^{n+1} \mu_s = 5(0.98)^{n+1}$$

$$\text{var}(s(n)) = a^{2n+2} \sigma_s^2 + \sigma_u^2 \sum_{k=0}^n a^{2k}$$

$$= (0.98)^{2n+2} + 0.1 \sum_{k=0}^n (0.98)^{2k}$$

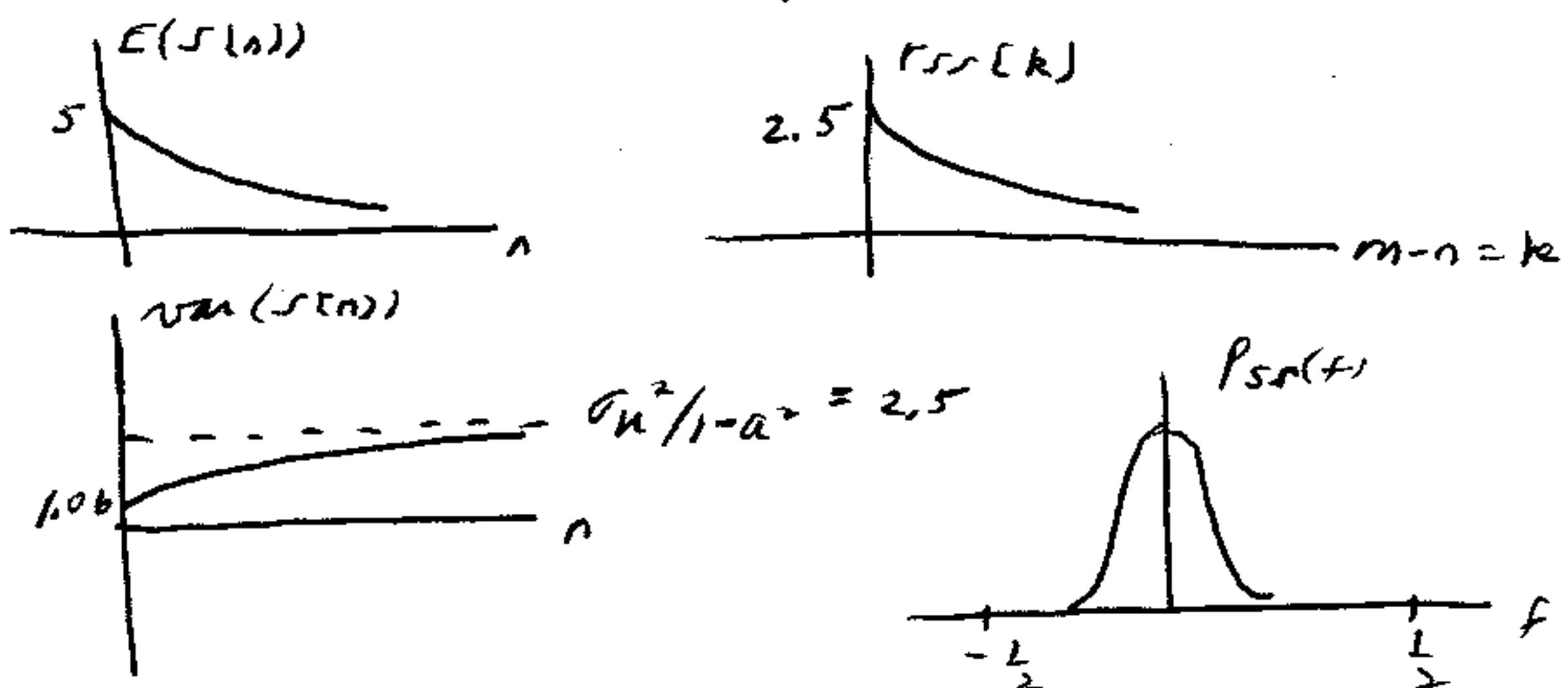
$$C_s(m, n) \rightarrow \frac{\sigma_u^2}{1-a^2} a^{m-n}$$

$$= \frac{0.1}{1-0.98^2} 0.98^{m-n}$$

PSD is just that of an AR(1) process or

$$P_{ss}(f) = \frac{\sigma_u^2}{|1-a e^{-j2\pi f}|^2}$$

$$= \frac{0.1}{|1-0.98 e^{-j2\pi f}|^2}$$



4) From (13.5)

$$C_s(m, n) = a^{m-n} \underbrace{\left[ a^{2n+2} \sigma_s^2 + \sigma_u^2 \sum_{k=0}^n a^{2k} \right]}_{\text{var}(s(n))}$$

$$\underline{c}(m, n) = \underline{a}^{m-n} \underline{c}(n, n)$$

5)  $\underline{s}(n) = \underline{A} \underline{s}(n-1) + \underline{B} \underline{u}(n)$

$$\begin{aligned} E(\underline{s}(n)) &= \underline{A} E(\underline{s}(n-1)) + \underline{B} E(\underline{u}(n)) \\ &= \underline{A} E(\underline{s}(n-1)) \end{aligned}$$

$$\begin{aligned} \underline{c}(n) &= E((\underline{s}(n) - E(\underline{s}(n))) (\underline{s}(n) - E(\underline{s}(n))^T)) \\ &= E((\underline{A} \underline{s}(n-1) + \underline{B} \underline{u}(n) - \underline{A} \underline{s}(n-1))^T) \\ &\quad ( " )^T \end{aligned}$$

$$= E((\underline{A} (\underline{s}(n-1) - E(\underline{s}(n-1))) + \underline{B} \underline{u}(n))^T)$$

$$= \underline{A} \underline{c}(n-1) \underline{A}^T + \underline{B} \underline{Q} \underline{B}^T \text{ since}$$

$$E(\underline{s}(n-1) \underline{u}(n)^T) = 0$$

( $\underline{s}(n-1)$  depends on  $\underline{u}(k)$  for  $k \leq n-1$ ).

6) From (3.14)  $E(\underline{s}(n)) = \underline{A}^{n+1} \underline{u}_s$

Let  $\underline{V}$  be the modal matrix for  $\underline{A}$

and  $\underline{\Lambda}$  the diagonal matrix of eigenvalues.

Thus,  $\underline{V}^T \underline{A} \underline{V} = \underline{\Lambda}$  or  $\underline{A} = \underline{V} \underline{\Lambda} \underline{V}^T$

$$\Rightarrow \underline{A}^{n+1} = \underline{V} \underline{\Lambda}^{n+1} \underline{V}^T$$

$$E(\underline{s}(n)) = \underline{V} \underline{\Lambda}^{n+1} \underline{V}^T \underline{u}_s$$

$$= \sum_{i=1}^p a_i \lambda_i^{n+1} \underline{v}_i$$

where  $a_i = (\underline{v}^T \underline{u}_i)_i$

$$\underline{v} = (\underline{v}_1 \underline{v}_2 \dots \underline{v}_p)$$

If any  $|a_i| > 1$   $E(\underline{s}(n)) \rightarrow \infty$

If all  $|a_i| < 1$   $E(\underline{s}(n)) \rightarrow 0$

$$7) \quad \underline{A}^n = (\underline{v} \underline{A} \underline{v}^T)^n = \underline{v} \underline{A}^n \underline{v}^T$$

$$\underline{A}^n \underline{s} \underline{A}^{nT} = \underline{v} \underline{A}^n \underline{v}^T \underline{s} \underline{A}^{nT} \underline{v}$$

$$\underline{e}_i^T \underline{A}^n \underline{s} \underline{A}^{nT} \underline{e}_j = \underbrace{\underline{e}_i^T \underline{v} \underline{A}^n (\underline{v}^T \underline{s} \underline{v}) \underline{A}^n \underline{v}^T \underline{e}_j}_{b^T}$$

$$\text{But } b = \underline{A}^n \underline{v} \underline{e}_i \rightarrow 0 \text{ if } |a_i| < 1$$

$$a = \underline{A}^n \underline{v} \underline{e}_j \rightarrow 0 \text{ if } |a_j| < 1$$

$$\Rightarrow b^T \underline{v}^T \underline{s} \underline{v} a = [A^n \underline{s} A^{nT}]_{ij} \rightarrow 0$$

for all  $i, j$

$$8) \quad \underbrace{\begin{bmatrix} r(n-p+1) \\ r(n-p+2) \\ \vdots \\ r(n) \end{bmatrix}}_{\underline{s}(n)} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & & \\ -a(p) & -a(p-1) & \dots & -a(1) & \end{bmatrix}}_A \underbrace{\begin{bmatrix} r(n-p) \\ r(n-p+1) \\ \vdots \\ r(n-1) \end{bmatrix}}_{\underline{s}(n-1)}$$

$$+ \underbrace{\begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & & & \\ b(q) & b(q-1) & \dots & 1 \end{bmatrix}}_B \underbrace{\begin{bmatrix} u(n-q) \\ u(n-q+1) \\ \vdots \\ u(n) \end{bmatrix}}_{\underline{u}(n)}$$

Not a Gauss-Markov process since  $\underline{u}(n)$  is not vector WGN. This is because  $\underline{u}(n)$  is correlated in time. If  $g=1$  for example

$$\begin{aligned} E(\underline{u}(n) \underline{u}(n+1)^T) &= E \begin{bmatrix} u(n-1) \\ u(n) \end{bmatrix} \begin{bmatrix} u(n) u(n+1) \end{bmatrix} \\ &= \begin{bmatrix} E[u(n-1)u(n)] & E[u(n-1)u(n+1)] \\ E[u^2(n)] & E[u(n)u(n+1)] \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ \sigma_u^2 & 0 \end{bmatrix} \neq 0 \end{aligned}$$

$$g) r(n) = s(n) + \sum_{l=1}^g b(l) s(n-l)$$

$$\begin{aligned} \sum_{k=0}^p a(k) r(n-k) &= \sum_{k=0}^p a(k) s(n-k) \\ &\quad + \sum_{l=1}^g b(l) \sum_{k=0}^p a(k) s(n-k+l) \end{aligned}$$

$$= u(n) + \sum_{l=1}^g b(l) u(n-l)$$

Now let the state

be given by (13.10)

$$\Rightarrow \underline{s}(n) = A \underline{s}(n-1) + B \underline{u}(n) \quad (\text{see development leading to (13.11)})$$

Since  $\underline{s}(n) = \begin{bmatrix} s(n-p+1) \\ s(n-p+2) \\ \vdots \\ s(n) \end{bmatrix}$ , if  $g \leq p-1$

$$r(n) = (\underbrace{0 \dots 0}_{h^T(n)} b(g) \dots b(1) \ 1) \underline{s}(n)$$

$$\Rightarrow \underline{x}(n) = \underline{h}^T \underline{s}(n) + \underline{w}(n)$$

Now use (13.51) - (13.56) with

$$\underline{A} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & & \vdots \\ -a(p) & -a(p-1), \dots, -a(1) \end{bmatrix} \quad \underline{B} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\underline{Q} = \sigma_u^2$$

- 10) From (13.38) - (13.42) with  $a=1$ ,  $\sigma_u^2 = 0$   
 so that  $s(n) = s(n-1) = A$ . We can skip  
 the prediction stage since

$$\hat{A}(n|n-1) = \hat{A}(n-1|n-1)$$

$$M(n|n-1) = M(n-1|n-1)$$

$$\Rightarrow K(n) = \frac{M(n-1|n-1)}{\sigma^2 + M(n-1|n-1)}$$

$$\hat{A}(n|n) = \hat{A}(n-1|n-1) + K(n)(x(n) - \hat{A}(n-1|n-1))$$

$$M(n|n) = (1 - K(n)) M(n-1|n-1)$$

or changing the notation we have

$$K(n) = \frac{M(n-1)}{\sigma^2 + M(n-1)}$$

$$\hat{A}(n) = \hat{A}(n-1) + K(n)(x(n) - \hat{A}(n-1))$$

$$M(n) = (1 - K(n)) M(n-1)$$

These equations are just (12.34) - (12.36) with  
 obvious changes in notation.

