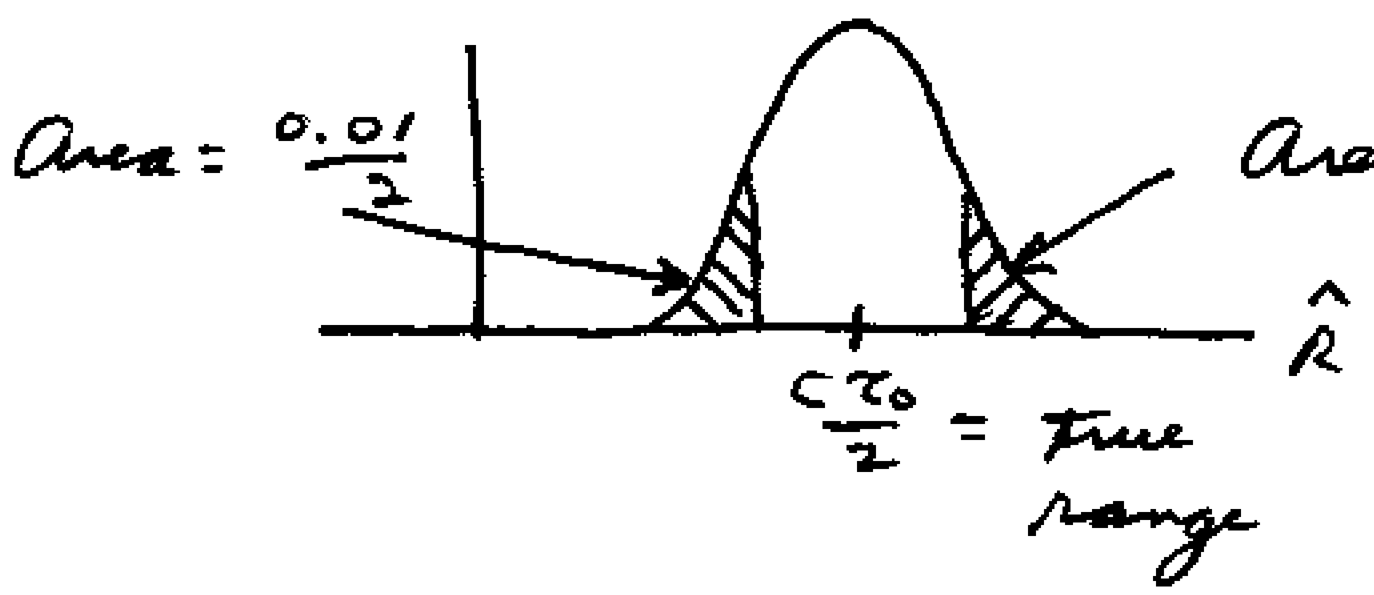
PROBLEM SOLUTIONS

FUNDAMENTALS OF STATISTICAL SIGNAL PROCESSING: ESTIMATION THEORY

BY STEVEN KAY

Chapter 1

1) Since $R = c\tau_0/2$, we use $\hat{R} = c\hat{\tau}_0/2$.
The PDF is from $\hat{\tau}_0 \sim N(\tau_0, \sigma^2_{\hat{\tau}_0})$,
$\hat{R} \sim N(c\tau_0/2, \frac{c^2}{4}\sigma^2_{\hat{\tau}_0})$



To be within 100 m we must have
$$Pr\{|\hat{R} - c\tau_0/2| < 100\} = 0.99$$
$$\Rightarrow Pr\left\{ \left| \underbrace{\frac{\hat{R} - c\tau_0/2}{\frac{c}{2}\sigma_{\hat{\tau}_0}}}_{N(0,1)} \right| < \frac{100}{\frac{c}{2}\sigma_{\hat{\tau}_0}} \right\} = 0.99$$

$$\Rightarrow \quad \frac{100}{\frac{c}{2}\sigma_{\hat{\tau}_0}} = 2.58 \text{ or } \sigma_{\hat{\tau}_0} = 2.6 \times 10^{-7} \text{ sec}$$
$$= 0.26 \ \mu\text{sec}$$

2) No, in fact θ could have been any value.
If θ 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 θ 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 θ 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,

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

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

$$\mathrm{var}(\check{A}) = \frac{1}{(N+2)^2}\left[4\sigma^2 + \sum_{n=1}^{N-2}\sigma^2 + 4\sigma^2\right]$$

$$= \frac{N+6}{(N+2)^2}\sigma^2 = \frac{N+6}{(N+2)^2}$$

$$\mathrm{var}(\check{A}) - \mathrm{var}(\hat{A}) = \frac{N+6}{(N+2)^2} - \frac{1}{N}$$

$$= \frac{N(N+6) - (N+2)^2}{N(N+2)^2}$$

$$= \frac{2N-4}{N(N+2)^2} > 0 \quad \text{for } N > 2$$

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 SNR).

Chapter 2

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

$$\mathrm{var}(\hat{\sigma}^2) = \frac{1}{N^2}\mathrm{var}\left(\sum_n x^2[n]\right)$$

$$= \frac{1}{N^2} N \,\mathrm{var}(x^2[n]) \quad (x[n]\text{'s are IID} \Rightarrow x^2[n] \text{ are IID})$$

$$= \frac{1}{N}\mathrm{var}(x^2[n])$$

$$\mathrm{var}(x^2[n]) = E(x^4[n]) - E(x^2[n])^2$$
$$= 3\sigma^4 - \sigma^4 = 2\sigma^4$$
$$\Rightarrow \mathrm{var}(\hat{\sigma}^2) = 2\sigma^4/N \to 0 \text{ as } N \to \infty$$
Hence, the PDF of $\hat{\sigma}^2$ collapses about the true value as $N \to \infty$.

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

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

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

$$= \frac{1}{N^2} N\, var(x[n])$$

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

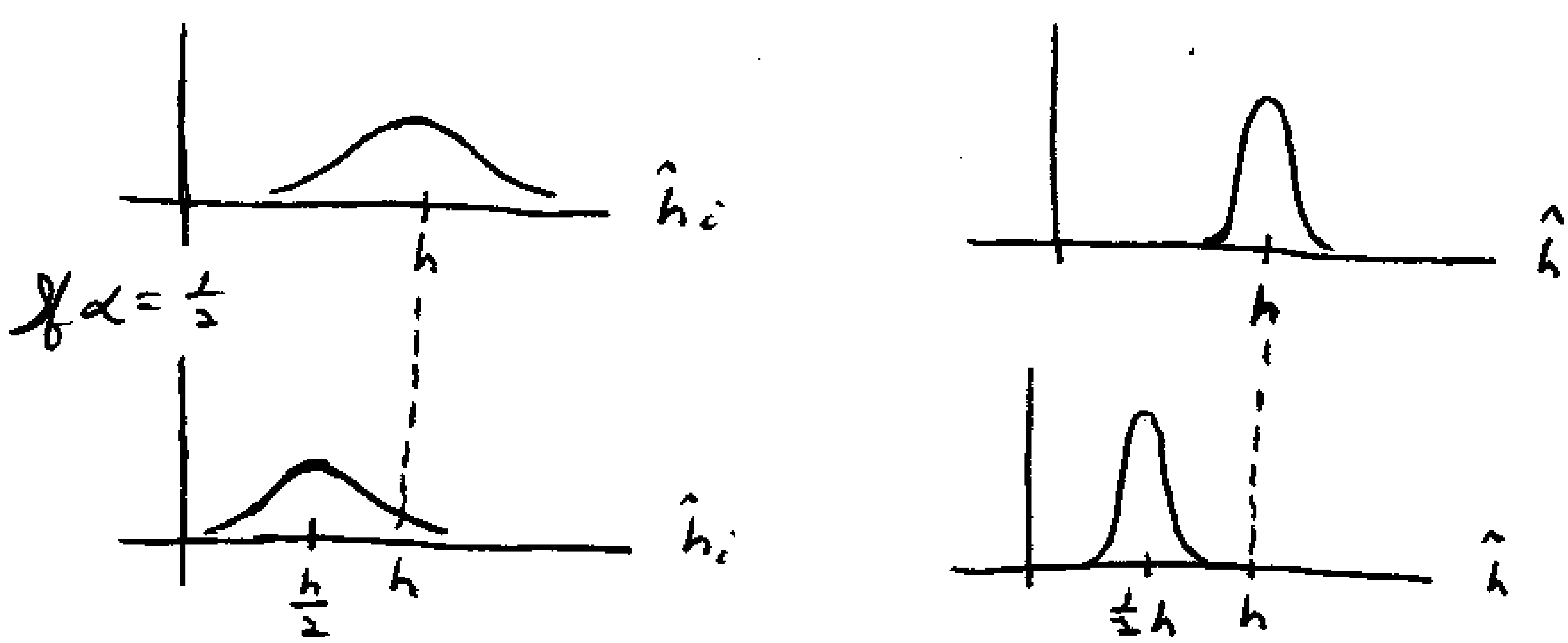
$$\Rightarrow var(\hat{A}) = \sigma^2/N.$$

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

$= \alpha h$

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

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) $x[n] \sim \mathcal{N}(0, \sigma^2)$ or $\frac{x[n]}{\sigma} \sim \mathcal{N}(0,1)$

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

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

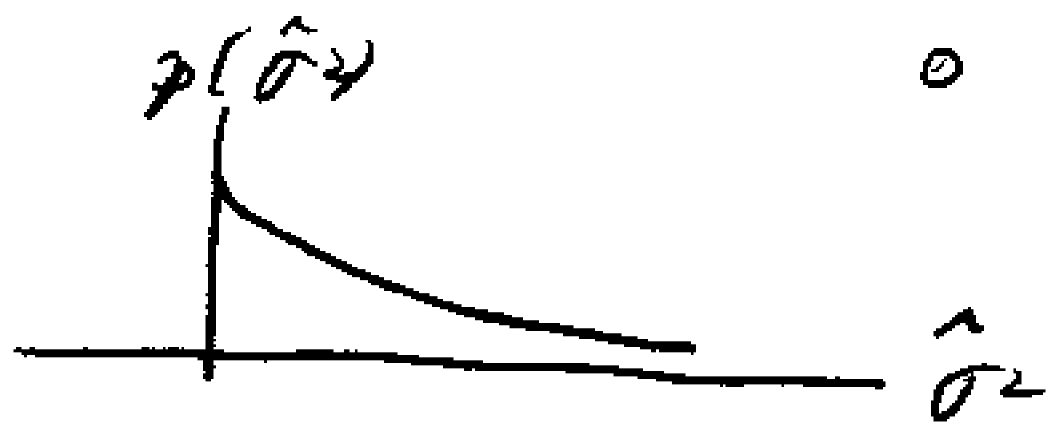
or $p(y) = \begin{cases} \frac{1}{2} e^{-y/2} & y > 0 \\ 0 & y < 0 \end{cases}$

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

$$p(\hat{\sigma}^2) = \frac{p_y(y(\hat{\sigma}^2))}{|d\hat{\sigma}^2/dy|}$$

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

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



Clearly, not symmetric.

But $E(\hat{\sigma}^2) = \int_0^\infty \hat{\sigma}^2 \frac{1}{\sigma^2} e^{-\hat{\sigma}^2/\sigma^2} d\hat{\sigma}^2$

$$= \sigma^2 \int_0^\infty u e^{-u} du$$
$$= \sigma^2 \left[-u e^{-u} - e^{-u} \right] \Big|_0^\infty$$
$$= \sigma^2$$

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

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

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

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

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

$\Rightarrow a_i = -\lambda / 2\sigma^2$ 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) $\dfrac{\hat{\theta}-\theta}{\sqrt{\mathrm{var}(\hat{\theta})}} \sim N(0,1) \quad \dfrac{\check{\theta}-\theta}{\sqrt{\mathrm{var}(\check{\theta})}} \sim N(0,1)$

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

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

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

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

If $\mathrm{var}(\hat{\theta}) < \mathrm{var}(\check{\theta})$

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

or $P_r\{|\hat{\theta}-\theta|>\epsilon\} < P_r\{|\check{\theta}-\theta|>\epsilon\}$

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

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

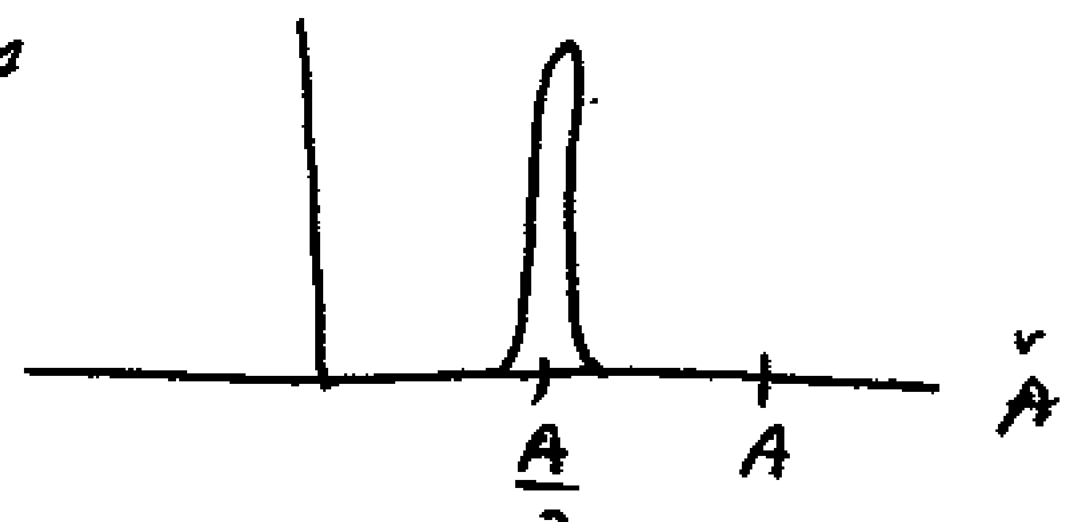
$$= 2\,\Phi\left(\frac{-\epsilon}{\sqrt{\sigma^2/N}}\right) \to 0$$

since $\dfrac{-\epsilon}{\sqrt{\sigma^2/N}} \to -\infty$ as $N\to\infty$

If $\check{A} = \dfrac{1}{2N}\sum_{n=0}^{N-1} x[n]$ is used,

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

as $N \to \infty$, the variance $\to 0$ and the PDF approaches



Thus, $\lim_{N\to\infty} \Pr\{|\check{A} - A| > \epsilon\} = 1$ for any $\epsilon > 0$

$\hat{A}$ is <u>consistent</u> while $\check{A}$ is inconsistent.