Hence, from Section 12.6

$$\hat{A}(n) = \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/n+1} \cdot \frac{1}{n+1} \sum_{k=0}^n x(k)$$

$$M(n-1) = \frac{\sigma_A^2 \sigma^2}{n \sigma_A^2 + \sigma^2}$$

$$K(n) = \frac{\sigma_A^2}{(n+1) \sigma_A^2 + \sigma^2}$$

(1) From (13.39), (13.40), (13.42)

$$M(n|n-1) = 0.81 M(n-1|n-1) + 1$$

$$K(n) = \frac{M(n|n-1)}{\sigma_n^2 + M(n|n-1)}$$

$$M(n|n) = (1 - K(n)) M(n|n-1)$$

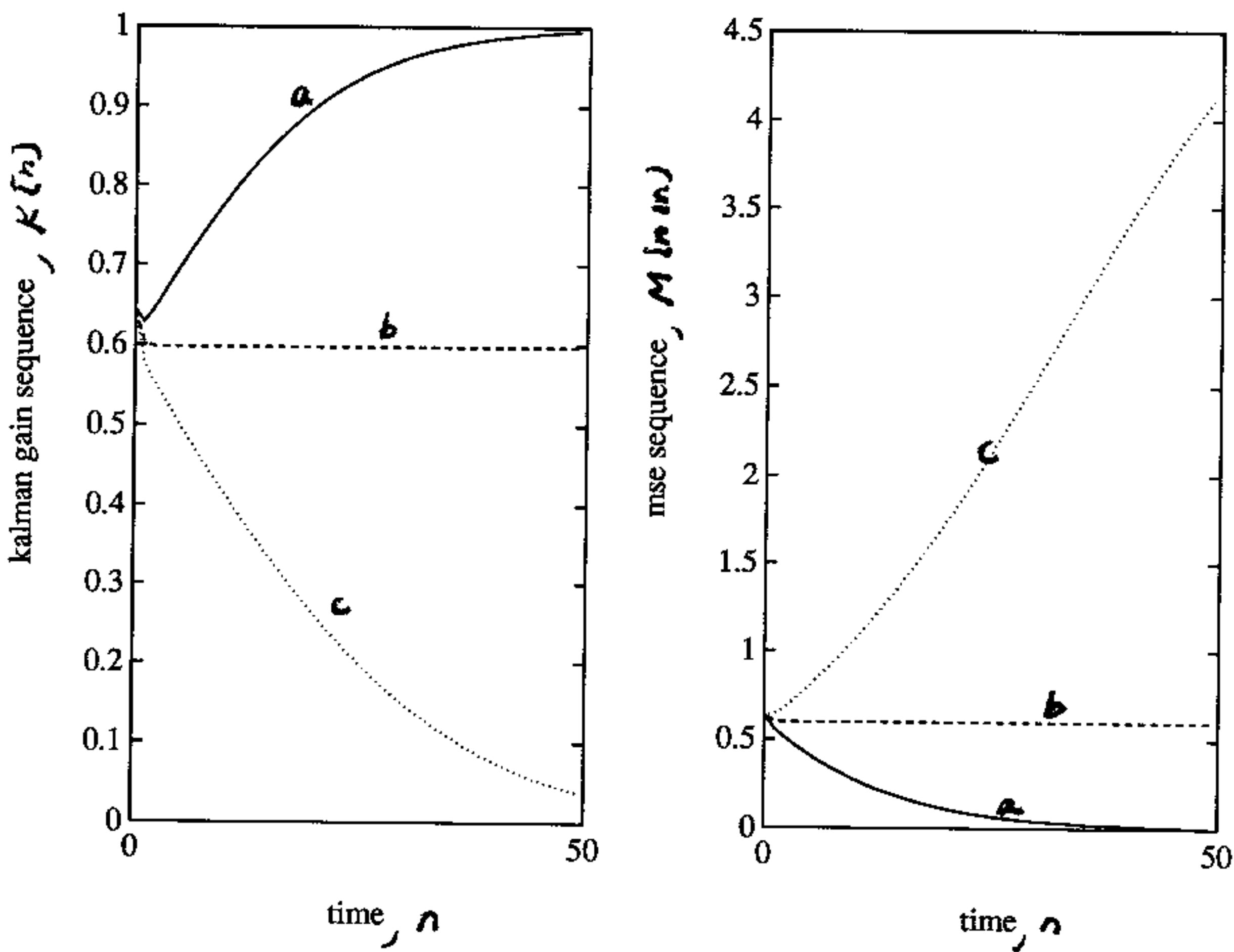
$$\text{where } M(-1|n-1) = \sigma_S^2 = 1$$

See next page for plots. For (a) the gain  $\rightarrow 1$  since the observations become less noisy and thus  $\hat{s}(n|m) \rightarrow x(n)$ . Also, the signal estimate improves with time since  $M(n|n) \rightarrow 0$ . For (c) the gain  $\rightarrow 0$  since the observations become noisier with time, and also  $\hat{s}(n|m) \rightarrow \hat{s}(n|n-1)$ .

$$= 0.9 \hat{s}(n-1|n-1)$$

The signal estimate will decay to zero so that the MMSE will be  $E[s^2(n)] = \sigma_u^2 / 1 - a^2 = 5.26$ , which is the steady-state variance of  $s(n)$ . For (b) the gain and MMSE sequences attain a steady state after a few

iterations. This is the steady-state Kalman filter or Wiener filter (see Section 13.5).

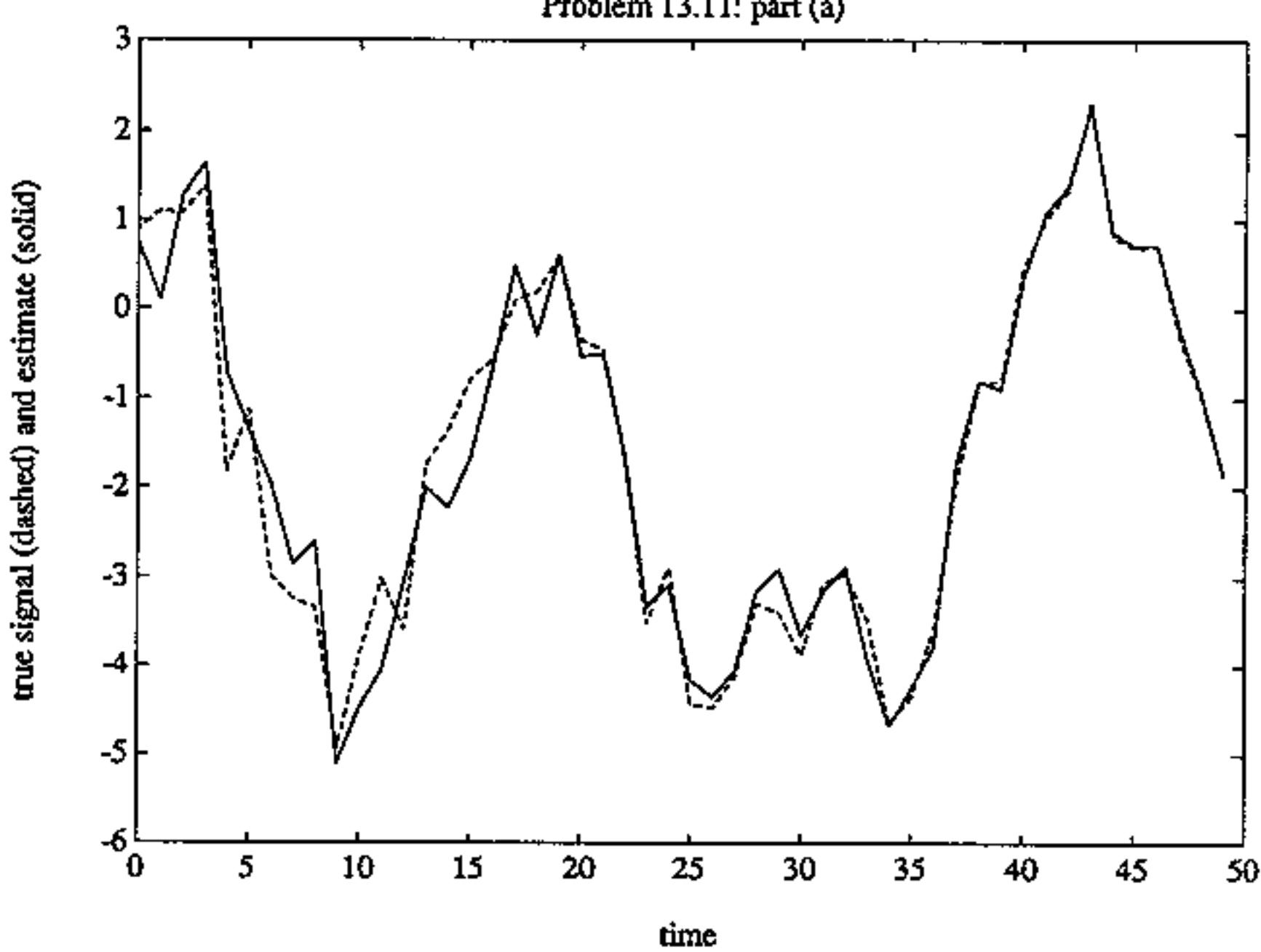


A Monte Carlo simulation produces the plots on the following page.

$$\begin{aligned}
 12) \quad & \text{If } \sigma_n^2 = 0, K(n) = 1 \text{ and } \hat{x}[n|n] = x[n] \\
 & \text{Also, } \hat{s}[n|n-1] = a \hat{s}[n-1|n-1] \\
 & \qquad \qquad \qquad = a \times x[n-1] = a s[n-1]
 \end{aligned}$$

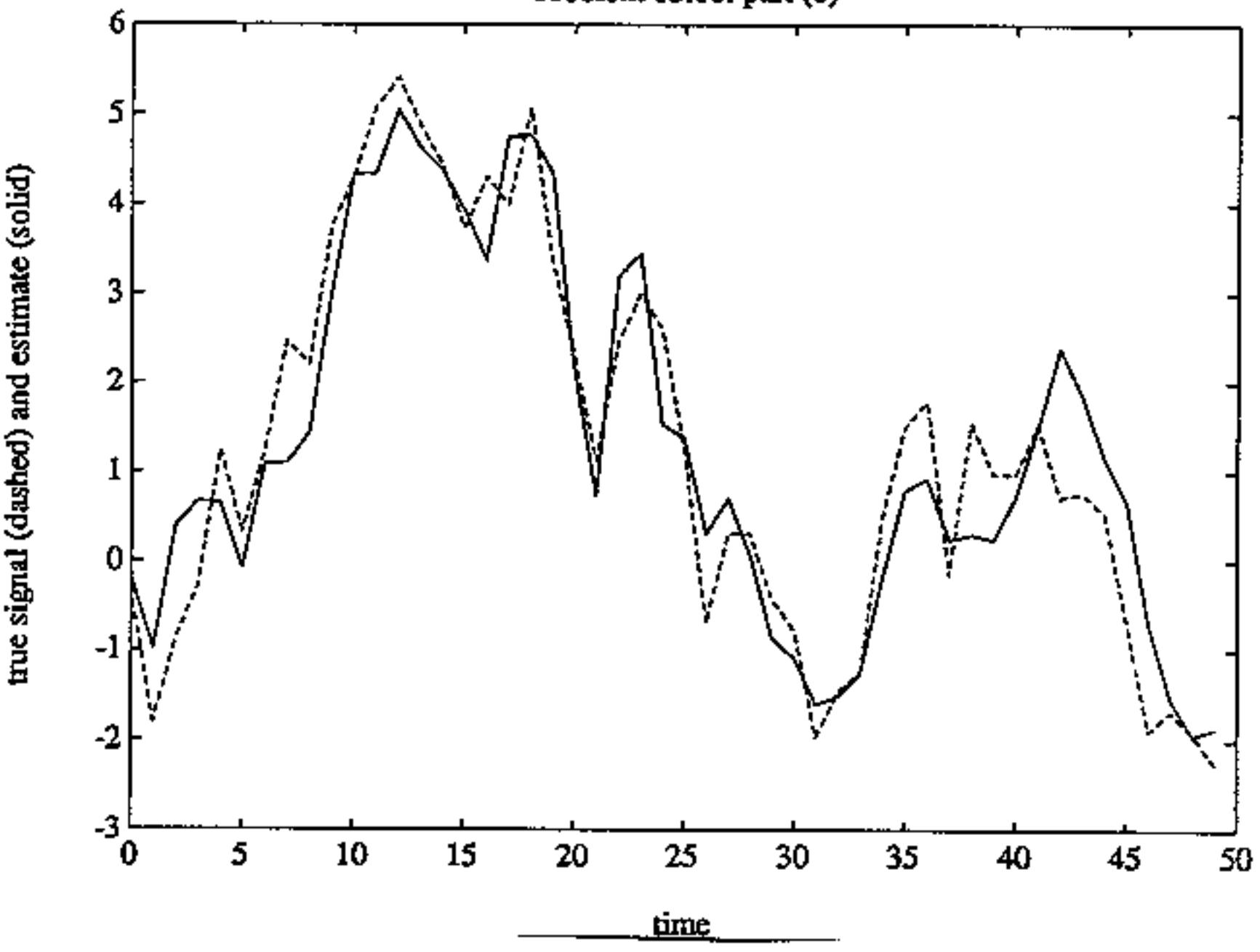
$$\text{Thus, } \hat{x}[n] = x[n] - \hat{s}[n|n-1]$$

Problem 13.11: part (a)

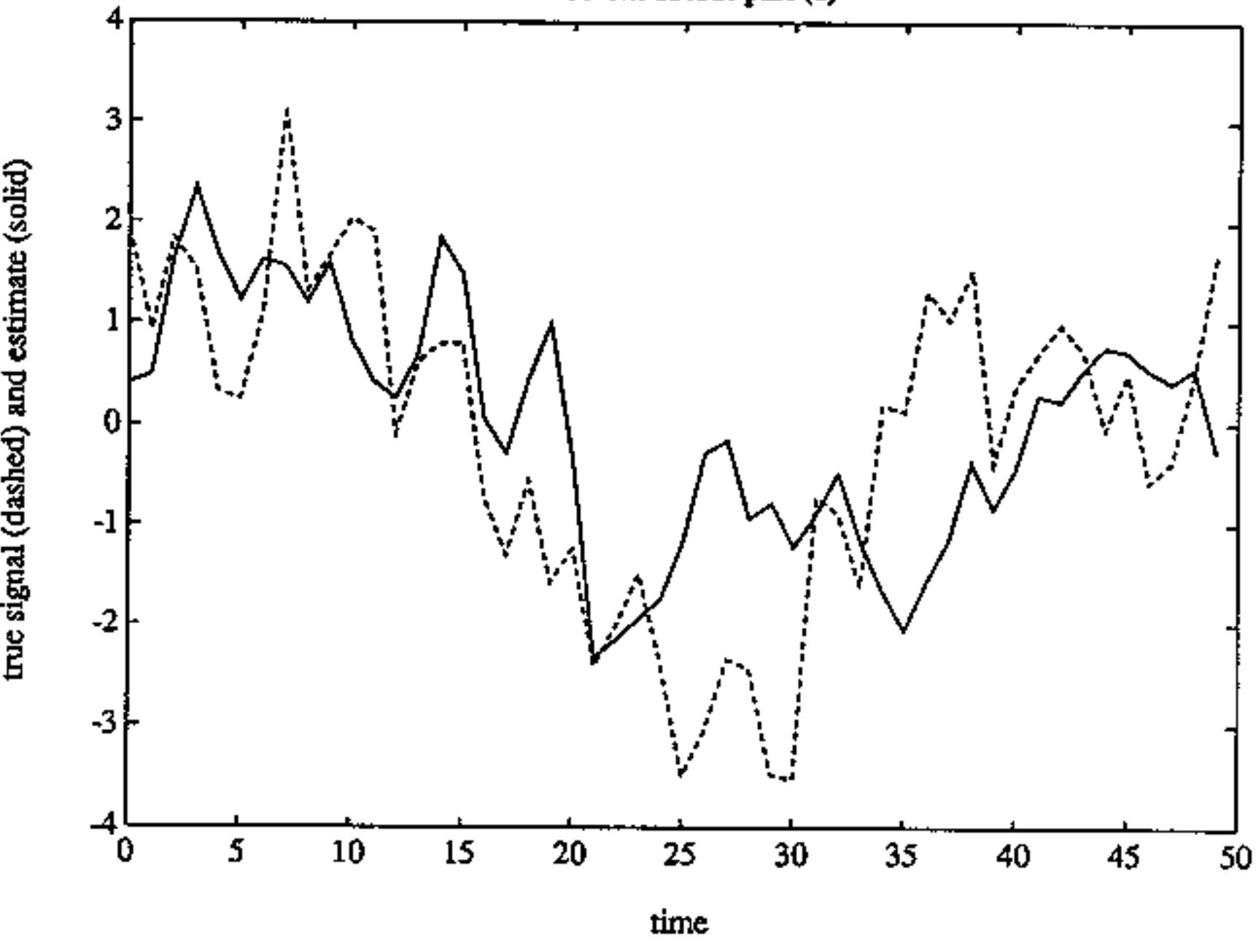


$s(n) = \text{DASHED}$   
 $\hat{s}(n) = \text{SOLID}$

Problem 13.11: part (b)



Problem 13.11: part (c)



$$= s(n) - a s(n-1) = u(n)$$

Yes.

- 13) Since  $E(s(n))$  is known, the MSE estimator of  $s'(n) = s(n) - E(s(n))$  is  $\hat{s}'(n) = \hat{s}(n) - E(s(n))$  and the minimum MSE is the same as for  $\hat{s}(n)$ . Thus,  $M(n|n)$  and  $M(n|n-1)$  do not change. Also, then  $K(n)$  does not change.

Prediction:  $\hat{s}'(n|n-1) = a \hat{s}'(n-1|n-1)$   
 $\hat{s}(n|n-1) - E(s(n)) = a (\hat{s}(n-1|n-1) - E(s(n-1)))$   
or  $\hat{s}(n|n-1) = a \hat{s}(n-1|n-1)$   
since  $E(s(n)) = a E(s(n-1))$

Correction:  $\hat{s}'(n|n) = \hat{s}'(n|n-1) + K(n)(x(n) - \hat{s}'(n|n-1))$   
 $\hat{s}(n|n) - E(s(n)) = \hat{s}(n|n-1) - E(s(n)) + K(n)(x(n) - E(x(n)) - \hat{s}(n|n-1) + E(s(n)))$

$$\text{But } E(x(n)) = E(s(n)) + E(w(n)) = E(s(n))$$

Thus, we have the same equations as before.  
The only difference arises in the initialization since  $u_0 \neq 0$ .

- 14) In steady state we have  $M(n|n) = M(\infty)$ ,  $M(n|n-1) = M_p(\infty)$  and from (13, 42)

$$M(\infty) = (1 - K(\infty)) M_p(\infty)$$

$$< M_p(\infty)$$

since  $K(\infty) < 1$ . Thus, for large  $n$   
 $M(n|n-1) > M(n-1|n-1)$ . This is  
reasonable since  $s(n)$  is harder to  
estimate than  $s(n-1)$  based on  $\{x(0), x(1),$   
 $\dots, x(n-1)\}$  due to the added variability  
of the  $u(n)$  noise term.

- 15) From (13.38)  $\hat{s}(n+1|n) = a \hat{s}(n|n)$   
Now if  $\sigma_{n+1}^2 \rightarrow \infty$ , the future measure-  
ments will be useless so that the  
corrected estimates will be predictions or  
will be based on only  $\{x(0), \dots, x(n)\}$ .

$$\Rightarrow \hat{s}(n+1|n+1) \rightarrow \hat{s}(n+1|n) = a \hat{s}(n|n)$$

$$\hat{s}(n+2|n+1) = a \hat{s}(n+1|n+1) = a^2 \hat{s}(n|n)$$

etc.

- 16) From (13.47)

$$H_{\infty}(z) = \frac{K(\infty)}{1 - a(1 - K(\infty)) z^{-1}}$$

- From (13.46)

$$M(\infty) = \frac{0.64 M(\infty) + 1}{0.64 M(\infty) + 2}$$

Let  $x = M(\infty)$

$$0.64x^2 + 2x = 0.64x + 1$$

$$x^2 + 2.125x - 1.5625 = 0$$

$$\Rightarrow x = -0.5781, -2.703$$

$$M(\infty) = -0.5781$$

$$M_p(\infty) = a = M(\infty) + \sigma_u^2$$

$$= 0.64M(\infty) + 1 = 1.37$$

$$K(\infty) = \frac{M_p(\infty)}{1 + M_p(\infty)} = 0.5781$$

$$H_{\infty}(z) = \frac{0.5781}{1 - 0.3375z^{-1}}$$

$$\hat{s}(n|n) = 0.3375 \hat{s}(n-1|n-1) + 0.5781 x(n)$$

17) See plot on next page.