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

$E(\hat{\theta}) = E[\hat{A}^2] = \mathrm{var}(\hat{A}) + E(\hat{A})^2$

$= \sigma^2/N + A^2 = \theta + \sigma^2/N \neq \theta$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

or $\int_0^{1/\theta} g(x[0])\, \theta \, dx[0] = \theta$

$\int_0^{1/\theta} g(u)\, du = 1$ for all $\theta > 0$

Now suppose a g could be found. Then

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

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

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

and subtracting the two gives

$$\int_{1/\theta_1}^{1/\theta_2} g(u)\, du = 0 \quad \text{for any } \theta_2 < \theta_1$$

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

Chapter 3

1) $$p(x[n];\theta) = \frac{1}{\theta}\left(u(x[n]) - u(x[n]-\theta)\right)$$

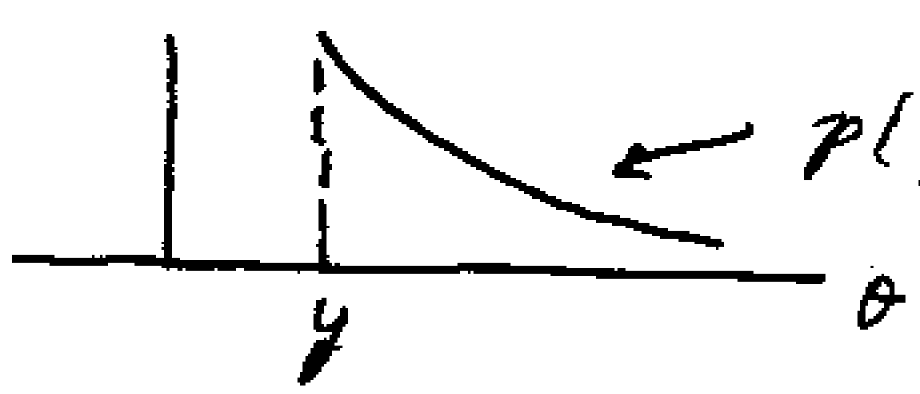
where $u(x) = 1 \quad x > 0$
$\qquad\qquad\quad 0 \quad x < 0$

Since $p(\underline{x};\theta) = \prod_{n=0}^{N-1} p(x[n];\theta)$, 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

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



$p(y;\theta)$ as function of θ with y fixed and $y>0$

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

2) $x[0] = A + w[0]$

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

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

$$= E\left[\left(\frac{\partial \ln p(x[0]-A)}{\partial (x[0]-A)}(-1)\right)^2\right]$$

$$= E\left[\left(\frac{1}{p(x[0]-A)}\frac{\partial p(x[0]-A)}{\partial (x[0]-A)}\right)^2\right]$$

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

$$= \int_{-\infty}^{\infty}\left(\frac{\partial p(x[0]-A)}{\partial (x[0]-A)}\right)^2 \frac{1}{p^2(x[0]-A)}\, p(x[0]-A)\, dx[0]$$

Letting $u \equiv x[0] - A$

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

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

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

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

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

$$= 2/\sigma^2$$

$\Rightarrow \operatorname{var}(\hat{A}) \geq \sigma^2/2$ 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) $$p(\underline{x};A) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x(n)-Ar^n)^2}$$

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

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

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

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

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

To show that an efficient estimator exists

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

$$= \underbrace{\frac{\sum_n r^{2n}}{\sigma^2}}_{I(A)}\left(\underbrace{\frac{\sum_n x(n)r^n}{\sum_n r^{2n}}}_{\hat{A}} - A\right)$$

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

$$\mathrm{var}(\hat{A}) \to \sigma^2(1-r^2) \qquad 0<r<1$$

$$\rightarrow 0 \qquad r \geq 1$$

as $N \to \infty$.

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.

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

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

$$= -\frac{1}{\sigma^2}\sum_n n^2 r^{2n-2}$$

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

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

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

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

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

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

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

6)

$$x[0] \sim N(\theta, 1)$$

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

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

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

For $\theta \geq 0$

$$\frac{\partial \ln p}{\partial \theta} = -\frac{1}{2}\left[2(x[0]-\theta)(-1) + 2(x[1]-\theta)(-1)\right]$$

$$= (x[0]-\theta) + (x[1]-\theta)$$

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

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

For $\theta < 0$

$$\frac{\partial \ln p}{\partial \theta} = -\frac{1}{2}\left(-2(x[0]-\theta) - (x[1]-\theta)\right)$$

$$= (x[0]-\theta) + \frac{1}{2}(x[1]-\theta)$$

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

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

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

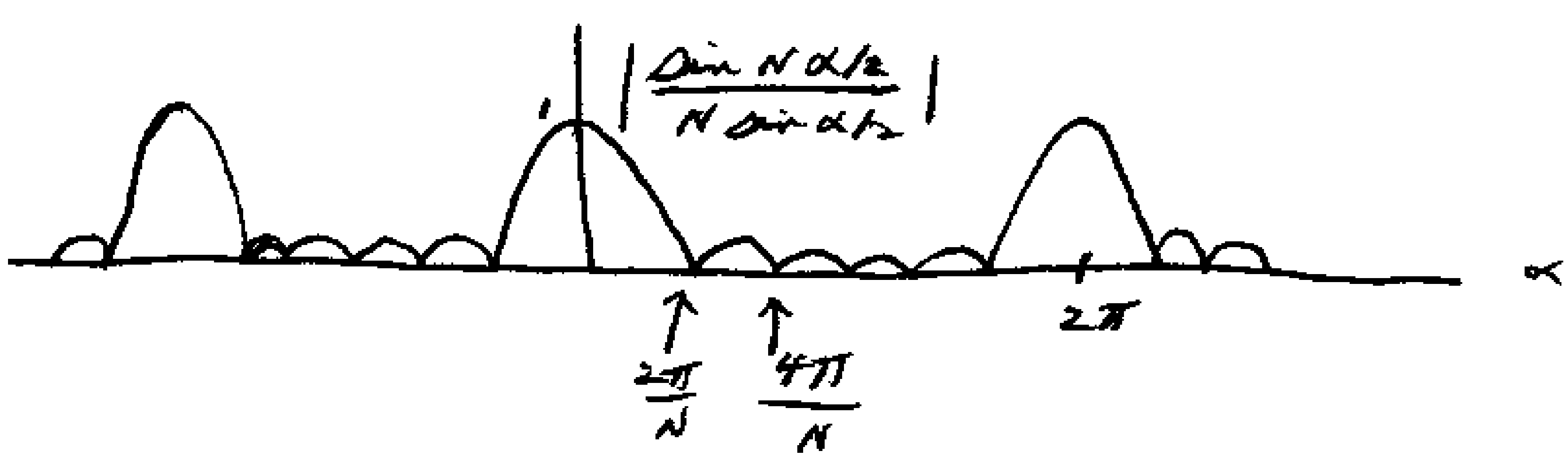
$$\frac{1}{N}\sum_{n=0}^{N-1} \cos(\alpha n + \beta) = \frac{1}{N} \mathrm{Re}\left(\sum_n e^{j(\alpha n + \beta)} \right)$$

$$= \frac{1}{N} \mathrm{Re}\left[e^{j\beta} \frac{1 - e^{j\alpha N}}{1 - e^{j\alpha}} \right]$$

$$= \frac{1}{N} \mathrm{Re}\left(e^{j\beta} \frac{e^{j\alpha N/2}}{e^{j\alpha/2}} \frac{e^{-j\alpha N/2} - e^{j\alpha N/2}}{e^{-j\alpha/2} - e^{j\alpha/2}} \right)$$

$$= \frac{1}{N} \mathrm{Re}\left[e^{j\beta} e^{j\alpha\left(\frac{N-1}{2}\right)} \frac{\sin N\alpha/2}{\sin \alpha/2} \right]$$

$$= \frac{\sin N\alpha/2}{N \sin \alpha/2} \cos\left[\alpha\left(\frac{N-1}{2}\right) + \beta\right)$$



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

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

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

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

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

$$= \frac{1}{\sigma^4}\sum_n \sigma^2 = N/\sigma^2$$

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

9) Using (3.32)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\left(\underline{I}^{-1}(\underline{\theta})\right)_{11} = \frac{c}{ac-b^2} = \frac{1}{a-b^2/c} \geq \frac{1}{a}$$

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) $$1^2 = \left(\underline{e}_i^T \sqrt{\underline{I}(\underline{\theta})}\sqrt{\underline{I}^{-1}(\underline{\theta})}\,\underline{e}_i\right)^2$$

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

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

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

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

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

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

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left(\frac{\partial}{\partial \theta_i} \sum_{k=0}^{p-1} A_k n^k \right) \left(\frac{\partial}{\partial \theta_j} \sum_{k=0}^{p-1} A_k n^k \right)$$

$$= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} n^{i-1} n^{j-1} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} n^{i+j-2}$$

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

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

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

$$E(\hat{A} \mid A = A_0) = A_0$$
$$\mathrm{var}(\hat{A} \mid A = A_0) = \sigma^2/N$$

and since $\sigma^2/N \to 0$ as $N \to \infty$, $\hat{A} \to A_0$. Now consider A as a random variable

As $N \to \infty$

$$\mathrm{var}(\hat{\sigma}_A^2) = \mathrm{var}(\hat{A}^2)$$
$$\to \mathrm{var}(A^2)$$

But $\mathrm{var}(A^2) = E(A^4) - E(A^2)^2$

$$= 3\sigma_A^4 - \sigma_A^4 = 2\sigma_A^4$$

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

$N \to \infty$, although we can nullify the noise effects by averaging ($\hat{A} \to A_0$), 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 $\underline{x}(n)$'s are independent

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

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

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

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

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

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

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

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

16) $$I(P_0) = \frac{1}{2} \mathrm{tr}\left[\left(\underline{C}^{-1}(P_0)\frac{\partial \underline{C}(P_0)}{\partial P_0}\right)^2\right]$$

Let $r_{xx}[k] = \mathcal{F}^{-1}\{P_{xx}(f)\}$
$= \mathcal{F}^{-1}\{P_0 Q(f)\}$
$= P_0 \mathcal{F}^{-1}\{Q(f)\}$

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

$$(\underline{C}_q)_{ij} = q[i-j]$$

Then, $\underline{C}(P_0) = P_0 \underline{C}_q$ and

$$\underline{C}^{-1}(P_0)\frac{\partial \underline{C}(P_0)}{\partial P_0} = \frac{1}{P_0}\underline{C}_q^{-1}\underline{C}_q = \frac{1}{P_0}\underline{I}$$

$$I(P_0) = \frac{1}{2}\mathrm{tr}\left(1/P_0^2\,\underline{I}^2\right) = \frac{N}{2P_0^2}$$

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

Using the asymptotic form

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

$$= \frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}}\left(\frac{\partial \ln P_0 Q(f)}{\partial P_0}\right)^2 df$$

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

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

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

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

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

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

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

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

But $\displaystyle\sum_{n=-M}^{M} n^2 = 2\sum_{n=1}^{M} n^2 = \frac{2M(M+1)(2M+1)}{6}$

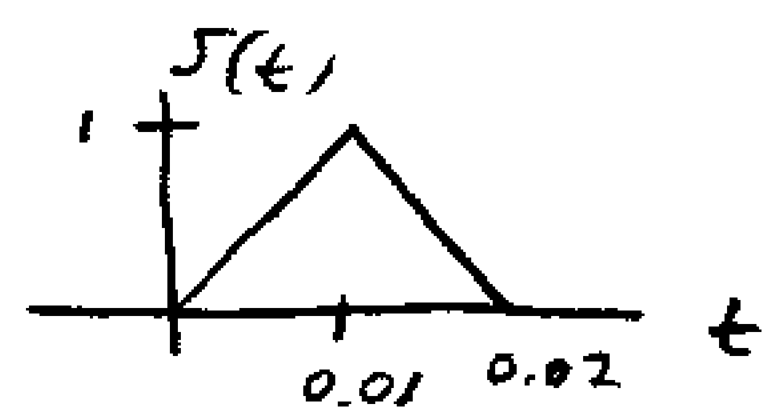
$$= \frac{1}{3}\left(\frac{N-1}{2}\right)\left(\frac{N+1}{2}\right)N = \frac{N(N^2-1)}{12}$$