18) Here  $h(n) = r^n$ . Using (13.51) - (13.54)

$$x(n) = h(n)A + w(n)$$

$$Q = 0$$

$$A = I$$

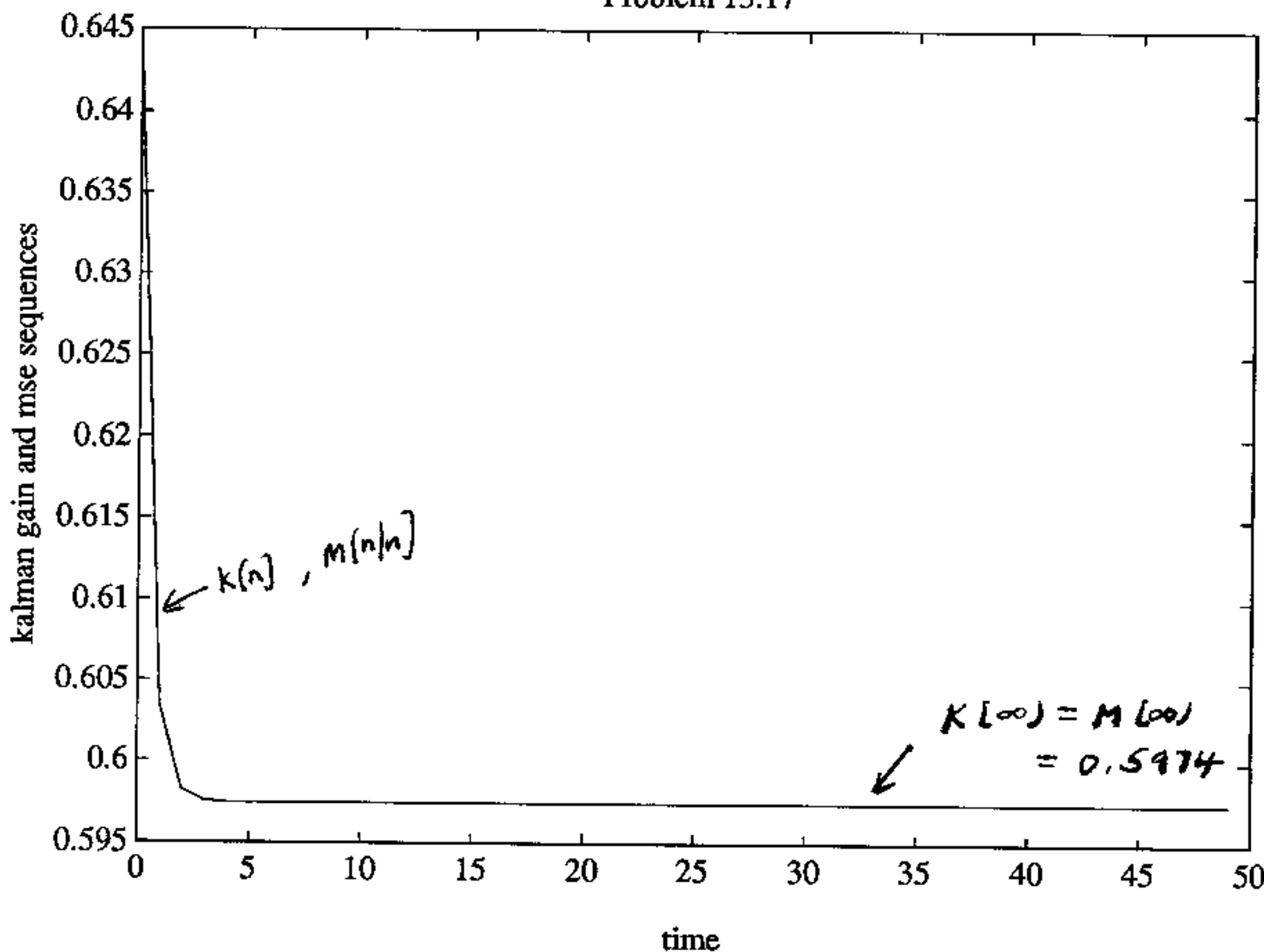
Can omit prediction stage.

$$K(n) = \frac{M(n|n-1)r^n}{\sigma^2 + r^{2n}M(n|n-1)}$$

$$M(n|n) = (1 - K(n)r^n)M(n|n-1)$$

$$\hat{A}(n|n) = \hat{A}(n|n-1) + K(n)(x(n) - r^n\hat{A}(n|n-1))$$

Problem 13.17



where  $\hat{A}(-1|1-1) = \underline{M}_A$ ,  $M(-1|1-1) = \sigma_A^2$ .

19)  $\underline{x}(n) = \underline{0}$ , from (13.60)

$$\begin{aligned} K(n) &= \underline{M}(n|n-1) \underline{H}^T(n) (\underline{H}(n) \underline{M}(n|n-1) \underline{H}^T(n))^{-1} \\ &= \underline{M}(n|n-1) \underline{H}^T(n) \underline{H}^{T-1}(n) \underline{M}^{-1}(n|n-1) \underline{H}^T(n) \\ &= \underline{H}^{-1}(n) \end{aligned}$$

$$\begin{aligned} \hat{x}(n|n) &= \hat{x}(n|n-1) + \underline{H}^{-1}(n) (\underline{x}(n) - \underline{H}(n) \hat{x}(n|n-1)) \\ &= \underline{H}^{-1}(n) \underline{x}(n) \end{aligned}$$

$\Rightarrow$  disregard all previous data since

$$\underline{x}(n) = \underline{H}(n) \hat{x}(n) \text{ and thus } \hat{x}(n) = \underline{H}^{-1}(n) \underline{x}(n)$$

$\Rightarrow$  no error in  $\hat{x}(n|n)$ .

If  $\underline{L}(n) \rightarrow \infty$ ,  $K(n) \rightarrow 0 \Rightarrow \hat{x}(n|n) \rightarrow \hat{x}(n|n-1)$   
or we ignore the data sample  $x(n)$ .

20) Same approach as in Problem 13.15.

21) Note that the state equation is linear  
but the observation equation is not.

From (13.67) - (13.71)

$$\hat{f}_o(n|n-1) = a \hat{f}_o(n-1|n-1)$$

$$M(n|n-1) = a^2 M(n-1|n-1) + \sigma_u^2$$

$$K(n) = \frac{M(n|n-1) H(n)}{\sigma^2 + H^2(n) M(n|n-1)}$$

$$\text{where } H(n) = \frac{\partial h}{\partial f_o(n)} \Big|_{f_o(n) = \hat{f}_o(n|n-1)}$$

$$\text{But } h(f_o(n)) = \cos 2\pi f_o(n)$$

$$H(n) = -2\pi \sin 2\pi \hat{f}_o(n|n-1)$$

$$\hat{f}_o(n|n) = \hat{f}_o(n|n-1) + K(n)(x(n) - \cos 2\pi \hat{f}_o(n|n-1))$$

$$M(n|n) = (1 - K(n) H(n)) M(n|n-1)$$

22)  $Nx(n) = Nx(n-1) + u_x(n)$

$$\Rightarrow Nx(n) = \sum_{k=0}^n u_x(k) + Nx(n-1)$$

$$\text{var}(Nx(n)) = \sum_{k=0}^n \text{var}(u_x(k)) + \text{var}(Nx(n-1))$$

Since  $u_x(n)$  is WGN and is independent  
of  $Nx(n-1)$

$$\text{var}(w_{xL_n}) = (n+1) \text{var}(u_{xL_n}) + \sigma_v^2 \\ = (n+1) \sigma_u^2 + \sigma_v^2$$

A better model would be

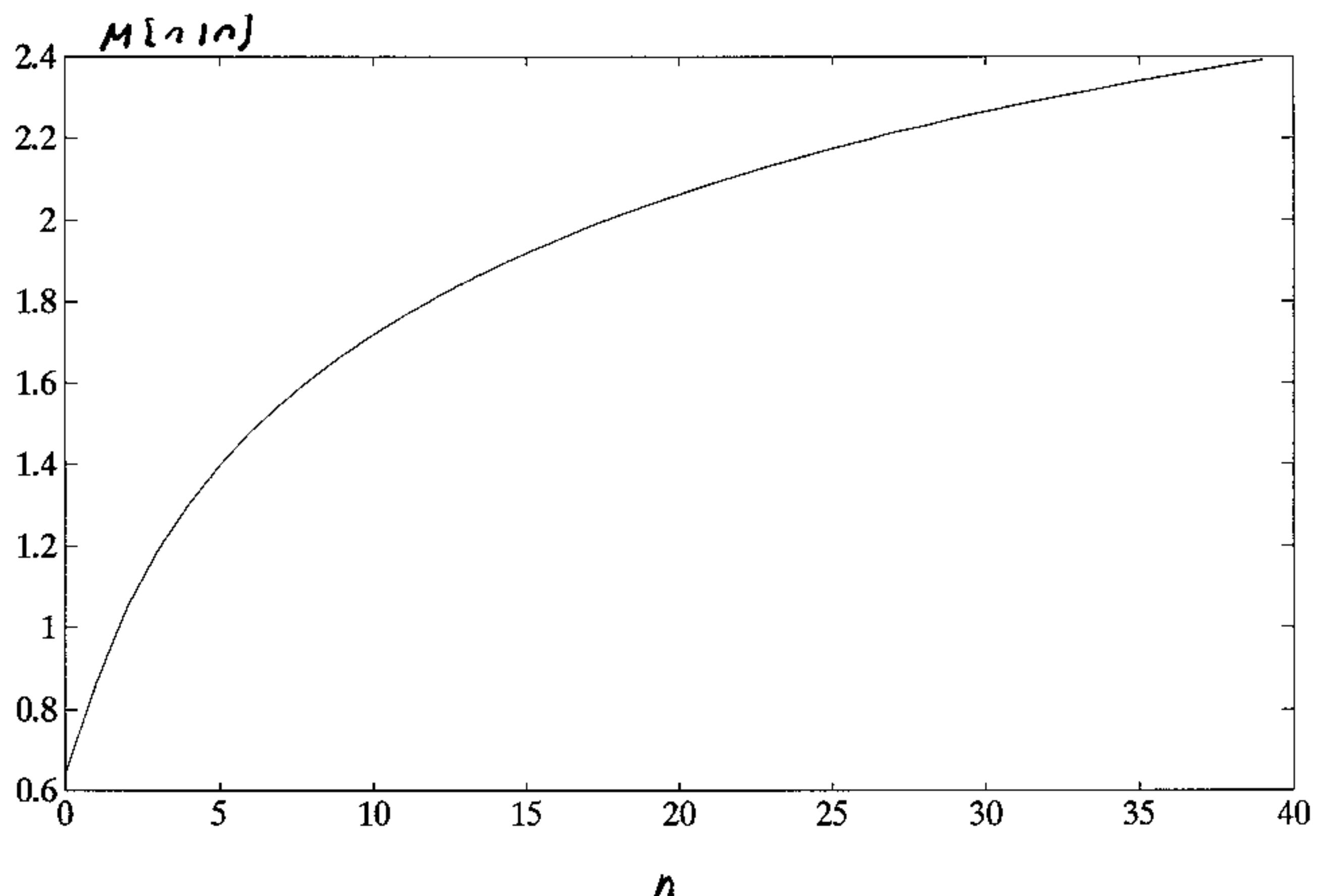
$$w_x(n) = a w_x(n-1) + u_{xn}$$

for  $0 < a < 1$  and  $a$  should be near one.  
Then, in steady-state the variance  
would be a constant.

$$\begin{aligned} 23) \quad M[n]_n &= (1 - K[n]) M[n]_{n-1} \\ &= \left( 1 - \frac{M[n]_{n-1}}{\sigma_n^2 + M[n]_{n-1}} \right) M[n]_{n-1} \\ &= \frac{\sigma_n^2 M[n]_{n-1}}{\sigma_n^2 + M[n]_{n-1}} \\ &= \frac{\sigma_n^2 (a^2 M[n-1]_{n-1} + \sigma_u^2)}{\sigma_n^2 + a^2 M[n-1]_{n-1} + \sigma_u^2} \\ &= \frac{(n+1)(0.81 M[n-1]_{n-1} + 1)}{n+1 + 0.81 M[n-1]_{n-1} + 1} \end{aligned}$$

where  $M[-1]_{-1} = 1$

See plot on next page. Since the observations  
are progressively noisier, the MHSE  
increases.



### Chapter 15

1)  $\text{Cov}(\tilde{x}_1, \tilde{x}_2) = E((\tilde{x}_1 - E(\tilde{x}_1))^* (\tilde{x}_2 - E(\tilde{x}_2)))$

The expectation is with respect to  $p(u_1, v_1, u_2, v_2)$

But  $p(u_1, v_1, u_2, v_2) = p(u_1, v_1) p(u_2, v_2)$

Thus,

$$\begin{aligned}\text{Cov}(\tilde{x}_1, \tilde{x}_2) &= E_{u_1, v_1} ((\tilde{x}_1 - E(\tilde{x}_1)) E_{u_2, v_2} (\tilde{x}_2 - E(\tilde{x}_2))) \\ &= 0\end{aligned}$$

2) Consider  $\tilde{y} = \underline{a}^H \tilde{x}$

$$E(|\tilde{y}|^2) = E(\underline{a}^H \tilde{x} \tilde{x}^H \underline{a}) = \underline{a}^H \underline{\Sigma} \underline{x} \underline{x}^H \underline{a} \geq 0$$

for all  $\underline{a}$ . It will be positive definite if and only if  $E(|\tilde{y}|^2) > 0$  or  $\tilde{y} \neq 0$  for any  $\underline{a}$ . Thus, if any random variable (element of  $\tilde{x}$ ) can be expressed as a linear combination of the others,  $\underline{\Sigma} \underline{x}$  will only be positive semidefinite.

3)  $\underline{\Sigma} \underline{x} = \frac{1}{2} \begin{bmatrix} A & -B \\ B & A \end{bmatrix}$  where  $A = \begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix}$   
 $B = \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix}$

Note that  $A^T = A$ ,  $B^T = -B$  as required.

$$\Rightarrow \underline{\Sigma} \underline{x} = \underline{A} + j \underline{B} = \begin{bmatrix} 4 & 2+2j \\ 2-2j & 4 \end{bmatrix}$$

The complex Gaussian vector  $\tilde{x} = \begin{bmatrix} u_1 + jv_1 \\ u_2 + jv_2 \end{bmatrix}$  will have the covariance matrix  $\underline{\Sigma} \underline{x}$ .

$$4) \text{ a) } \underline{C} \underline{x}^{-1} = \frac{1}{2} \begin{bmatrix} 2 & -1 & 0 & 1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 1 & 0 & -1 & 2 \end{bmatrix}$$

$$\underline{C} \tilde{\underline{x}}^{-1} = \frac{1}{4} \begin{bmatrix} 2 & -1-j \\ -1+j & 2 \end{bmatrix}$$

Let  $\underline{x} = [\underline{u}^T \underline{v}^T]^T$  where  $\underline{u} = (u_1, u_2)$ ,  $\underline{v} = (v_1, v_2)^T$

$$2 \underline{x}^T \underline{C} \tilde{\underline{x}}^{-1} \underline{x} = (\underline{u}^T \underline{v}^T) \begin{bmatrix} 2 & -1 & 0 & 1 \\ -1 & 2 & -1 & 0 \\ \hline 0 & -1 & 2 & -1 \\ 1 & 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} \underline{u} \\ \underline{v} \end{bmatrix}$$

$$= [\underline{u}^T \underline{v}^T] \begin{bmatrix} \underline{C} & \underline{D}^T \\ \underline{D} & \underline{C} \end{bmatrix} \begin{bmatrix} \underline{u} \\ \underline{v} \end{bmatrix}$$

$$= \underline{u}^T \underline{C} \underline{u} + \underline{v}^T \underline{C} \underline{v} + \underline{u}^T \underline{D}^T \underline{v} + \underline{v}^T \underline{D} \underline{u}$$

$$= \underline{u}^T \underline{C} \underline{u} + \underline{v}^T \underline{C} \underline{v} + \underline{v}^T \underline{D} \underline{u} + \underline{v}^T \underline{D} \underline{u}$$

$$= \underline{u}^T \underline{C} \underline{u} + \underline{v}^T \underline{D} \underline{v} + 2 \underline{v}^T \underline{D} \underline{u}$$

$$\begin{aligned} \tilde{\underline{x}}^H \underline{C} \tilde{\underline{x}}^T \tilde{\underline{x}} &= (\underline{u} - j \underline{v})^T \frac{1}{4} (\underline{C} + j \underline{D}) (\underline{u} + j \underline{v}) \\ &= \frac{1}{4} (\underline{u} - j \underline{v})^T (\underline{C} \underline{u} + j \underline{C} \underline{v} + j \underline{D} \underline{u} - \underline{D} \underline{v}) \\ &= \frac{1}{4} (\underline{u}^T \underline{C} \underline{u} + j \underline{u}^T \underline{C} \underline{v} + j \underline{v}^T \underline{D} \underline{u} - \underline{u}^T \underline{D} \underline{v} \\ &\quad - j \underline{v}^T \underline{C} \underline{u} + \underline{v}^T \underline{C} \underline{v} + \underline{v}^T \underline{D} \underline{u} + j \underline{v}^T \underline{D} \underline{v}) \end{aligned}$$

$$\text{But } \underline{u}^T \underline{D} \underline{u} = (\underline{u}^T \underline{D} \underline{u})^T = \underline{u}^T \underline{D}^T \underline{u} = -\underline{u}^T \underline{D} \underline{u} = 0$$

since  $\underline{D}^T = -\underline{D}$  and similarly for  $\underline{v}^T \underline{D} \underline{v} = 0$

$$\text{Also, } \underline{v}^T \underline{C} \underline{u} = \underline{u}^T \underline{C} \underline{v} \text{ since } \underline{C}^T = \underline{C}$$

$$\begin{aligned}
 \tilde{x}^H \tilde{x}' \tilde{x} &= 1/4 (\underline{u}^T \underline{u} - \underline{u}^T D \underline{v} + \underline{v}^T \underline{v} + \underline{v}^T D \underline{u}) \\
 &= 1/4 (\underline{u}^T \underline{u} - \underline{v}^T D^T \underline{u} + \underline{v}^T \underline{v} + \underline{v}^T D \underline{u}) \\
 &= 1/4 \underbrace{(\underline{u}^T \underline{u} + \underline{v}^T \underline{v} + 2 \underline{v}^T D \underline{u})}_{2 \underline{x}^T \underline{x}' \underline{x}} \\
 &= \frac{1}{2} \underline{x}^T \underline{x}' \underline{x}
 \end{aligned}$$

b)  $\det(C_x) = 4$

$$\begin{aligned}
 \det(C_x) &= \det \begin{pmatrix} 4 & 2+2j \\ 2-2j & 4 \end{pmatrix} \\
 &= 16 - |2+2j|^2 = 16 - 8 = 8
 \end{aligned}$$

$$\frac{\det^2(C_x)}{16} = \frac{64}{16} = 4$$

5)  $m_1 = \begin{bmatrix} a-b \\ b-a \end{bmatrix} \quad m_2 = \begin{bmatrix} c-d \\ d-c \end{bmatrix}$   
 $m_1 \rightarrow a+jb \quad m_2 \rightarrow c+jd$

a)  $\alpha(a+jb) = \alpha a + j\alpha b \rightarrow \begin{bmatrix} \alpha a - \alpha b \\ \alpha b \alpha a \end{bmatrix} = \alpha m_1$

b)  $(a+jb) + (c+jd) = (a+c) + j(b+d)$   
 $\rightarrow \begin{bmatrix} a+c & -(b+d) \\ b+d & a+c \end{bmatrix} = m_1 + m_2$

c)  $(a+jb)(c+jd) = (ac-bd) + j(ad+bc)$   
 $\rightarrow \begin{bmatrix} ac-bd & -(ad+bc) \\ ad+bc & ac-bd \end{bmatrix} = m_1 m_2$

6) Transform  $2 \times 2$  blocks or let

$$\underline{A} = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} \rightarrow \begin{bmatrix} a_1 + jb_1 \\ a_2 + jb_2 \\ a_3 + jb_3 \end{bmatrix}$$

$$\text{But } \underline{A}^T \underline{A} = m_1^T m_1 + m_2^T m_2 + m_3^T m_3$$

(Note that if  $m \rightarrow a+jb$ ,  $m^T \rightarrow a-jb$ )

$$\begin{aligned} \Rightarrow \underline{A}^T \underline{A} &\rightarrow (a_1 - jb_1)(a_1 + jb_1) \\ &+ (a_2 - jb_2)(a_2 + jb_2) \\ &+ (a_3 - jb_3)(a_3 + jb_3) \\ &= a_1^2 + b_1^2 + a_2^2 + b_2^2 + a_3^2 + b_3^2 + j^0 \end{aligned}$$

Transforming back we have

$$\underline{A}^T \underline{A} = \begin{bmatrix} a_1^2 + b_1^2 + a_2^2 + b_2^2 + a_3^2 + b_3^2 & 0 \\ 0 & \xrightarrow{\quad} \end{bmatrix}$$

7) Since  $E(\tilde{x}^{*3}) = E(\tilde{x}^3)^*$

$$E(\tilde{x}^* \tilde{x}^2) = E(\tilde{x} \tilde{x}^{*2})^*$$

we need only show that  $E(\tilde{x}^3) = 0$ ,  $E(\tilde{x}^* \tilde{x}^2) = 0$ .

$$\text{But } \frac{\partial^3 \phi_{\tilde{x}}(\tilde{w})}{\partial \tilde{w}^{*3}} = (\mathcal{J}_{12})^3 E(\tilde{x}^3)$$

(see development of (15B.3) in App 15B)

$$\phi_{\tilde{x}}(\tilde{\omega}) = e^{-\frac{1}{4}\sigma^2 \tilde{\omega} \tilde{\omega}^*}$$

$$\frac{\partial \phi}{\partial \tilde{\omega}^*} = e^{-\frac{1}{4}\sigma^2 |\tilde{\omega}|^2} (-\frac{1}{4}\sigma^2 \tilde{\omega})$$

$$\frac{\partial^2 \phi}{\partial \tilde{\omega}^{*2}} = e^{-\frac{1}{4}\sigma^2 |\tilde{\omega}|^2} (-\frac{1}{4}\sigma^2 \tilde{\omega})^2$$

$$\frac{\partial^3 \phi}{\partial \tilde{\omega}^{*3}} = e^{-\frac{1}{4}\sigma^2 |\tilde{\omega}|^2} (-\frac{1}{4}\sigma^2 \tilde{\omega})^3$$

$$\Rightarrow \frac{\partial^3 \phi}{\partial \tilde{\omega}^{*3}} \Big|_{\tilde{\omega}=0} = 0 \Rightarrow E(\tilde{x}^3) = 0$$

Similarly,

$$E(\tilde{x}^* \tilde{x}^2) = (\mathcal{J}_2)^3 \frac{\partial^3 \phi_{\tilde{x}}(\tilde{\omega})}{\partial \tilde{\omega} \partial \tilde{\omega}^{*2}} \Big|_{\tilde{\omega}=0}$$

$$\frac{\partial^2 \phi}{\partial \tilde{\omega}^{*2}} = e^{-\frac{1}{4}\sigma^2 \tilde{\omega} \tilde{\omega}^*} (-\frac{1}{4}\sigma^2 \tilde{\omega})^2$$

$$\begin{aligned} \frac{\partial^3 \phi}{\partial \tilde{\omega} \partial \tilde{\omega}^{*2}} &= e^{-\frac{1}{4}\sigma^2 |\tilde{\omega}|^2} \left( \frac{1}{16} \sigma^4 \tilde{\omega} \right) \\ &\quad + (-\frac{1}{4}\sigma^2 \tilde{\omega})^2 e^{-\frac{1}{4}\sigma^2 |\tilde{\omega}|^2} (-\frac{1}{4}\sigma^2 \tilde{\omega}^*) \end{aligned}$$

which when evaluated at  $\tilde{\omega} = 0$  is zero.