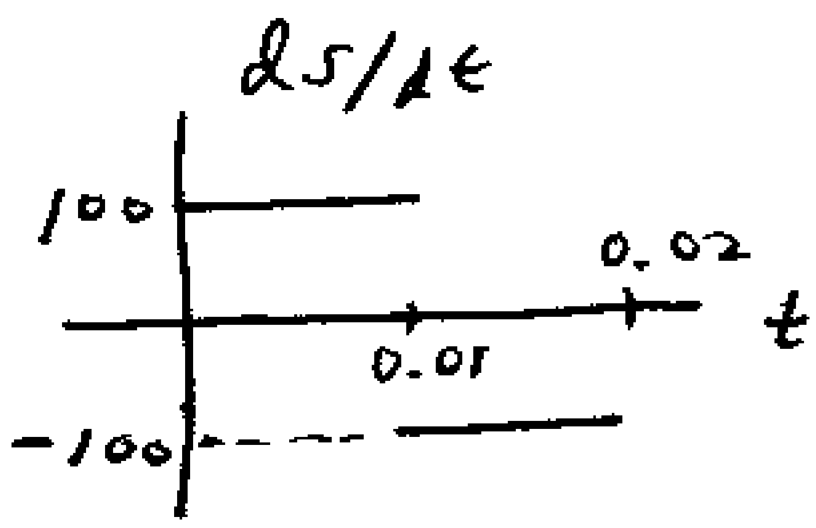
$$\mathrm{var}(\hat{f}_0) \geq \frac{6\sigma^2}{A^2\pi^2 N(N^2-1)} = \frac{3}{\eta\pi^2 N(N^2-1)}$$

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

18) $$\mathrm{var}(\hat{R}) \geq \frac{c^2/4}{\frac{\mathcal{E}}{N_0/2}\,\overline{F^2}}$$

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





$$\overline{F^2} = \frac{\int_0^{0.02} (100)^2\,dt}{\mathcal{E}} = \frac{200}{\mathcal{E}}$$

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

$$= 0.00281$$

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

19) From Example 3.15

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

For $\beta = 90°$, $\eta = 1$, $F_0 = 10$, $L = (M-1)d$
$= (M-1)\lambda/2$

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

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

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

or $M \geq 6$. But then

$$L = (M-1)\lambda/2 = (M-1)\frac{c}{2F_0} = \frac{5(3\times10^8)}{2\times10^6}$$

$= 750$ m

This is clearly impossible.

20) $$\mathrm{var}(\hat{P}_{xx}(f)) \geq \frac{\left(\frac{\partial P_{xx}(f)}{\partial a(1)}\right)^2}{I(a(1))}$$

$$= \frac{(\partial P_{xx}(f)/\partial a(1))^2}{N/(1-a(1))^2}$$

using results from Example 3.16.

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

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

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

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

$$= \frac{-\sigma_u^2}{|A(f)|^4} \, 2\underbrace{\mathrm{Re}\left(A(f)e^{j2\pi f} \right)}_{a[1]+\cos 2\pi f}$$

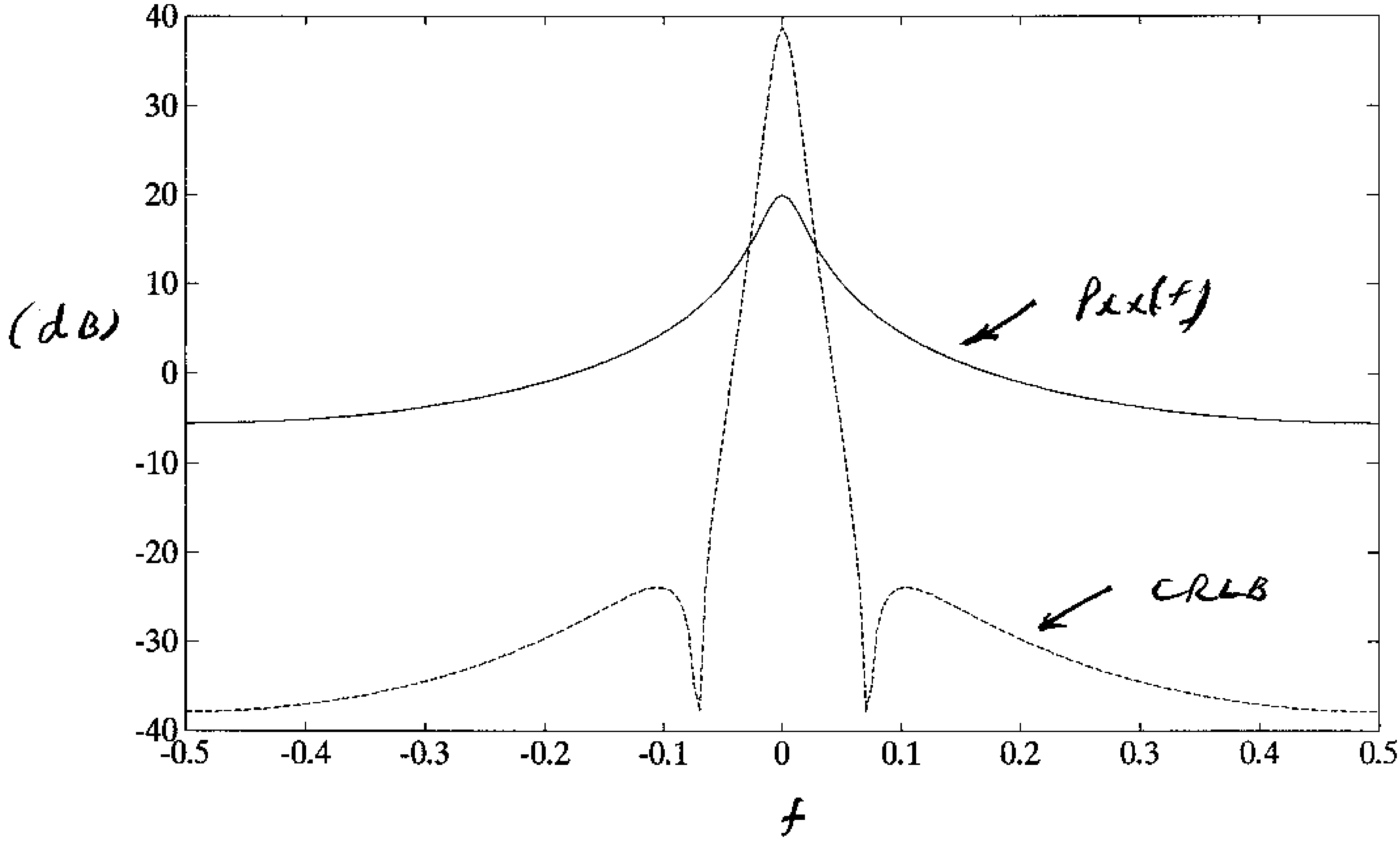
$$\mathrm{var}(\hat{P}_{xx}(f)) \geq \frac{4\sigma_u^4}{N}(1-a^2[1]) \frac{(a[1]+\cos 2\pi f)^2}{|A(f)|^8}$$

For the given values

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

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

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

$$\frac{\partial P_{xx}(f)}{\partial a(1)}\bigg|_{f=0} = \frac{-\sigma_u^2 \; 2(1+a(1))}{(1+a(1))^4}$$

$$= \frac{-2\sigma_u^2}{(1+a(1))^3}$$

and for this example

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

Chapter 4

1) This fits linear model form.

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

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

$$\hat{\underline{\theta}} = (\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{x} \qquad \underline{C}_{\hat{\theta}} = \sigma^2(\underline{H}^T\underline{H})^{-1}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$= \frac{1}{2\epsilon^2}\begin{bmatrix} 3+2\epsilon+\epsilon^2 & -(3+\epsilon) \\ -(3+\epsilon) & 3 \end{bmatrix}\begin{bmatrix} 6 \\ 6+2\epsilon \end{bmatrix}$$

$$= \frac{1}{2\epsilon^2}\begin{bmatrix} 18+12\epsilon+6\epsilon^2-18-6\epsilon-6\epsilon-2\epsilon^2 \\ -18-6\epsilon+18+6\epsilon \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

$$= \frac{1 - e^{j\frac{2\pi}{N}(k+l)N}}{1 - e^{j\frac{2\pi}{N}(k+l)}} = \frac{1 - e^{j2\pi(k+l)}}{1 - e^{j\frac{2\pi}{N}(k+l)}} = 0$$

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

6) From Example 4.2

$\hat{a}_k \sim N(a_k, 2\sigma^2/N) \quad \hat{b}_k \sim N(b_k, 2\sigma^2/N)$

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

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

$$= \frac{\mathrm{var}(\hat{a}_k) + E(\hat{a}_k)^2 + \mathrm{var}(\hat{b}_k) + E(\hat{b}_k)^2}{2}$$

$$= \frac{a_k^2 + b_k^2 + 4\sigma^2/N}{2} = P + \frac{2\sigma^2}{N}$$

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

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

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

If $z \sim N(\mu, \sigma^2)$

$$\mathrm{var}(z^2) = 4\mu^2\sigma^2 + 2\sigma^4$$

(see development preceding (3.19))

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

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

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

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

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

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

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

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 $N >> p$, these terms will be negligible.
Now assume $i \geq j$ and let $m = n-i$

$$[\underline{H}^T\underline{H}]_{ij} = \sum_{m=-\infty}^{\infty} u[m]u[m+i-j]$$

$$= \sum_{m=0}^{N-1-(i-j)} u[m]u[m+i-j]$$

Similarly for $i < j$

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

$$\Rightarrow \quad [\underline{H}^T\underline{H}]_{ij} = \sum_{m=0}^{N-1-|i-j|} u[m]u[m+|i-j|]$$

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

$$\begin{aligned}
r_{ux}[k] &= E\{u[n]\,x[n+k]\} \\
&= E\{u[n]\sum_{l=0}^{\infty} h[l]\,u[n+k-l]\} \\
&= \sum_{l=0}^{\infty} h[l]\,E\{u[n]\,u[n+k-l]\} \\
&= \sum_{l=0}^{\infty} h[l]\,r_{uu}[k-l] \\
&= h[k] \star r_{uu}[k]
\end{aligned}$$

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

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) $$p(\underline{x};\underline{\theta}) = \frac{1}{(2\pi)^{N/2}\det(\underline{C})^{1/2}} e^{-\frac{1}{2}(\underline{x}-\underline{H}\underline{\theta})^T\underline{C}^{-1}(\underline{x}-\underline{H}\underline{\theta})}$$

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

$$= -\frac{1}{2}\left[-2\underline{H}^T\underline{C}^{-1}\underline{x} + 2\underline{H}^T\underline{C}^{-1}\underline{H}\underline{\theta}\right) \quad \text{using (4.3)}$$

$$= \underline{H}^T\underline{C}^{-1}\underline{x} - \underline{H}^T\underline{C}^{-1}\underline{H}\underline{\theta}$$

$$= \underbrace{\underline{H}^T\underline{C}^{-1}\underline{H}}_{\underline{I}(\underline{\theta})}\left(\underbrace{(\underline{H}^T\underline{C}^{-1}\underline{H})^{-1}\underline{H}^T\underline{C}^{-1}\underline{x}}_{\hat{\underline{\theta}}} - \underline{\theta}\right)$$

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

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

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

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

$\underline{C}^{-1} = \text{diag}(1/\sigma_0^2, 1/\sigma_1^2, \ldots, 1/\sigma_{N-1}^2)$

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

$$\hat{A} = \sum_{n=0}^{N-1} d_n x'[n]$$

$$\text{where } d_n = \frac{[\underline{D}\underline{1}]_n}{\underline{1}^T\underline{D}^T\underline{D}\underline{1}} = \frac{1/\sigma_n}{\sum_{m=0}^{N-1} 1/\sigma_m^2}$$

Since $\underline{C}$ 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 all 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 a $\sigma_n^2 = 0$, say σ_m^2, then we cannot prewhiten the data. In this case, however, as $\sigma_m^2 \to 0$, $d_n \to 0$ for $n \neq m$ and $d_n \to \sigma_m$ for $n = m$. Thus, $\hat{A} \to d_m x'[m] = x[m]$, as expected.

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

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

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

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

$$\underline{H}^T \underline{C}^{-1} \underline{x} = \frac{\underline{1}^T}{\sigma^2(1-\rho^2)} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \end{bmatrix}$$

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

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

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

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

We don't need prewhitening here because $\underline{H}$ is an eigenvector of $\underline{C}$. Hence,

$$(\underline{H}^T\underline{C}^{-1}\underline{H})^{-1}\underline{H}^T\underline{C}^{-1} = (\underline{H}^T\tfrac{1}{\lambda}\underline{H})^{-1}\tfrac{1}{\lambda}\underline{H}^T$$

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

As $\rho \to 1$, $\mathrm{var}(\hat{A}) \to \sigma^2$
As $\rho \to -1$, $\mathrm{var}(\hat{A}) \to 0$.
See Prob 3.9 for explanation.