8)  $E(\tilde{x}^2) = 0$

$$\Rightarrow E((u+jv)(u+jv)) = 0$$

$$E(u^2) - E(v^2) + j E(uv) + j E(vu) = 0$$

$$\Rightarrow E(u^2) = E(v^2), E(uv) = 0$$

$$\text{or } \text{var}(u) = \text{var}(v), \text{cov}(u, v) = 0$$

This is the usual complex Gaussian random variable assumption.

$$\begin{aligned}
 9) \quad & E((\underline{\tilde{x}} - \underline{\mu}) (\underline{\tilde{x}} - \underline{\mu})^T) = \\
 & E((\underline{u} - \underline{\mu}_u) + j(\underline{v} - \underline{\mu}_v)) (\underline{u} - \underline{\mu}_u)^T \\
 = & E((\underline{u} - \underline{\mu}_u)(\underline{u} - \underline{\mu}_u)^T) + E((\underline{v} - \underline{\mu}_v)(\underline{v} - \underline{\mu}_v)^T) \\
 & + j(E((\underline{u} - \underline{\mu}_u)(\underline{v} - \underline{\mu}_v)^T) + E((\underline{v} - \underline{\mu}_v)(\underline{u} - \underline{\mu}_u)^T)) \\
 = & 0
 \end{aligned}$$

$$\Rightarrow \underline{\Sigma}_{uu} = \underline{\Sigma}_{vv}, \quad \underline{\Sigma}_{uv} = -\underline{\Sigma}_{vu} \\
 = A/2 \quad \quad \quad = -B/2$$

$$\text{or } \underline{\Sigma}_x = E\left(\left(\begin{bmatrix} \underline{u} \\ \underline{v} \end{bmatrix} - E\left(\begin{bmatrix} \underline{u} \\ \underline{v} \end{bmatrix}\right)\right)\left(\begin{bmatrix} \underline{u} \\ \underline{v} \end{bmatrix} - E\left(\begin{bmatrix} \underline{u} \\ \underline{v} \end{bmatrix}\right)\right)^T\right) \\
 = \begin{bmatrix} A/2 & -B/2 \\ B/2 & A/2 \end{bmatrix}$$

$$\begin{aligned}
 10) \quad E(\hat{\sigma}^2) &= E(\underline{\tilde{x}}^H \underline{A} \underline{\tilde{x}}) = \text{tr}(\underline{A} \underline{\sigma}^2 \underline{Q}) \quad (\text{see (15.29)}) \\
 &= \sigma^2 \text{tr}(\underline{A} \underline{B}) \\
 \Rightarrow \text{tr}(\underline{A} \underline{B}) &= 1
 \end{aligned}$$

$$\text{var}(\hat{\sigma}^2) = \text{tr}(\underline{A} \underline{\sigma}^2 \underline{B} \underline{A} \underline{\sigma}^2 \underline{B}) \quad (\text{see (15.30)}) \\
 = \sigma^4 \text{tr}((\underline{A} \underline{B})^2)$$

$$\begin{aligned}
 \text{But } \text{tr}(\underline{A} \underline{B}) &= \sum_{i=1}^N \lambda_i = 1 \\
 \text{tr}((\underline{A} \underline{B})^2) &= \sum_{i=1}^N \lambda_i^2
 \end{aligned}$$

Using the constraint we have to minimize

$$\text{tr}((\underline{A}\underline{B})^2) = \sum_{i=2}^n \lambda_i^2 + \left(1 - \sum_{i=2}^n \lambda_i\right)^2$$

$$\frac{\partial \text{tr}((\underline{A}\underline{B})^2)}{\partial \lambda_k} = 2\lambda_k + 2\left(1 - \sum_{i=2}^n \lambda_i\right)(-1) = 0$$

$$\Rightarrow \lambda_k = 1 - \sum_{i=2}^n \lambda_i \quad k = 2, 3, \dots, N$$

Since  $\sum_{i=2}^n \lambda_i$  is a constant,  $\lambda_k$  must be the same for all  $k$ . Thus,

$$\lambda_k = \frac{1}{N} \text{ for } \text{tr}(\underline{A}\underline{B}) = 1.$$

Since

$$\underline{Y}^T \underline{A} \underline{B} \underline{Y} = \underline{\Lambda} \quad \underline{Y} = \text{modal matrix}$$

$$= \frac{1}{N} \underline{I} \quad \underline{\Lambda} = \text{eigenvalue matrix}$$

$$\Rightarrow \underline{A}\underline{B} = \frac{1}{N} \underline{Y}^{-1} \underline{Y}^{-1} = \frac{1}{N} \underline{I}$$

$$\Rightarrow \underline{A} = \frac{1}{N} \underline{B}^{-1} \text{ or } \hat{\sigma}^2 = \frac{1}{N} \tilde{\underline{x}}^H \underline{B}^{-1} \tilde{\underline{x}}$$

$$\text{if } \underline{B} = \underline{I}, \quad \hat{\sigma}^2 = \frac{1}{N} \tilde{\underline{x}}^H \tilde{\underline{x}}.$$

$$(1) \quad \sum_{n=0}^{N-1} |\tilde{\underline{x}}(n)|^2 = \tilde{\underline{x}}^H \tilde{\underline{x}}$$

$$E(\tilde{\underline{x}}^H \tilde{\underline{x}}) = N \sigma^2$$

$$\begin{aligned} \text{var}(\tilde{\underline{x}}^H \tilde{\underline{x}}) &= \text{tr}(\underline{C}_{\tilde{\underline{x}}} \underline{C}_{\tilde{\underline{x}}}) \quad (\text{see (15.30)}) \\ &= \text{tr}(\sigma^4 \underline{I}) \\ &= N \sigma^4 \end{aligned}$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{N} \bar{x}^H \bar{x}$$

where  $\text{var}(\hat{\sigma}^2) = \sigma^4/N$

For real WGN,  $\hat{\sigma}^2 = \frac{1}{N} \bar{x}^T \bar{x}$  and  
 $\text{var}(\hat{\sigma}^2) = 2\sigma^4/N$ . Now, since  $\bar{x}(n) \sim \mathcal{CN}(0, \sigma^2)$ ,  
 $\bar{x}(n) = u(n) + jv(n)$ , where  $u(n) \sim \mathcal{N}(0, \sigma^2/2)$ ,  
 $v(n) \sim \mathcal{N}(0, \sigma^2/2)$ . Hence, for the complex  
Case  $\hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} (u^2(n) + v^2(n))$  and we  
are averaging  $2N$  independent samples as  
opposed to only  $N$  for the real case. Thus,  
 $\text{var}(\hat{\sigma}^2)/\text{complex} = \frac{1}{2} \text{var}(\hat{\sigma}^2)/\text{real}$ .

$$(2) v(n) = \sum_{h=-\infty}^{\infty} h(h) u(n-h)$$

$$E(v(n)) = \sum_h h(h) E(u(n-h)) = 0$$

$$E(v(n)v(n+m)) = E \left[ \sum_h h(h) u(n-h) \cdot \sum_l h(l) u(n+m-l) \right]$$

$$= \sum_k \sum_l h(k) h(l) \underbrace{E(u(n-k) u(n+m-l))}_{R_{uu}(m+k-l)}$$

doesn't depend on  $n \Rightarrow$   $R_{uu}$

Also, since  $u(n)$  is Gaussian and  $v(n)$   
is the result of a linear transformation,  
 $v(n)$  is Gaussian ( $u, v$  are jointly  
Gaussian)

To verify the ACF relationships

$$\begin{aligned} P_{vv}(f) &= |H(f)|^2 P_{uu}(f) \\ &= 1 \cdot P_{uu}(f) \\ \Rightarrow r_{vu}[k] &= r_{vv}[k] \end{aligned}$$

$$\begin{aligned} \text{also, } P_{uv}(f) &= H(f) P_{uu}(f) \\ &= -j P_{uu}(f) \quad f \geq 0 \\ &\quad j P_{uu}(f) \quad f < 0 \end{aligned}$$

$$\begin{aligned} \text{But } P_{vu}(f) &= P_{uv}^*(f), \\ \Rightarrow P_{vu}(f) &= \begin{cases} j P_{uu}(f) & f \geq 0 \\ -j P_{uu}(f) & f < 0 \end{cases} \\ &= -P_{uv}(f) \end{aligned}$$

and thus,  $r_{uv}[k] = -r_{vu}[k]$ . The PSD of  $\tilde{x}(n)$  is from (15.3).

$$\begin{aligned} \tilde{P}_{xx}(f) &= 2(P_{uu}(f) + j P_{uv}(f)) \\ &= 2(P_{uu}(f) + j(-j P_{uu}(f))) \quad f \geq 0 \\ &\quad 2(P_{uu}(f) + j(j P_{uu}(f))) \quad f < 0 \\ &= 4P_{uu}(f) \quad f \geq 0 \\ &\quad 0 \quad f < 0 \end{aligned}$$

$$(3) \quad \frac{\partial \theta^*}{\partial \theta} = \frac{1}{2} \left( \frac{\partial}{\partial \alpha} - j \frac{\partial}{\partial \beta} \right) (\alpha - j \beta)$$

$$= \frac{1}{2} \left( \frac{\partial \alpha}{\partial \alpha} - j \frac{\partial \beta}{\partial \alpha} - j \cdot j \frac{\partial \alpha}{\partial \beta} - \frac{\partial \beta}{\partial \beta} \right)$$

$$= \frac{1}{2} (1 - 0 - 0 - 1) = 0$$

$$18) \quad \frac{\partial}{\partial \theta} = \frac{\partial}{\partial \alpha} + j \frac{\partial}{\partial \beta}$$

$$\frac{\partial \theta}{\partial \theta} = \left( \frac{\partial}{\partial \alpha} + j \frac{\partial}{\partial \beta} \right) (\alpha + j\beta)$$

$$= 1 - 1 = 0$$

$$\frac{\partial \theta^*}{\partial \theta} = \left( \frac{\partial}{\partial \alpha} + j \frac{\partial}{\partial \beta} \right) (\alpha - j\beta)$$

$$= 1 + 1 = 2$$

$$14) \quad \frac{\partial \underline{\theta}^H b}{\partial \theta_k} = \frac{\partial}{\partial \theta_k} \sum_L b_L \theta_L^*$$

$$= \sum_L b_L \frac{\partial \theta_L^*}{\partial \theta_k} = 0 \text{ all } k$$

$$\Rightarrow \frac{\partial \underline{\theta}^H b}{\partial \theta} = 0$$

15) Same proof as in deriving (15.46)

$$\frac{\partial \underline{\theta}^H A \underline{\theta}}{\partial \theta} = A^T \underline{\theta}^*$$

Since  $A^H \neq A$ ,  $\frac{\partial \underline{\theta}^H A \underline{\theta}}{\partial \theta} = (A^H \underline{\theta})^*$

16) To find LSE

$$J = \sum_{n=0}^{N-1} \|\tilde{x}(n) - \tilde{A} s^n\|^2$$

From Example 15.2

$$\hat{A} = \frac{\sum_{n=0}^{N-1} \tilde{x}(n) \tilde{s}^*(n)}{\sum_{n=0}^{N-1} |\tilde{s}(n)|^2}$$

$$= \frac{\sum_{n=0}^{N-1} \tilde{x}(n) s^{*n}}{\sum_{n=0}^{N-1} |s|^n}$$

MLE will be the same.

$$E(\hat{A}) = \frac{\sum_n E(\tilde{x}(n)) s^{*n}}{\sum_n |s|^n}$$

$$= \frac{\tilde{A} \sum_n s^n s^{*n}}{\sum_n |s|^{2n}} = \tilde{A}$$

$$\text{var}(\hat{A}) = \frac{\sum_n \text{var}(\tilde{x}(n)) |s|^n}{(\sum_n |s|^n)^2}$$

$$= \frac{\sigma^2}{\sum_n |s|^{2n}}$$

$$\text{If } |s| < 1, \quad \text{var}(\hat{A}) \rightarrow \frac{\sigma^2}{(1 - |s|^2)^{-1}}$$

$$= \sigma^2 (1 - |s|^2)$$

$$\text{If } |s| \geq 1, \quad \text{var}(\hat{A}) \rightarrow 0$$

17) Equivalently, we have

$$\tilde{x}(n) = A + \tilde{w}(n) \text{ or}$$

$$u(n) = A + w_I(n)$$

$$v(n) = w_R(n)$$

Now let  $\underline{x} = \begin{pmatrix} u \\ v \end{pmatrix}$  and  $\underline{w} = \begin{pmatrix} w_R \\ w_I \end{pmatrix}$  so that

$$\underline{x} = A \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} + \underline{w}$$

$$\text{where } E(\underline{w}) = \underline{0}, \quad \underline{\Sigma}_w = \sigma^2 I_2 \quad \underline{\Sigma} = \underline{\Sigma}_x$$

The real PLVE can now be used.

$$\begin{aligned} \hat{A} &= \frac{[\underline{1}^T \underline{0}^T] \underline{\Sigma}^{-1} \underline{x}}{[\underline{1}^T \underline{0}^T] \underline{\Sigma}^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix}} \\ &= \frac{1}{N} \sum_{n=0}^{N-1} u(n) = \frac{1}{N} \sum_{n=0}^{N-1} \operatorname{Re}(\tilde{x}(n)) \end{aligned}$$

In Example 15.7  $\tilde{A}$  is complex and if  $\underline{\Sigma} = \sigma^2 \underline{I}$ , which conforms to the assumptions for this problem,

$$\hat{A} = \frac{\underline{1}^T \tilde{\underline{x}}}{\underline{1}^T \underline{1}} = \frac{1}{N} \sum_{n=0}^{N-1} \tilde{x}(n)$$

18) From (15.52)

$$I(\theta) = 2 \operatorname{Re} \left( \frac{\partial \tilde{u}^H(\theta)}{\partial \theta} \underline{\Sigma}^{-1} \underline{\tilde{x}}' \frac{\partial \tilde{u}(\theta)}{\partial \theta} \right)$$

$$= \frac{2}{\sigma^2} \operatorname{Re} \left[ \sum_{n=0}^{N-1} \left| \frac{\partial \tilde{u}^H(\theta)}{\partial \theta} \right|_n \right]^2$$

$$\text{var}(\hat{\theta}) \geq \frac{\sigma^2}{\sum_{n=0}^{N-1} \left( \frac{\partial \tilde{s}(n; \theta)}{\partial \theta} \right)^2 + \left( \frac{\partial s(n; \theta)}{\partial \theta} \right)^2}$$

For real data (see (3.14))

$$\text{var}(\hat{\theta}) \geq \frac{\sigma^2}{\sum_{n=0}^{N-1} \left( \frac{\partial s(n; \theta)}{\partial \theta} \right)^2}$$

In complex case there is information in both the real and imaginary parts of the signal. Also, the  $\sigma^2/2$  factor accounts for the variance of each real noise sample.

19) From (15.52) with  $\underline{\Sigma} = \sigma^2 \underline{\mathbb{I}}$

$$\begin{aligned} I(\sigma^2) &= \text{tr} \left[ \underline{\Sigma}^{-1} \frac{\partial \underline{\Sigma}}{\partial \sigma^2} \underline{\Sigma}^{-1} \frac{\partial \underline{\Sigma}}{\partial \sigma^2} \right] \\ &= \text{tr} \left[ \left( \frac{1}{\sigma^2} \underline{\mathbb{I}} \right) (\underline{\mathbb{I}}) \left( \frac{1}{\sigma^2} \underline{\mathbb{I}} \right) (\underline{\mathbb{I}}) \right] \\ &= N/\sigma^4 \end{aligned}$$

$$\text{var}(\hat{\sigma}^2) \geq \sigma^4/N$$

$$P(\tilde{\underline{x}}; \sigma^2) = \frac{1}{\pi^N \det(\sigma^2 \underline{\mathbb{I}})} e^{-\frac{1}{\sigma^2} \tilde{\underline{x}}^H \tilde{\underline{x}}}$$

$$\begin{aligned}
 \frac{\partial \ln p}{\partial \sigma^2} &= -\frac{\partial \ln \det(\sigma^2 \mathbf{I})}{\partial \sigma^2} - \frac{\partial}{\partial \sigma^2} \left( \frac{1}{\sigma^2} \underline{\mathbf{x}}^H \underline{\mathbf{x}} \right) \\
 &= -\frac{\partial \ln \sigma^{2N}}{\partial \sigma^2} - \frac{\partial}{\partial \sigma^2} \left( \frac{1}{\sigma^2} \underline{\mathbf{x}}^H \underline{\mathbf{x}} \right) \\
 &= -N/\sigma^2 + \frac{1}{\sigma^4} \underline{\mathbf{x}}^H \underline{\mathbf{x}} \\
 &= N/\sigma^4 \underbrace{\left( \frac{1}{N} \sum_{n=0}^{N-1} \underline{\mathbf{x}}^H (\underline{\mathbf{x}}^H - \sigma^2) \right)}_{\text{effcient estimator}}
 \end{aligned}$$

20)  $\hat{\theta}_i = \underline{\alpha}_i^H \underline{\tilde{\mathbf{x}}} \Rightarrow \hat{\theta} = \underline{A} \underline{\tilde{\mathbf{x}}} \quad \underline{A} = \begin{bmatrix} \underline{\alpha}_1^H \\ \vdots \\ \underline{\alpha}_p^H \end{bmatrix}$

$$\begin{aligned}
 E(\hat{\theta}) &= \underline{\theta} \Rightarrow E(A(\underline{\theta} + \underline{\tilde{w}})) \\
 &= A \underline{\theta} E(\underline{\theta}) = \underline{\theta}
 \end{aligned}$$

or  $A^H A = \underline{\mathbf{I}}$

$$\begin{aligned}
 \text{var}(\hat{\theta}_i) &= E((\theta_i - \hat{\theta}_i)^2) \\
 &= E((\underline{\alpha}_i^H \underline{\tilde{\mathbf{x}}} - \underline{\alpha}_i^H E(\underline{\mathbf{x}}))^2) \\
 &= E(\underline{\alpha}_i^H (\underline{\mathbf{x}} - E(\underline{\mathbf{x}})) (\underline{\mathbf{x}} - E(\underline{\mathbf{x}}))^H \underline{\alpha}_i) \\
 &= \underline{\alpha}_i^H \underline{\alpha}_i
 \end{aligned}$$

$$A^H A = \underline{\mathbf{I}} \Rightarrow \begin{bmatrix} \underline{\alpha}_1^H \\ \vdots \\ \underline{\alpha}_p^H \end{bmatrix}^H \underline{\alpha}_i = \underline{\mathbf{I}}$$

or  $A^H (\underline{\alpha}_1, \dots, \underline{\alpha}_p) = \underline{\mathbf{I}}$

$$\underline{H}^H \underline{\alpha}_i = \underline{e}_i \quad \text{where } \underline{e}_i = (0 \dots 0, 1, 0 \dots 0)^T$$

↑  $i^{\text{th}}$  place

Using (15.51) we have

$$\underline{\alpha}_{i,\text{opt}} = \underline{\mathbf{C}}^{-1} \underline{H} (\underline{H}^H \underline{\mathbf{C}}^{-1} \underline{H})^{-1} \underline{e}_i$$

$$\hat{\theta}_i = \underline{a}_{i, \text{opt}} \tilde{x} = \underline{e}_i^H (\underline{H}\underline{C}^{-1}\underline{H})^{-1} \underline{H}^H \underline{C}^{-1} \tilde{x}$$

$$\Rightarrow \hat{\underline{\theta}} = (\underline{H}\underline{C}\underline{H})^{-1} \underline{H}^H \underline{C}^{-1} \tilde{x}$$

21)  $\tilde{x} = \tilde{A}_1 e_1 + \tilde{A}_2 e_2 + \tilde{w} = E \tilde{A} + \tilde{w}$   
where  $E = (e_1, e_2)$ ,  $\tilde{A} = (\tilde{A}_1, \tilde{A}_2)^T$

The MLE will minimize

$$\begin{aligned} J(\tilde{A}, f) &= (\tilde{x} - E \tilde{A})^H (\tilde{x} - E \tilde{A}) \\ &= \tilde{x}^H \tilde{x} - \tilde{x}^H E \tilde{A} - \tilde{A}^H E^H \tilde{x} + \tilde{A}^H E^H E \tilde{A} \end{aligned}$$

Taking the complex gradient we have

$$\frac{\partial J}{\partial \tilde{A}} = 0 - \frac{\partial}{\partial \tilde{A}} \tilde{x}^H E \tilde{A} - 0 + \frac{\partial}{\partial \tilde{A}} \tilde{A}^H E^H E \tilde{A}$$