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

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

$$= \underline{A}\underline{\theta} + \underline{w}' = \underline{H}'\underset{r\times 1}{\underline{\alpha}} + \underline{w}' \qquad \text{where } \underline{H}' = \underline{I}$$

$$\underline{C}' = E(\underline{w}'\underline{w}'^T) = E(\underline{A}(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{w}\underline{w}^T\underline{H}(\underline{H}^T\underline{H})^{-1}\underline{A}^T)$$

$$= \sigma^2\underline{A}(\underline{H}^T\underline{H})^{-1}\underline{A}^T$$

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

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

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

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

$$\underline{C}_{\hat{\theta}} = E\{(\hat{\underline{\theta}}-\underline{\theta})(\hat{\underline{\theta}}-\underline{\theta})^T\}$$
$$= E\{((\underline{H}^T\underline{H})^{-1}\underline{H}^T(\underline{x}-\underline{H}\underline{\theta}))((\underline{H}^T\underline{H})^{-1}\underline{H}^T(\underline{x}-\underline{H}\underline{\theta}))^T\}$$
$$= E\left[(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{w}\underline{w}^T\underline{H}(\underline{H}^T\underline{H})^{-1}\right]$$
$$= E_{H|W} E_W \left\{(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{w}\underline{w}^T\underline{H}(\underline{H}^T\underline{H})^{-1}\right\}$$

$$= E_{H|W}\left[(\underline{H}^T\underline{H})^{-1}\underline{H}^T\sigma^2\underline{I}\,\underline{H}(\underline{H}^T\underline{H})^{-1}\right]$$
$$= E_{H|W}\left[\sigma^2(\underline{H}^T\underline{H})^{-1}\right] = \sigma^2 E_H\left[(\underline{H}^T\underline{H})^{-1}\right]$$

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

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

$$\underline{H} = [\underbrace{1\,1\ldots1}_{M}\;\underbrace{0\,0\ldots0}_{N-M}]^T \quad \text{with probability } \epsilon$$

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

$$= \frac{1}{N}\sum_{n=0}^{N-1} x[n] \qquad \text{no fade}$$

$$\frac{1}{M}\sum_{n=0}^{M-1} x[n] \qquad \text{fade}$$

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

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

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

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

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

Chapter 5

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

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

But $\sum_n x^2[n] = s\sigma^2$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

But the product is zero unless $-\theta \le x[n] \le \theta$ for all $x[n]$ or $\min x[n] \ge -\theta$, $\max x[n] \le \theta$

or $\max |x[n]| \le \theta$ 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.

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

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

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

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

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) $P_r\{x[n]\} = \theta^{x[n]}(1-\theta)^{1-x[n]} \qquad x[n] = 0, 1$

$$P_r\{\underline{x}\} = \prod_{n=0}^{N-1} \theta^{x[n]}(1-\theta)^{1-x[n]}$$

$$= \underbrace{\theta^{\sum_n x[n]}(1-\theta)^{N-\sum_n x[n]}}_{g(T(\underline{x}), \theta)} \cdot \underbrace{1}_{h(\underline{x})}$$

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

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

10) At high SNR we ignore the noise so that

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

$$= \sum_n \frac{A}{2}\left[\cos\phi + \cos(4\pi f_0 n + \phi)\right]$$

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

Also,

$$T_2(\underline{x}) \approx \sum_n A\cos(2\pi f_0 n + \phi)\sin 2\pi f_0 n$$

$$= \sum_n \frac{A}{2}\left[-\sin\phi + \sin(4\pi f_0 n + \phi)\right]$$

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

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

$$= \phi$$

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 <u>not</u> true that

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

$$\underset{\substack{\uparrow \\ \text{incorrect}}}{=} -\arctan \frac{E[T_2(\underline{x})]}{E[T_1(\underline{x})]}$$

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

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

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

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

$$= \underbrace{\frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\left[N\left(\frac{\theta-1}{2}\right)^2 - (\theta-1)\sum_n x(n)\right]}}_{g(T(\underline{x}),\theta)} \underbrace{e^{-\frac{1}{2\sigma^2}\sum_n x^2(n)}}_{h(\underline{x})}$$

Thus $\sum_{n=0}^{N-1} x(n)$ is a sufficient statistic for θ. To make it unbiased

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

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

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

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 θ.
Examining $\bar{x}^3$ we have

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

since all odd-order moments are zero.

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

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

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

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.

13) $p(\underline{x};\theta) = e^{-(\sum_n x[n] - \theta)} u(\min x[n] - \theta)$

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

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

$$\Pr\{T \le \xi\} = 1 - \Pr\{T \ge \xi\}$$
$$= 1 - \Pr\{x[0] \ge \xi, \ldots, x[N-1] \ge \xi\}$$
$$= 1 - \prod_{n=0}^{N-1} \Pr\{x[n] \ge \xi\}$$
$$= 1 - \Pr^N\{x[n] \ge \xi\}$$

$$p_T(\xi) = \frac{d\Pr\{T \le \xi\}}{d\xi} = -N \Pr\{x[n] \ge \xi\}^{N-1} \cdot \frac{d\Pr\{x[n] \ge \xi\}}{d\xi}$$

But
$$\frac{d\Pr\{x[n] \ge \xi\}}{d\xi} = \frac{d[1 - \Pr\{x[n] \le \xi\}]}{d\xi}$$

$$= -\frac{d\Pr\{x[n] \le \xi\}}{d\xi} = -\begin{cases} e^{-(\xi-\theta)} & \xi > \theta \\ 0 & \xi < \theta \end{cases}$$

and
$$\Pr\{x[n] \ge \xi\} = \begin{cases} 1 - \int_\theta^\xi e^{-(x-\theta)}\,dx & \text{for } \xi > \theta \\ 1 & \text{for } \xi < \theta \end{cases}$$

For $\xi > \theta$
$$= 1 + e^{\theta} e^{-x}\Big|_\theta^\xi = 1 + e^{-(\xi-\theta)} - 1$$
$$= e^{-(\xi-\theta)}$$

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

$$= \begin{cases} N e^{-N(\mathfrak{z}-\theta)} & \mathfrak{z} > \theta \\ 0 & \text{otherwise} \end{cases}$$

$$E(T) = \int_\theta^\infty \mathfrak{z} N e^{-N(\mathfrak{z}-\theta)} d\mathfrak{z}$$
$$= N e^{N\theta} \int_\theta^\infty \mathfrak{z} e^{-N\mathfrak{z}} d\mathfrak{z}$$
$$= N e^{N\theta}\left(-\frac{\mathfrak{z}}{N} e^{-N\mathfrak{z}} - \frac{1}{N^2} e^{-N\mathfrak{z}}\right)\Big|_\theta^\infty$$

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

$\Rightarrow \quad \hat{\theta} = T - 1/N = \min x[n] - \frac{1}{N}$ is the MVU estimator.

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

$$= e^{\underbrace{x\mu}_{A(\mu)B(x)} \underbrace{- \frac{1}{2}x^2}_{C(x)} + \underbrace{(-\frac{1}{2}\mu^2 + \ln 1/\sqrt{2\pi})}_{D(\mu)}}$$

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

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

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

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

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

b) $E(x^2) = \int_0^\infty \frac{x^3}{\sigma^2} e^{-\frac{1}{2}x^2/\sigma^2} dx$

$= 2\sigma^2$

$$\Rightarrow \quad \hat{\sigma}^2 = \frac{1}{2N} \sum_n x^2(n)$$

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

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

16) $$p(\underline{x};\underline{\theta}) = \frac{1}{(2\pi)^{N/2} \det^{1/2}(\underline{C})} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^T \underline{C}^{-1}(\underline{x}-\underline{\mu})}$$

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

$$\det(\underline{C}) = \det(\underline{I}) \det(a - \underline{b}^T \underline{I}^{-1} \underline{b})$$
$$= a - \underline{b}^T\underline{b} = N-1+\sigma^2 - (-\underline{1})^T(-\underline{1})$$
$$= N-1+\sigma^2 - (N-1) = \sigma^2$$

$$\underline{C}^{-1} = \begin{bmatrix} (a - \underline{b}^T\underline{b})^{-1} & -(a-\underline{b}^T\underline{b})^{-1}\underline{b}^T \\ -\underline{b}(a-\underline{b}^T\underline{b})^{-1} & \left(\underline{I} - \frac{\underline{b}\,\underline{b}^T}{a}\right)^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sigma^2} & \frac{1}{\sigma^2}\underline{1}^T \\ \frac{1}{\sigma^2}\underline{1} & \left(\underline{I} - \frac{\underline{1}\,\underline{1}^T}{N-1+\sigma^2}\right)^{-1} \end{bmatrix}$$

But $$\left(\underline{I} - \frac{\underline{1}\,\underline{1}^T}{N-1+\sigma^2}\right)^{-1} = \underline{I} + \frac{\frac{\underline{1}\,\underline{1}^T}{N-1+\sigma^2}}{1 - \frac{\underline{1}^T\underline{1}}{N-1+\sigma^2}}$$

$$= \underline{I} + \frac{\underline{1}\underline{1}^T}{\sigma^2}$$

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

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

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

$$= \frac{1}{\sigma^2}\Big((x[0]-N\mu)^2 + (x[0]-N\mu)\,\underline{1}^T\underline{x}' + (x[0]-N(\mu))\,\underline{x}'^T\underline{1} + \sigma^2\underline{x}'^T\underline{x}' + (\underline{x}'^T\underline{1})^2\Big)$$

$$= \frac{1}{\sigma^2}\Big[(x[0]-N\mu + \underline{1}^T\underline{x}')^2 + \sigma^2\underline{x}'^T\underline{x}'\Big]$$

$$= \frac{1}{\sigma^2}\Big[\Big(\sum_{n=0}^{N-1} x[n] - N\mu\Big)^2 + \sigma^2\sum_{n=1}^{N-1} x^2[n]\Big]$$

$$= \frac{N^2}{\sigma^2}(\bar{x}-\mu)^2 + \sum_{n=1}^{N-1} x^2[n]$$

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

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

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

But $\sum_n (x[n] - A\cos 2\pi f_0 n)^2 =$

$$\sum_n x^2[n] - 2A\sum x[n]\cos 2\pi f_0 n + A^2 \sum_n \cos^2 2\pi f_0 n$$

$$p(\underline{x};A) = \underbrace{\frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\left(A^2\sum_n \cos^2 2\pi f_0 n - 2A\sum_n x[n]\cdot\cos 2\pi f_0 n\right)}}_{g(T(\underline{x}),A)} \cdot \underbrace{e^{-\frac{1}{2\sigma^2}\sum_n x^2[n]}}_{h(\underline{x})}$$

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

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

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

b) Let $\underline{\theta} = \begin{bmatrix} A \\ \sigma^2 \end{bmatrix}$

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 $T_1(\underline{x})$ 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 $T_2(\underline{x})$ unbiased:

$$E(T_2(\underline{x})) = \sum_n E(x^2[n])$$

$$= \sum_n E\left[(A\cos 2\pi f_0 n + w[n])^2\right]$$

$$= \sum_n A^2\cos^2 2\pi f_0 n + N\sigma^2$$

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

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

$$= \frac{\sum_m \sum_n E(x[m]x[n])\cos 2\pi f_0 m \cos 2\pi f_0 n}{\left(\sum_n \cos^2 2\pi f_0 n\right)^2}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

Alternatively, for the PDF to be nonzero

$\min x(n) \ge \theta_1$, $\max x(n) \le \theta_2$ so that

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

$$\cdot \underbrace{1}_{h(\underline{x})}$$

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

19) $(\underline{x}-\underline{H}\hat{\underline{\theta}})^T(\underline{x}-\underline{H}\hat{\underline{\theta}}) + (\underline{\theta}-\hat{\underline{\theta}})^T\underline{H}^T\underline{H}(\underline{\theta}-\hat{\underline{\theta}})$

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

$\quad + (\underline{\theta}-(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{x})^T\underline{H}^T\underline{H}(\underline{\theta}-(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{x})$

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

$\quad + \underline{\theta}^T\underline{H}^T\underline{H}\underline{\theta} - \underline{\theta}^T\underline{H}^T\underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{x}$

$\quad - \underline{x}^T\underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{H}\underline{\theta} + \underline{x}^T\underline{H}^T(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{x}$

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

$\quad + \underline{x}^T\underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{x} = (\underline{x}-\underline{H}\underline{\theta})^T(\underline{x}-\underline{H}\underline{\theta})$

$$p(\underline{x};\underline{\theta}) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}(\underline{x}-\underline{H}\underline{\theta})^T(\underline{x}-\underline{H}\underline{\theta})}$$

$$= \underbrace{\frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}(\underline{\theta}-\hat{\underline{\theta}})^T\underline{H}^T\underline{H}(\underline{\theta}-\hat{\underline{\theta}})}}_{g(T(\underline{x}),\underline{\theta})} \underbrace{e^{-\frac{1}{2\sigma^2}(\underline{x}-\underline{H}\hat{\underline{\theta}})^T(\underline{x}-\underline{H}\hat{\underline{\theta}})}}_{h(\underline{x})}$$

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

Chapter 6

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

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

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

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

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

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

2) $\mathrm{var}(\hat{A}) = \dfrac{1}{\sum_{n=0}^{N-1} 1/\sigma_n^2}$

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

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

If $\sigma_n^2 = (n+1)^2$, as $N \to \infty$ $\sum_{n=0}^{N-1} \frac{1}{(n+1)^2} \to$ Constant
$\Rightarrow \mathrm{var}(\hat{A}) \not\to 0$.
In this case the noise samples have such a large variance that the estimator

variance does not go to zero.

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

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

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

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

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

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

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

But $\underline{B}^T \underline{1} = \dfrac{\begin{bmatrix} 1-\rho \\ 1-\rho \end{bmatrix}}{1-\rho^2} \Rightarrow \underline{1}^T \underline{B} \underline{x}_i = \underline{x}_i^T \underline{B}^T \underline{1}$

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

$$= \frac{[\underline{x}_i]_1 + [\underline{x}_i]_2}{1+\rho}$$

$$\underline{1}^T \underline{C}^{-1} \underline{x} = \frac{1}{\sigma^2} \frac{1}{1+\rho} \sum_{i=1}^{N/2} [\underline{x}_i]_1 + [\underline{x}_i]_2$$

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

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

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

Since the subvectors $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_{N/2}$ 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 σ^2/N for $\rho = 0$ (our usual case), $2\sigma^2/N$ for $\rho \to 1$ since the samples of each subvector are equal and hence we have only $N/2$ uncorrelated samples, and $\to 0$ for $\rho \to -1$ since then the noise samples cancel (see Probs. 3.9 and 4.11).