Using (15.44) and (15.46)

$$\frac{\partial J}{\partial \tilde{A}} = - (E^H \tilde{x})^* + (E^H E \tilde{A})^* = 0$$

$$\Rightarrow \hat{\tilde{A}} = (E^H E)^{-1} E^H \tilde{x}$$

$$\begin{aligned} J(\hat{\tilde{A}}, f) &= (\tilde{x} - E(E^H E)^{-1} E^H \tilde{x})^H \\ &\quad \cdot (\tilde{x} - E(E^H E)^{-1} E^H \tilde{x}) \end{aligned}$$

$$= \tilde{x}^H (\mathbb{I} - E(E^H E)^{-1} E^H) \tilde{x}$$

Since  $\mathbb{I} - E(E^H E)^{-1} E^H$  is idempotent

To minimize over  $f$  we need to maximize

$$\mathcal{J}(f) = \tilde{\underline{x}}^H \underline{\xi} (\underline{\xi}^H \underline{\xi})^{-1} \underline{\xi}^H \tilde{\underline{x}}$$

This will require a 2-D search. An approximate MLE for  $|f_1 - f_2| \gg 1/N$  is found as follows:

$$\underline{\xi}^H \underline{\xi} = \begin{bmatrix} \underline{\xi}_1^H \\ \underline{\xi}_2^H \end{bmatrix} \begin{pmatrix} \underline{\xi}_1 & \underline{\xi}_2 \end{pmatrix} = \begin{pmatrix} N \underline{\xi}_1^H \underline{\xi}_2 \\ \underline{\xi}_2^H \underline{\xi}_1 \end{pmatrix}$$

$$\begin{aligned} \underline{\xi}_1^H \underline{\xi}_2 &= \sum_{n=0}^{N-1} e^{-j2\pi f_1 n} e^{j2\pi f_2 n} \\ &= \sum_{n=0}^{N-1} e^{j2\pi \Delta f n} \quad \Delta f = f_2 - f_1 \\ &= \frac{1 - e^{j2\pi \Delta f N}}{1 - e^{j2\pi \Delta f}} \\ &= \frac{e^{j\pi \Delta f N}}{e^{j\pi \Delta f}} \quad \frac{e^{-j\pi \Delta f N} - e^{j\pi \Delta f N}}{e^{-j\pi \Delta f} - e^{j\pi \Delta f}} \\ &= e^{j\pi \Delta f (N-1)} \quad \frac{\sin \pi \Delta f N}{\sin \pi \Delta f} \\ &\quad | \underline{\xi}_1^H \underline{\xi}_2 | \end{aligned}$$

For  $|\Delta f| \gg 1/N$ ,  $\underline{\xi}_1^H \underline{\xi}_2 \ll N$  and thus

$$\underline{\xi}^H \underline{\xi} \approx N \underline{\xi}$$

$$\mathcal{J}(f) \approx 1/N \tilde{\underline{x}}^H \underline{\xi}^H \underline{\xi} \tilde{\underline{x}} = \frac{1}{N} \| \underline{\xi}^H \tilde{\underline{x}} \|^2$$

$$\begin{aligned}
 &= \frac{1}{N} (|\underline{e}_1^H \hat{\underline{x}}|^2 + |\underline{e}_2^H \hat{\underline{x}}|^2) \\
 &= \frac{1}{N} \left| \sum_{n=0}^{N-1} \hat{x}(n) e^{-j2\pi f_1 n} \right|^2 \\
 &\quad + \frac{1}{N} \left| \sum_{n=0}^{N-1} \hat{x}(n) e^{-j2\pi f_2 n} \right|^2
 \end{aligned}$$

Approximate MLE finds peak locations of the two largest peaks of a single periodogram, with the constraint  $|f_1 - f_2| \gg 1/N$ .

Without constraint we could have



22) An MLE will minimize

$$\sum_{n=0}^{N-1} |\tilde{x}(n) - \tilde{A}\tilde{s}(n)|^2$$

$$\text{where } \tilde{s}(n) = e^{j2\pi(f_0 n + f_2 \alpha n)}$$

This is a linear LS problem with respect to  $\tilde{A}$ . Thus, from Example 15.2

$$\hat{A} = \frac{\sum_n \tilde{x}(n) \tilde{s}^*(n)}{\sum_n |\tilde{s}(n)|^2}$$

Substituting for  $\hat{A}$

$$\begin{aligned}
 & \sum_{n=0}^{N-1} (\tilde{x}(n) - \hat{A} \tilde{s}(n))^* (\tilde{x}(n) - \hat{A} \tilde{s}(n)) \\
 &= \sum_{n=0}^{N-1} \tilde{x}^*(n) (\tilde{x}(n) - \hat{A} \tilde{s}(n)) \\
 &\quad - \hat{A} \underbrace{\sum_{n=0}^{N-1} \tilde{s}^*(n) (\tilde{x}(n) - \hat{A} \tilde{s}(n))}_{= 0} \\
 &= \sum_n |\tilde{x}(n)|^2 - \hat{A} \sum_n \tilde{x}^*(n) \tilde{s}(n) \\
 &= \sum_n |\tilde{x}(n)|^2 - \frac{\left| \sum_n \tilde{x}(n) s^*(n) \right|^2}{\sum_n |s(n)|^2} \\
 &= \sum_n |\tilde{x}(n)|^2 - \frac{\left| \sum_n \tilde{x}(n) e^{-j2\pi(f_0 n + \frac{1}{2}\alpha n^2)} \right|^2}{N}
 \end{aligned}$$

Hence, we need to maximize over  $f_0, \alpha$

$$\left| \sum_{n=0}^{N-1} \tilde{x}(n) e^{-j2\pi(f_0 n + \frac{1}{2}\alpha n^2)} \right|^2. \text{ To}$$

Compute this efficiently let  $\tilde{y}(n) = \tilde{x}(n) e^{-j2\pi f_0 n}$

$$\Rightarrow \left| \sum_{n=0}^{N-1} \tilde{y}(n) e^{-j2\pi f_0 n} \right|^2, \text{ can use}$$

FFT on  $\tilde{y}(n)$  for each  $\alpha$ .

$$23) \hat{\alpha} = E(\alpha | u, v) = \int \alpha p(\alpha | u, v) d\alpha$$

$$\hat{\beta} = E(\beta | u, v) = \int \beta p(\beta | u, v) d\beta$$

$$\alpha \hat{\alpha} = \iint \alpha p(\alpha, \beta | u, v) d\alpha d\beta$$

$$\hat{\beta} = \iint \beta p(\alpha, \beta | u, v) d\alpha d\beta$$

$$\hat{\theta} = \hat{\alpha} + \hat{\beta} = \iint (\alpha + \beta) p(\alpha, \beta | u, v) d\alpha d\beta$$

$$= \iint \theta p(\alpha, \beta | u, v) d\alpha d\beta$$

$$= E(\theta | u, v) = E(\varrho | \tilde{x})$$

Note that  $p(\varrho | \tilde{x})$  is just  $p(\alpha, \beta | u, v)$  in disguise.

24) From (15.52) with  $\tilde{x} = P_0 Q$  where  $Q$  is the covariance matrix corresponding to a process with PSD  $Q(t)$ ,

$$I(P_0) = \text{tr} \left( \underline{C}_{\tilde{x}}' \frac{\partial \underline{C}_{\tilde{x}}}{\partial P_0} \underline{C}_{\tilde{x}}' \frac{\partial \underline{C}_{\tilde{x}}}{\partial P_0} \right)$$

$$= \text{tr} \left( \frac{1}{P_0} Q^{-1} Q \frac{1}{P_0} Q^{-1} Q \right)$$

$$= \frac{1}{P_0^2} \text{tr}(I) = N/P_0^2$$

$$\Rightarrow \text{var}(\hat{P}_0) \geq P_0^2/N$$

From (15.68)

$$I(P_0) = N \int_{-\frac{L}{2}}^{\frac{L}{2}} \left( \frac{\partial \ln P_{X|X}(t)}{\partial P_0} \right)^2 dt$$

$$= N \int_{-\frac{1}{2}}^{\frac{1}{2}} \left( \frac{\partial \ln \frac{P_0 Q(t)}{P_0}}{\partial P_0} \right)^2 dt$$

$$= N \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{P_0^2} dt = N/P_0^2$$

Same result (not true in general)

$$\begin{aligned} 25) \quad X(f_k) &= \sum_n \tilde{x}(n) e^{-j2\pi f_k n} \quad (f_k = k/N) \\ &= \sum_n (\tilde{A} e^{j2\pi f_c n} + \tilde{w}(n)) e^{-j2\pi f_k n} \\ &= \sum_n \tilde{w}(n) e^{-j2\pi f_k n} \quad f_k \neq f_c \\ &\quad \sum_n (\tilde{A} + \tilde{w}(n)) e^{-j2\pi f_k n} \quad f_k = f_c \\ &= W(f_k) \quad k \neq c \\ &\quad N\tilde{A} + W(f_k) \quad k = c \end{aligned}$$

where  $W(f_k)$  is the DFT of  $\tilde{w}(n)$ .

But  $\tilde{w} \sim CN(0, N\sigma^2 I)$  from Sect 15.9

$$\Rightarrow \underline{X} \sim CN \left( \begin{bmatrix} 0 \\ \vdots \\ 0 \\ N\tilde{A} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, N\sigma^2 I \right)$$

where  $\underline{X} = [X(f_0) \ X(f_1) \ \dots \ X(f_{N-1})]^T$

$$\begin{aligned} \text{Now, } SNR(\text{input}) &= |E(\tilde{x}(n))|^2 / \sigma^2 \\ &= |\tilde{A}|^2 / \sigma^2 \end{aligned}$$

$$SNR(\text{output}) = \frac{|E(X(f_c))|^2}{\text{var}(X(f_c))}$$

$$= \frac{|N\tilde{A}|^2}{N\sigma^2} = \frac{N|\tilde{A}|^2}{\sigma^2}$$

Processing gain =  $10 \log_{10} N$  dB

To detect a sinusoid of unknown frequency we could form

$$\frac{1}{N} |x(f_k)|^2 = \frac{1}{N} \left| \sum_{n=0}^{N-1} \tilde{x}(n) e^{-j2\pi f_k n} \right|^2$$

and choose the maximum over  $f_k$  for comparison to a threshold. If no signal is present,

$$\begin{aligned} E\left(\frac{1}{N} |x(f_k)|^2\right) &= \frac{1}{N} \text{var}(x(f_k)) \\ &= \sigma^2 \quad \text{for all } k \end{aligned}$$

and for a signal present

$$\begin{aligned} E\left(\frac{1}{N} |x(f_k)|^2\right) &= \frac{1}{N} [\text{var}(x(f_k)) \\ &\quad + |E(x(f_k))|^2] \end{aligned}$$

$$= \frac{1}{N} (N\sigma^2 + N^2 |\tilde{A}|^2)$$

$$= \sigma^2 + N|\tilde{A}|^2 \quad \text{for } k=2$$

$$= \sigma^2 \quad \text{for } k \neq 2$$

This statistic is just a sampled in frequency periodogram!