4) In either case we have the model

$$\underline{x} = \underline{1}\mu + \underline{w} \quad \text{where } E(\underline{w}) = \underline{0}$$
$$\text{and } E(\underline{w}\underline{w}^T) = \mathrm{var}(w[n])\, \underline{I}$$

The BLUE for each case is

$$\hat{\mu} = \frac{\underline{1}^T\underline{C}^{-1}\underline{x}}{\underline{1}^T\underline{C}^{-1}\underline{1}} = \frac{\underline{1}^T\underline{x}}{\underline{1}^T\underline{1}} = \bar{x}$$

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

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

Let $y = \ln x \quad dy/dx = 1/x \, dx \Rightarrow$
$dx = e^y dy$

$$E(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\theta)^2} e^y dy$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y^2 - 2y(\theta+1) + (\theta+1)^2)} \cdot e^{-\frac{1}{2}(\theta^2 - (\theta+1)^2)} dy$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-(\theta+1))^2} dy \cdot e^{-\frac{1}{2}(-2\theta-1)}$$

$$= e^{\theta + 1/2}$$

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

$$p(y) = \frac{p(x(y))}{|dy/dx|} = \frac{\frac{1}{\sqrt{2\pi}\, e^y} e^{-\frac{1}{2}(y-\theta)^2}}{e^{-y}}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\theta)^2}$$

$\Rightarrow y \sim N(\theta, 1)$

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

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

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

$$E(\hat{\theta}) = E\left[\sum_n a_n x[n] + b \right]$$

$$= \sum_n a_n (\theta s[n] + \beta) + b$$

$$= \theta$$

$$\Rightarrow \quad \sum_n a_n s[n] = 1, \quad \beta \sum_n a_n = -b$$

or $\quad \hat{\theta} = \sum_n a_n x[n] - \beta \sum_n a_n$

$$= \sum_n a_n (x[n] - \beta)$$

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

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

and since the covariance for $\underline{x} - \beta \underline{1}$ is $\underline{C}$,

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

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

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

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

$$\mathrm{var}(\hat{A}) = \frac{1}{\underline{s}^T \underline{C}^{-1} \underline{s}} = \frac{\lambda}{\underline{s}^T \underline{s}}$$

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

8) $$\mathrm{var}(\hat{A}) = \frac{1}{\underline{s}^T \underline{C}^{-1} \underline{s}}$$

$$= \frac{1}{\left(\sum_i \alpha_i \underline{v}_i\right)^T \underline{C}^{-1} \left(\sum_j \alpha_j \underline{v}_j\right)}$$

$$= \frac{1}{\sum_i \sum_j \alpha_i \alpha_j \underbrace{\underline{v}_i^T \underline{C}^{-1} \underline{v}_j}_{\underbrace{\underline{v}_i^T \, 1/\lambda_j \, \underline{v}_j}_{1/\lambda_j \, \delta_{ij}}}}$$

$$= \frac{1}{\sum_i \alpha_i^2 / \lambda_i}$$

$$\mathcal{E} = \underline{s}^T \underline{s} = \left(\sum_i \alpha_i \underline{v}_i\right)^T \left(\sum_j \alpha_j \underline{v}_j\right)$$

$$= \sum_i \sum_j \alpha_i \alpha_j \underbrace{\underline{v}_i^T \underline{v}_j}_{\delta_{ij}} = \sum_i \alpha_i^2$$

Must minimize $\dfrac{1}{\sum_i \alpha_i^2/\lambda_i}$ subject to constraint $\sum_i \alpha_i^2 = \varepsilon_0$. Equivalently we must maximize $\sum_i \alpha_i^2/\lambda_i$. Using Lagrangian multipliers

$$F = \sum_i \alpha_i^2/\lambda_i + \lambda\left(\sum_i \alpha_i^2 - \varepsilon_0\right)$$

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

$$\Rightarrow \alpha_k = 0 \quad \text{or}$$
$$\lambda = -1/\lambda_k \quad \text{all } k$$

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

$$\alpha_k = 0 \quad \text{except for } k = j$$
$$\lambda = -1/\lambda_j \quad \text{and } \alpha_j \neq 0.$$

Hence, $\underline{s} = \alpha_j \underline{v}_j$. To determine which eigenvector to use

$$\mathrm{var}(\hat{A}) = \frac{1}{\alpha_j^2/\lambda_j} = \lambda_j/\alpha_j^2$$

and since $\mathcal{E}_0 = \alpha_j^2$, $\text{var}(\hat{A}) = \lambda_j / \mathcal{E}_0$ so that λ_j should be the minimum eigenvalue. Hence, the optimal signal is

$$\underline{s} = c\,\underline{V}_{MIN} = \sqrt{\mathcal{E}_0}\,\underline{V}_{MIN}$$

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

9) $\theta = A$

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

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

$$= \frac{\sum_{n=0}^{N-1} x(n) \cos 2\pi f_1 n}{\sum_{n=0}^{N-1} \cos^2 2\pi f_1 n}$$

This is a scaled Fourier coefficient since

$$\hat{A} = \frac{N/2}{\sum_n \cos^2 2\pi f_1 n} \underbrace{\frac{2}{N} \sum_n x(n) \cos 2\pi f_1 n}_{\text{Fourier coefficient}}$$

and for large N and f_1 not near 0 or 1/2 the scale factor is one.

$$\mathrm{var}(\hat{A}) = \frac{1}{\underline{s}^T \underline{C}^{-1} \underline{s}} = \frac{\sigma^2}{\sum_{n=0}^{N-1} s^2[n]}$$

$$= \frac{\sigma^2}{\sum_{n=0}^{N-1} \cos^2 2\pi f_1 n} \geq \frac{\sigma^2}{N}$$

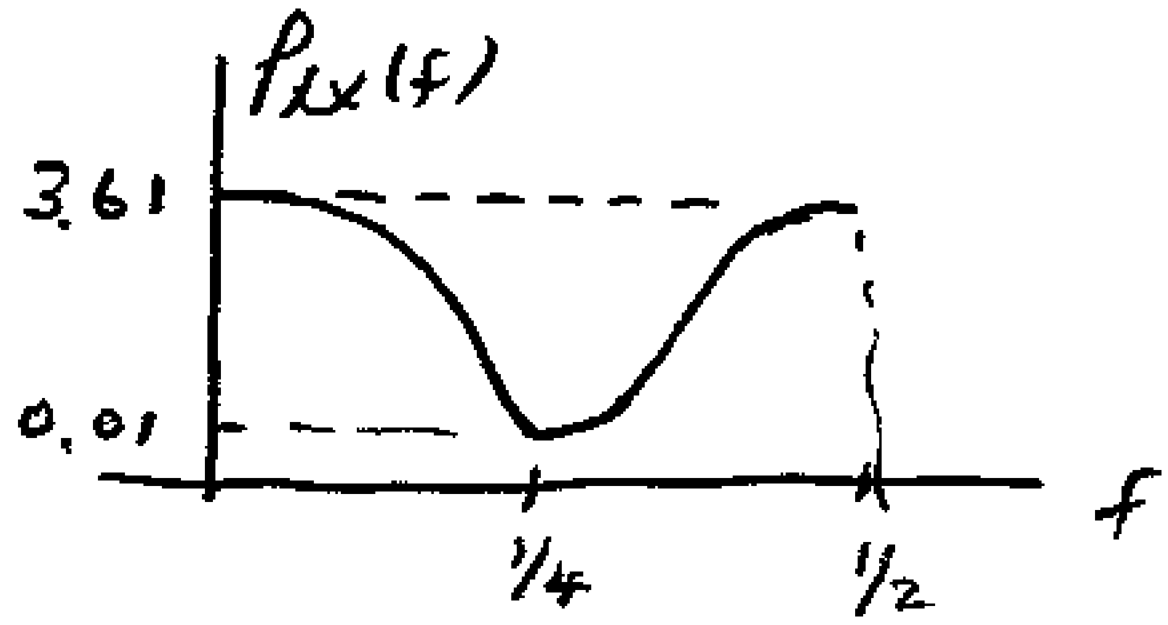
Since $\sum_{n=0}^{N-1} \cos^2 2\pi f_1 n \leq N$

$\Rightarrow$ variance is minimized if $f_1 = 0$. Note that this choice maximizes the signal energy.

10) $$P_{xx}(f) = r_{xx}[0] + r_{xx}[2] e^{-j4\pi f} + r_{xx}[2] e^{j4\pi f}$$

$$= 1.81 + 2(0.9)\cos 4\pi f$$

$$= 1.81 + 1.8 \cos 4\pi f$$



Need to minimize $\frac{1}{\underline{s}^T \underline{C}^{-1} \underline{s}} = \mathrm{var}(\hat{A})$

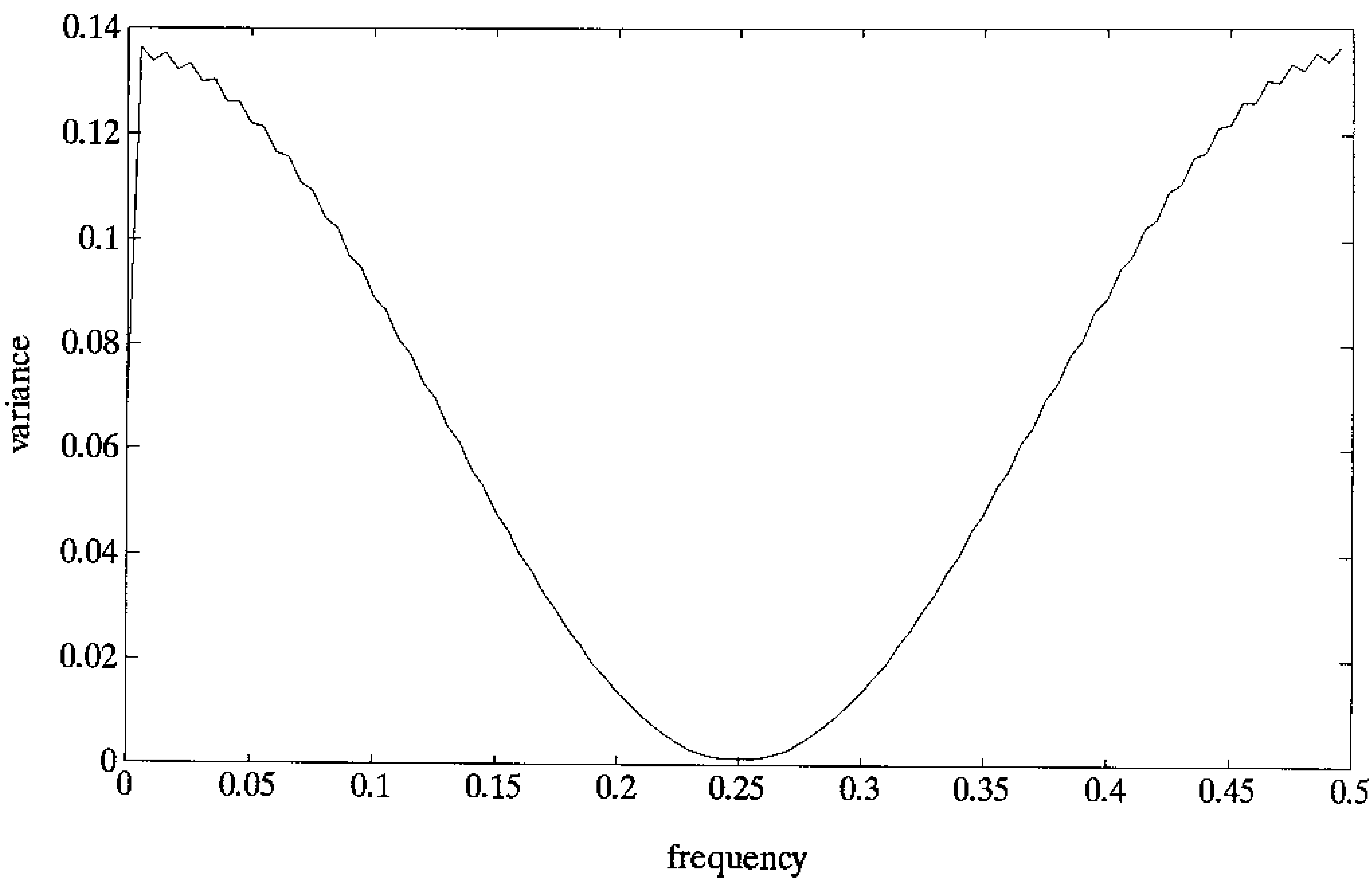
where $\underline{s} = [1 \ \cos 2\pi f_1 \ \ldots \ \cos 2\pi f_1 (N-1)]^T$

and

$$\underline{C} = \begin{bmatrix} r_{xx}[0] & r_{xx}[1] & \ldots & r_{xx}[N-1] \\ r_{xx}[1] & r_{xx}[0] & & \\ & & \ddots & \\ & & & r_{xx}[0] \end{bmatrix}$$

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

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



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

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

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

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

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

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

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

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

The noise power at the output is

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

$$= \frac{1}{\underline{1}^T \underline{C}^{-1} \underline{1}}$$

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

$$\begin{aligned}
\mathrm{var}(\hat{A}) &= E\left[(\hat{A} - E(\hat{A}))^2 \right] \\
&= E\left[\left(\sum_k h[k]\, x[N-1-k] - \sum_k h[k] A \right)^2 \right]
\end{aligned}$$

since $\sum_k h[k] = 1$

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

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

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

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

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

$$= (\underline{B}^{-1^T} \underline{H}^T \underline{C}^{-1} \underline{H} \underline{B}^{-1})^{-1} \underline{B}^{-1^T} \underline{H}^T \underline{C}^{-1} (\underline{x} + \underline{H}\underline{B}^{-1}\underline{b})$$

$$= \underline{B} (\underline{H}^T \underline{C}^{-1} \underline{H})^{-1} \underline{H}^T \underline{C}^{-1} (\underline{x} + \underline{H}\underline{B}^{-1}\underline{b})$$

$$= \underline{B}\hat{\underline{\theta}} + \underline{B}\underline{B}^{-1}\underline{b} = \underline{B}\hat{\underline{\theta}} + \underline{b}$$

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

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

Using (4.3) we have

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

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

14) Since $p(w[n])$ is even, $E(w[n]) = 0$.

$$\text{var}(w[n]) = E(w^2[n])$$

$$= \int_{-\infty}^{\infty} w^2 \left(\frac{1-\epsilon}{\sqrt{2\pi\sigma_B^2}} e^{-\frac{1}{2} w^2/\sigma_B^2} + \frac{\epsilon}{\sqrt{2\pi\sigma_I^2}} e^{-\frac{1}{2} w^2/\sigma_I^2} \right) dw$$

$$= (1-\epsilon)\int_{-\infty}^{\infty} w^2 \frac{1}{\sqrt{2\pi\sigma_B^2}} e^{-\frac{1}{2}w^2/\sigma_B^2}\,dw$$

$$+ \epsilon \int_{-\infty}^{\infty} w^2 \frac{1}{\sqrt{2\pi\sigma_I^2}} e^{-\frac{1}{2}w^2/\sigma_I^2}\,dw$$

$$= (1-\epsilon)\sigma_B^2 + \epsilon\sigma_I^2$$

The BLUE of σ^2 is $\hat{\sigma}^2 = 1/N \sum_{n=0}^{N-1} w^2(n)$ (see Section 6.3). This is because the $w(n)$'s are independent and thus $y(n) = w^2(n)$'s are independent. The mean is $E(y(n)) = \sigma^2$ and covariance matrix is $\underline{C} = \mathrm{var}(y(n))\,\underline{I}$ so that

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

Now using results from Prob. 6.12

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

$$\Rightarrow \quad \hat{\sigma}_I^2 = \frac{\hat{\sigma}^2 - (1-\epsilon)\sigma_B^2}{\epsilon}$$

$$= \frac{\frac{1}{N}\sum_{n=0}^{N-1} w^2(n) - (1-\epsilon)\sigma_B^2}{\epsilon}$$

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

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

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

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

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

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

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

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

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

If $\underline{C} = \underline{I}$, $\hat{C} = \begin{bmatrix} 1 & 0 \\ 0 & \alpha \end{bmatrix} \Rightarrow \hat{\underline{C}}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1/\alpha \end{bmatrix}$

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

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

Clearly, by the minimum variance property of the BLUE

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

Let $$\mathfrak{Z} = \frac{\mathrm{var}(\hat{A})}{\mathrm{var}(\hat{A})_{MIN}} = \frac{2(1 + 1/\alpha^2)}{(1 + 1/\alpha)^2}$$

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

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

Chapter 7

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

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

$$\frac{\partial^2 \ln p}{\partial A^2} = \frac{N}{2A^2} - \frac{1}{A^2}\sum_n x[n] - \frac{1}{A^2}\sum_n (x[n]-A)$$
$$- \frac{1}{A^3}\sum (x[n]-A)^2$$

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

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

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

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

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

Even as $N \to \infty$, $\bar{x}$ 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) $p(\underline{x};\mu) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x[n]-\mu)^2}$

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

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

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

(This is just a DC level, μ, in WGN).

b) $p(\underline{x};\lambda) = \begin{cases} \lambda^N e^{-\lambda \sum_n x[n]} & \text{all } x[n] > 0 \\ 0 & \text{otherwise} \end{cases}$

Assuming all $x[n] > 0$ we have

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

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

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

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

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

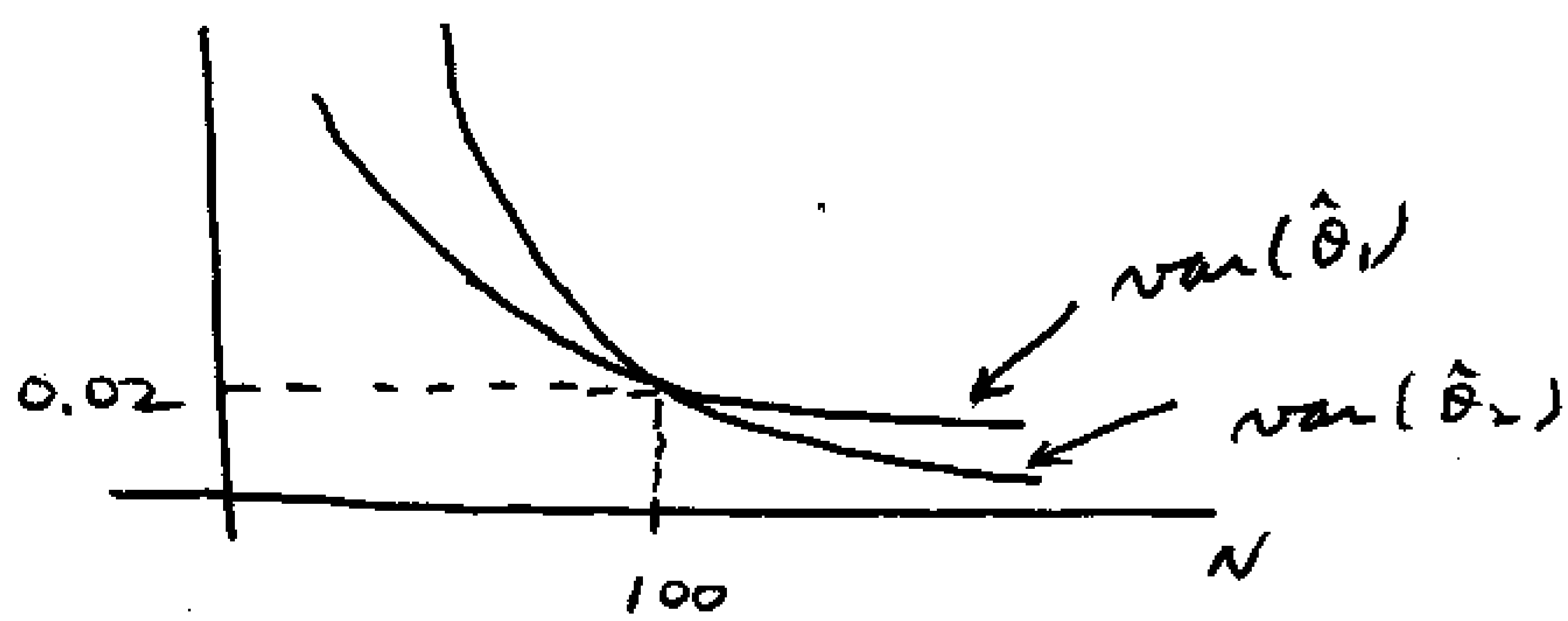
$$\frac{d \ln p}{d\lambda} = N/\lambda - N\bar{x}$$

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

$\Rightarrow$ $\ln p$ is concave function

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

This result is reasonable since the mean of x is $1/\lambda$.

4) 

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

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

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

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

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

$\Rightarrow$ $\bar{x}$ is consistent

6) Linearizing about the true value of θ, ie, θ_0 yields

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

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

$$Pr\{|\hat{\alpha}-\alpha|>\epsilon\} = Pr\left\{\left|\left.\frac{dg}{d\theta}\right|_{\theta=\theta_0}(\hat{\theta}-\theta)\right|>\epsilon\right\}$$

$$= Pr\left\{|\hat{\theta}-\theta| > \frac{\epsilon}{\left|\left.\frac{dg}{d\theta}\right|_{\theta=\theta_0}\right|}\right\} \to 0 \text{ as } N\to\infty$$

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

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

$$= e^{A(\theta)\sum_n B(x[n]) + \sum_n C(x[n]) + ND(\theta)}$$

To maximize p we must minimize
$$A(\theta)\sum_n B(x[n]) + ND(\theta)$$

Differentiating produces the necessary condition

$$\frac{dA(\theta)}{d\theta}\sum_{n=0}^{N-1} B(x(n)) + N\frac{dD(\theta)}{d\theta} = 0$$

For the PDFs of Prob. 7.2

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Clearly, $p(\underline{x})$ is maximized over θ when θ 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) $$p(\underline{x};A) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_n (x[n]-As[n])^2}$$

Minimize $\sum_n (x[n]-As[n])^2$ to find MLE.

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

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

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

$$= A$$

$$\mathrm{var}(\hat{A}) = \frac{\sum_n \mathrm{var}(x[n]) s^2[n]}{\left(\sum_n s^2[n]\right)^2}$$

$$= \sigma^2 \frac{\sum_n s^2[n]}{\left(\sum_n s^2[n]\right)^2} = \frac{\sigma^2}{\sum_{n=0}^{N-1} s^2[n]}$$

$$= I^{-1}(A) \qquad (\text{see Theorem } 4.1)$$

$\hat{A}$ 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.

11) $$p(\underline{x};\rho) = \prod_{n=0}^{N-1} \frac{1}{2\pi\sqrt{\det(\underline{C})}} e^{-\frac{1}{2}\underline{x}^T(n)\underline{C}^{-1}\underline{x}(n)}$$

$$= \frac{1}{(2\pi)^N[\det(\underline{C})]^{N/2}} e^{-\frac{1}{2}\sum_n \underline{x}^T[n]\underline{C}^{-1}\underline{x}(n)}$$

But $\underline{C} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \Rightarrow \underline{C}^{-1} = \dfrac{\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}}{1-\rho^2}$

$\det(\underline{C}) = 1-\rho^2$

$$\ln p = -N\ln 2\pi - \frac{N}{2}\ln(1-\rho^2) - \frac{1}{2(1-\rho^2)} \underbrace{\sum_n \underline{x}^T[n]\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}\underline{x}[n]}_{Q}$$

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

$$= \sum_n \left[x_1^2(n) + x_2^2(n) - 2\rho\, x_1(n)\, x_2(n)\right]$$

Let $C_{ij} = \sum_n x_i[n]\, x_j(n)$

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

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

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

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

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

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

as $N\to\infty$, $\frac{c_{11}}{N}\to 1$, $\frac{c_{22}}{N}\to 1$

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

for which a solution is $\hat{\rho}=\frac{1}{N}c_{12}$

$$=\frac{1}{N}\sum_{n=0}^{N-1}x_1(n)x_2(n)$$

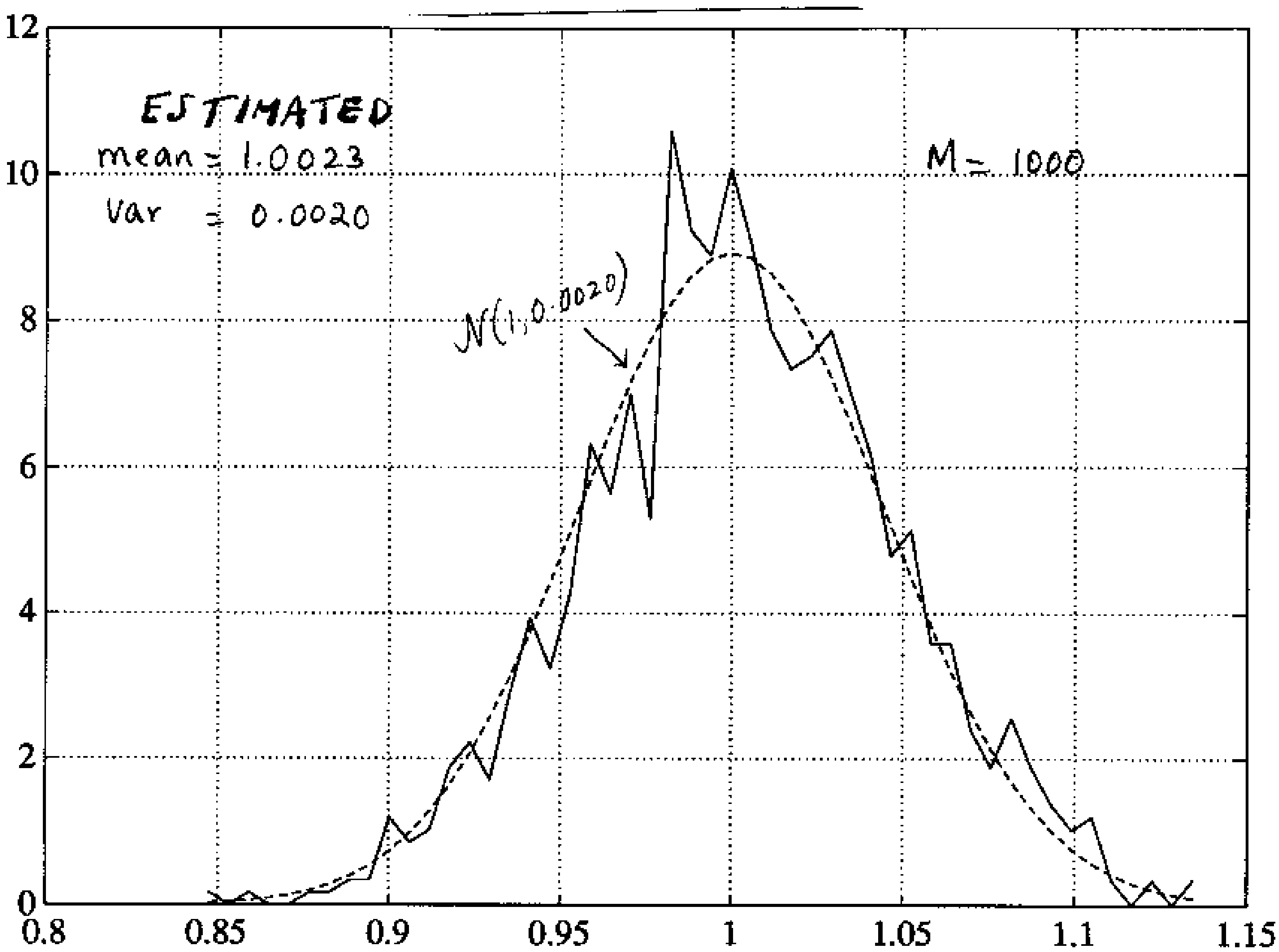
12) To find the MLE we maximize p or equivalently $\ln p$

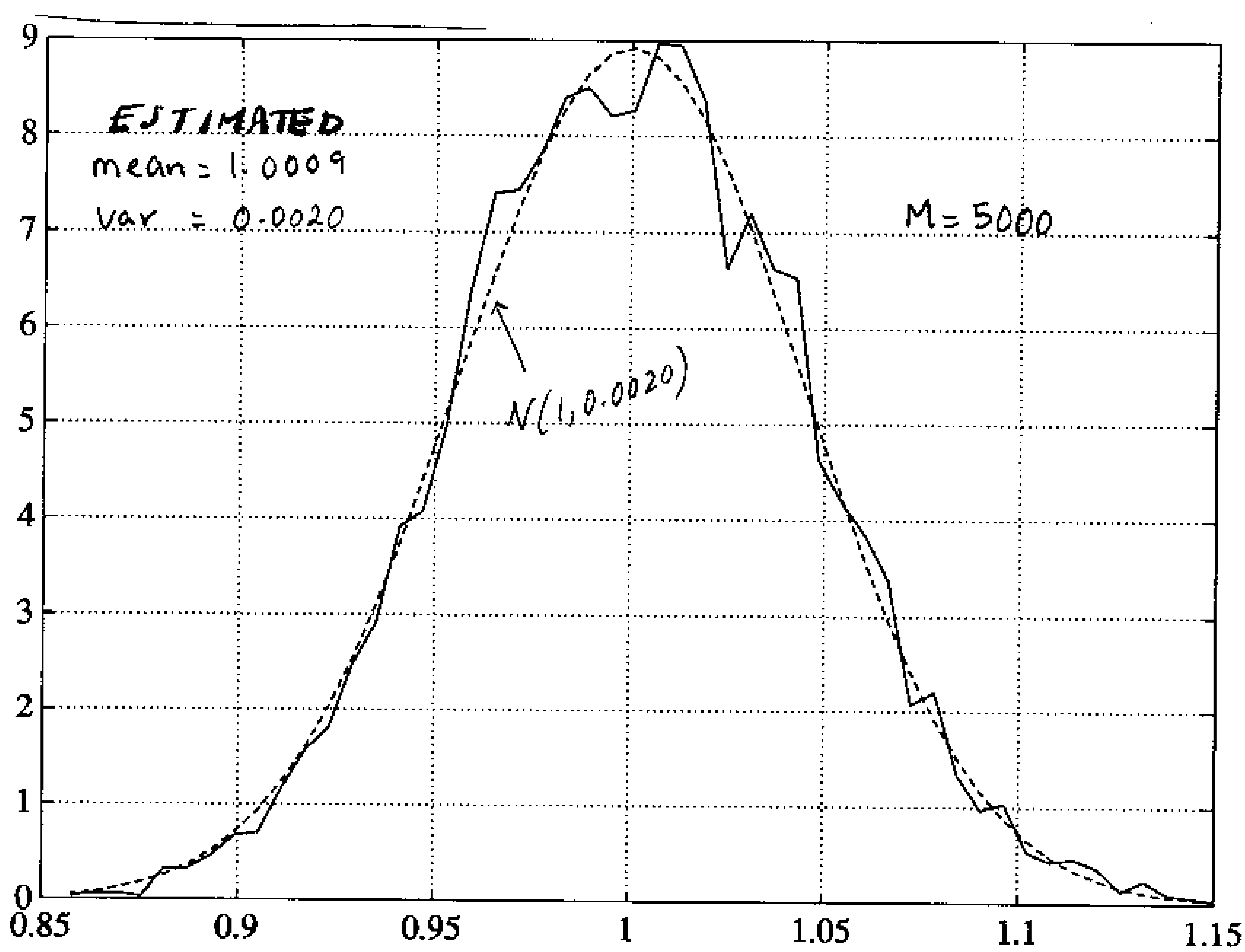
$$\Rightarrow\quad \frac{\partial \ln p}{\partial\theta}=0$$

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

Since $I(\theta)>0$ for all θ (we always assume this - otherwise PDF does not depend on θ), 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 M increases.

14) See plots on next two pages.

15) ϵ_s, ϵ_c are zero mean since $w(n)$ is zero mean.

$$E(\epsilon_s \epsilon_c) = \frac{-4}{N^2A^2} \sum_m \sum_n E(w(m)w(n)) \cdot \sin 2\pi f_0 m \cos 2\pi f_0 n$$

$$= \frac{-4}{N^2A^2} \sigma^2 \sum_n \sin 2\pi f_0 n \cos 2\pi f_0 n$$

$$= \frac{-2\sigma^2}{NA^2} \sum_n \sin 4\pi f_0 n \approx 0$$

Thus, ϵ_s, ϵ_c are independent Gaussian random variables with zero means and variances given as follows:

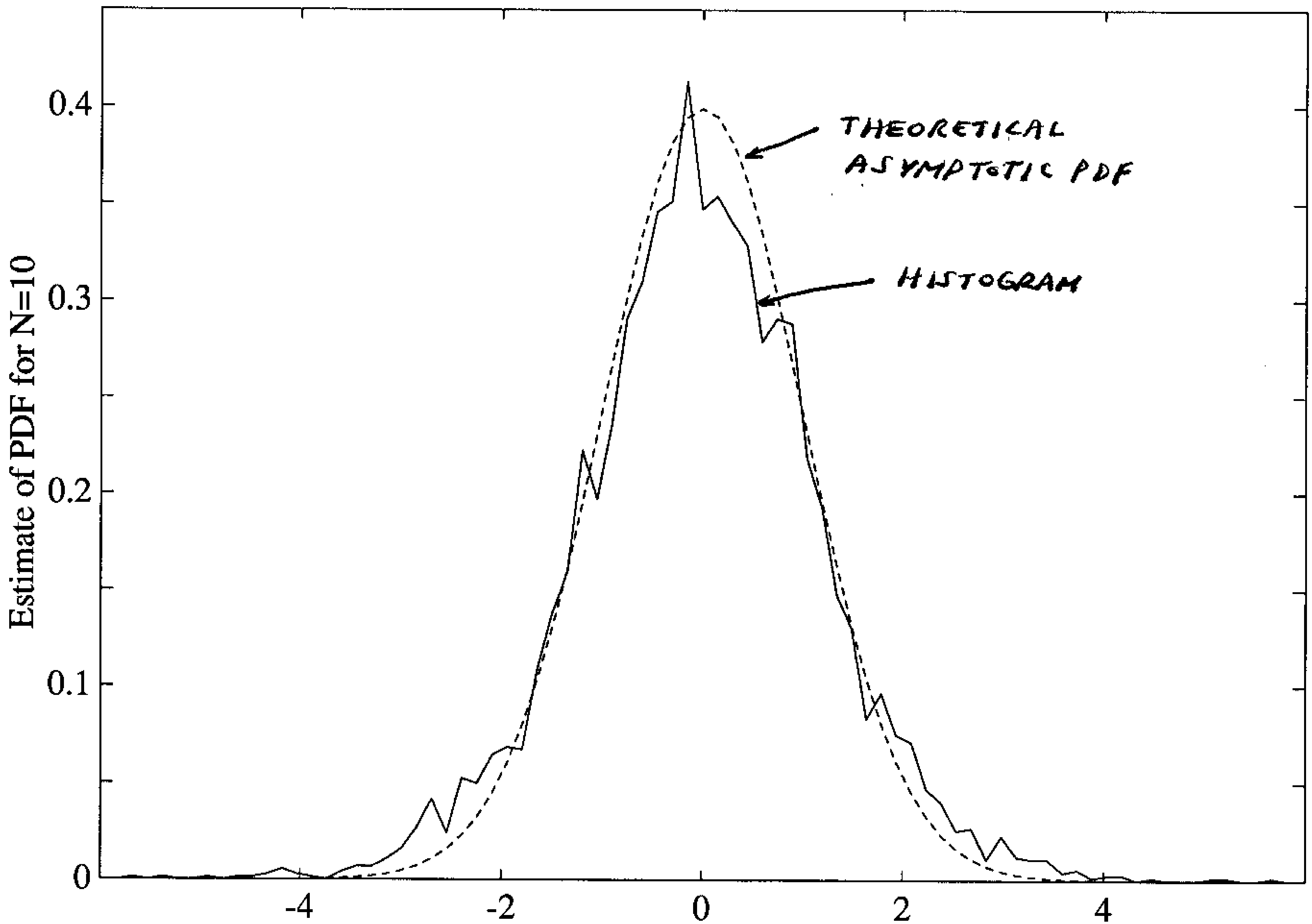
$$\operatorname{var}(\epsilon_s) = E(\epsilon_s^2) = \frac{4}{N^2A^2} \sum_m \sum_n E(w(m)w(n)) \cdot \sin 2\pi f_0 m \sin 2\pi f_0 n$$

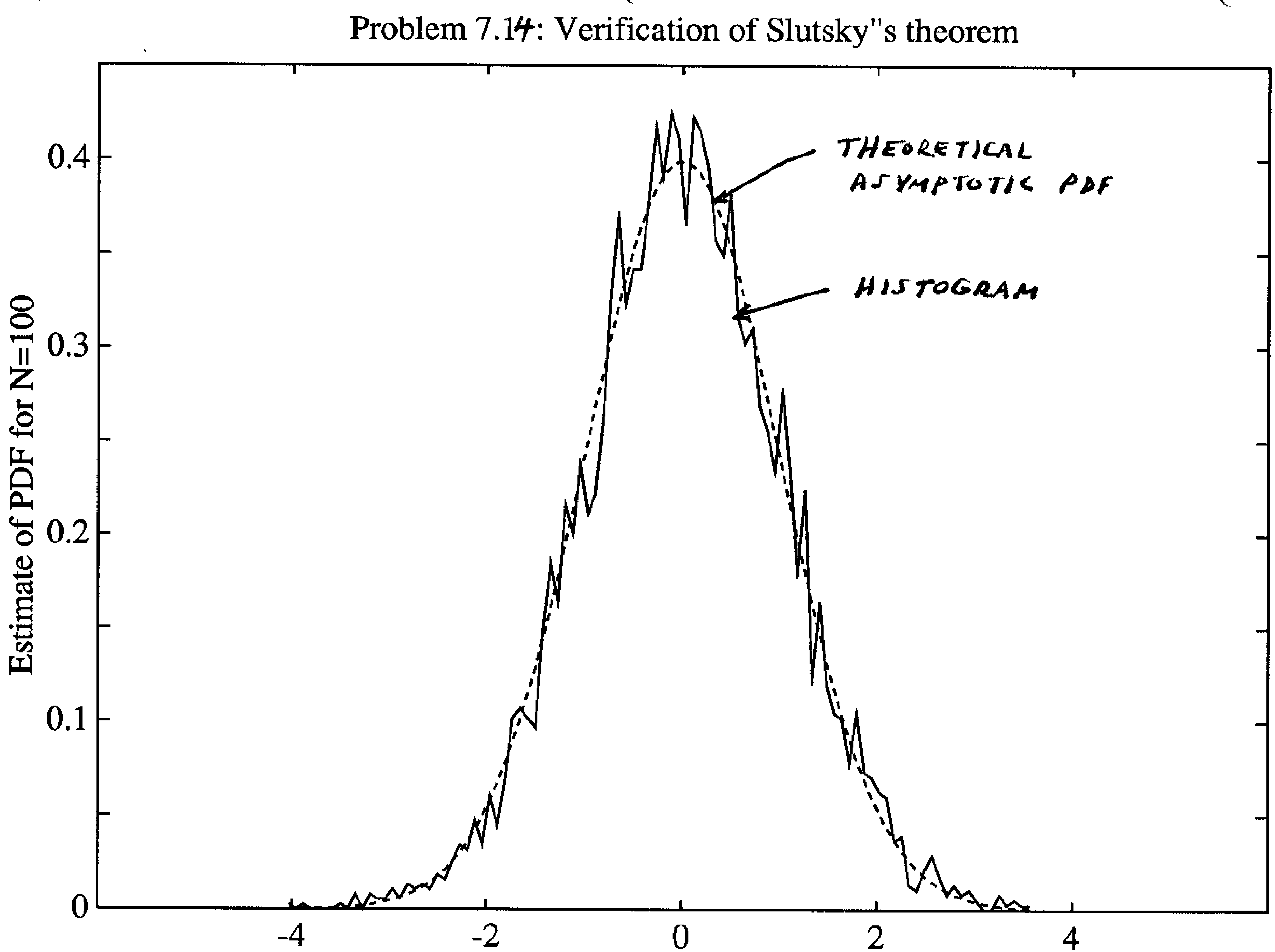
$$= \frac{4\sigma^2}{N^2A^2} \sum_n \sin^2 2\pi f_0 n$$

$$= \frac{4\sigma^2}{N^2A^2} \sum_n \left(\tfrac{1}{2} - \tfrac{1}{2}\cos 4\pi f_0 n\right)$$

$$\approx \frac{4\sigma^2}{N^2A^2}(N/2) = \frac{2\sigma^2}{NA^2}$$

Problem 7.14: Verification of Slutsky''s theorem




Problem 7.14: Verification of Slutsky"s theorem
Estimate of PDF for N=100
0.4
0.3
0.2
0.1
0
-4
-2
0
2
4
THEORETICAL ASYMPTOTIC PDF
HISTOGRAM

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

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

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

$$\left.\frac{\partial g}{\partial \epsilon_s}\right|_{\epsilon_s=\epsilon_c=0} = \frac{\cos^2\phi}{\cos^2\phi+\sin^2\phi} \frac{1}{\cos\phi}$$

$$= \cos\phi$$

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

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

$$\hat{\phi} \simeq \phi + \cos\phi\, \epsilon_s + \sin\phi\, \epsilon_c$$

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

The variance is just the CRLB.

16) $\operatorname{var}(\hat{A}) = \operatorname{var}(x[0])$

$$= \operatorname{var}(w(0))$$

$$= \int_{-\infty}^{\infty} u^2 \frac{1}{2} e^{-|u|} du$$

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

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

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

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

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

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

No, the MLE has variance 2 for all N.

17) Let $\alpha = 1/\theta$ so that by invariance of the MLE, $\hat{\alpha} = 1/\hat{\theta}$. But $\hat{\alpha} = 1/N \sum_{n=0}^{N-1} x^2[n]$ since

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

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

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

Thus, $$\hat{\theta} = \frac{1}{\frac{1}{N}\sum_{n=0}^{N-1} x^2[n]}$$

From Chapter 3

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

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

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

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

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

$$\Rightarrow \hat{\theta} \overset{a}{\sim} N(\theta, 2\theta^2/N)$$

18)

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

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

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

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

To find MLE minimize

$$\sum_n (x[n]-\cos 2\pi f_0 n)^2 = \sum_n x^2[n] - 2\sum_n x[n]\cos 2\pi f_0 n + \sum_n \cos^2 2\pi f_0 n$$

$$\approx \sum_n x^2[n] - 2\sum_n x[n]\cos 2\pi f_0 n + N/2$$

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

over $0 < f_0 < 1/2$

Using a Newton-Raphson iteration

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

$$dg/df_0 = -\sum_n 2\pi n\, x[n]\sin 2\pi f_0 n$$

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

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

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_0})	$E(\hat{f}_0)$	$\text{var}(\hat{f}_0)$
0.24	0.25	1.2×10^{-6}
0.26	0.25	1.2×10^{-6}
0.28	0.16	10

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

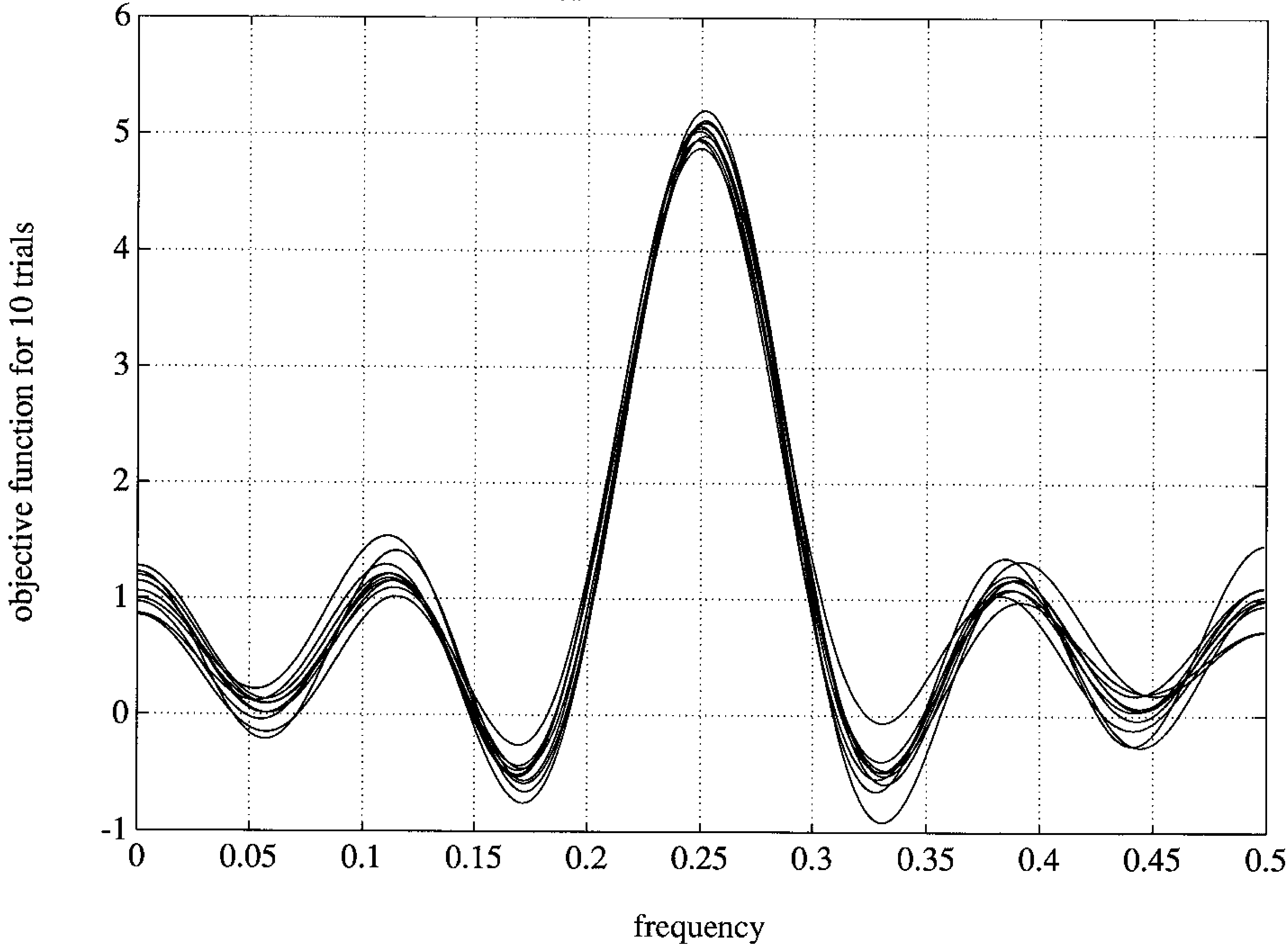
Maximized when $\sum_n (x[n]-s[n])^2$ is minimized. Clearly, then $\hat{s}[n] = x[n]$.

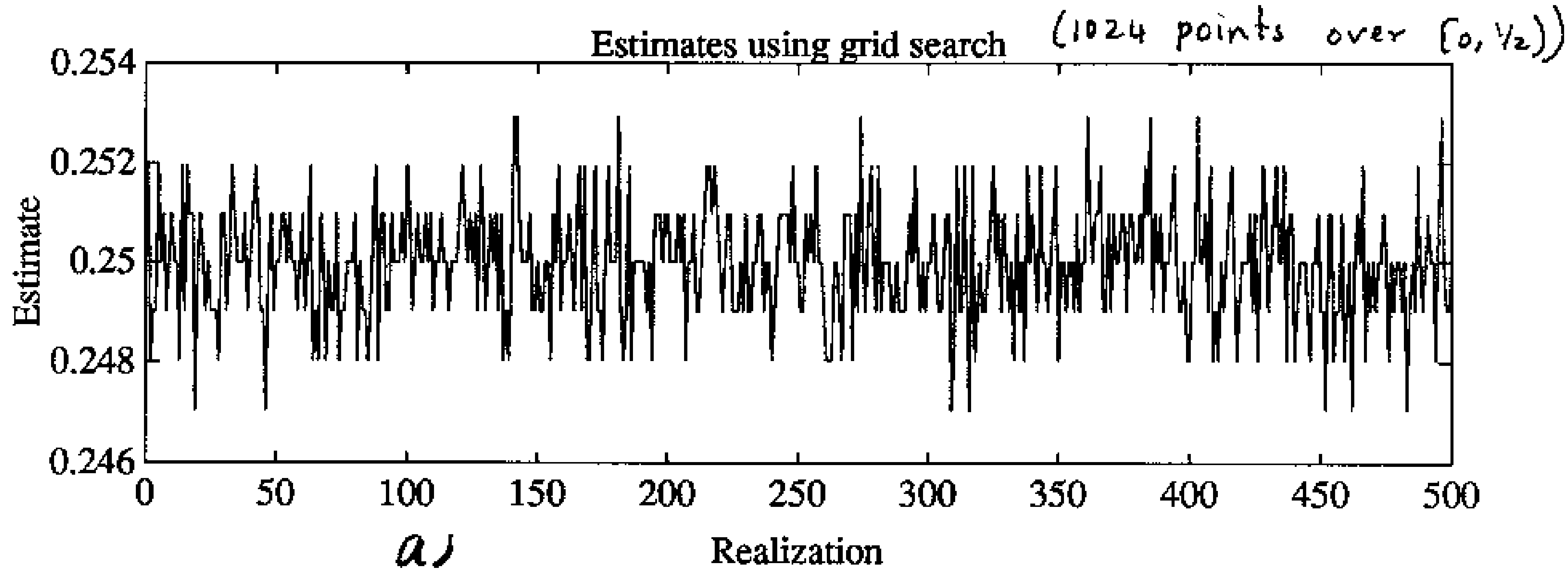
But $\hat{\underline{s}} = \underline{x} \sim \mathcal{N}(\underline{s}, \sigma^2 \underline{I})$ and thus $\hat{\underline{s}}$ is unbiased and Gaussian. To determine if it is efficient we find the CRLB.

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

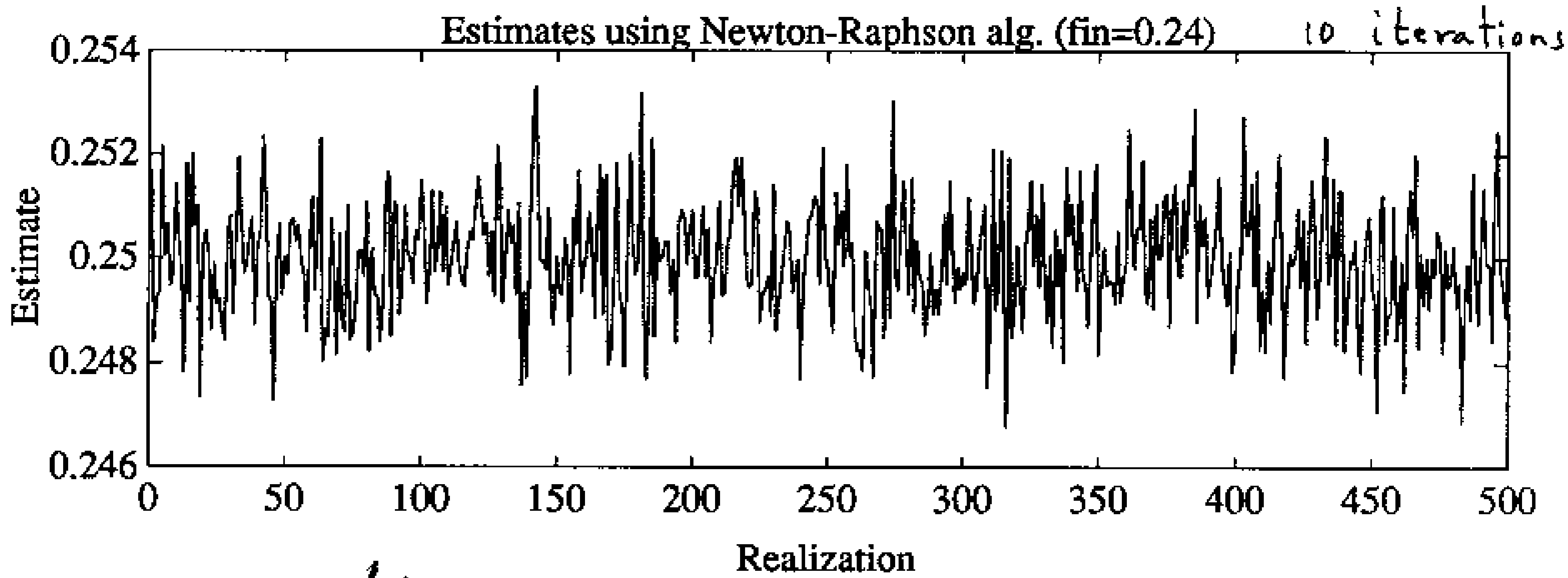
$$\frac{\partial^2 \ln p}{\partial s[k]\partial s[l]} = -\frac{1}{\sigma^2}\delta_{kl}$$

Problem 7.19: Approx. MLE function to be maximized

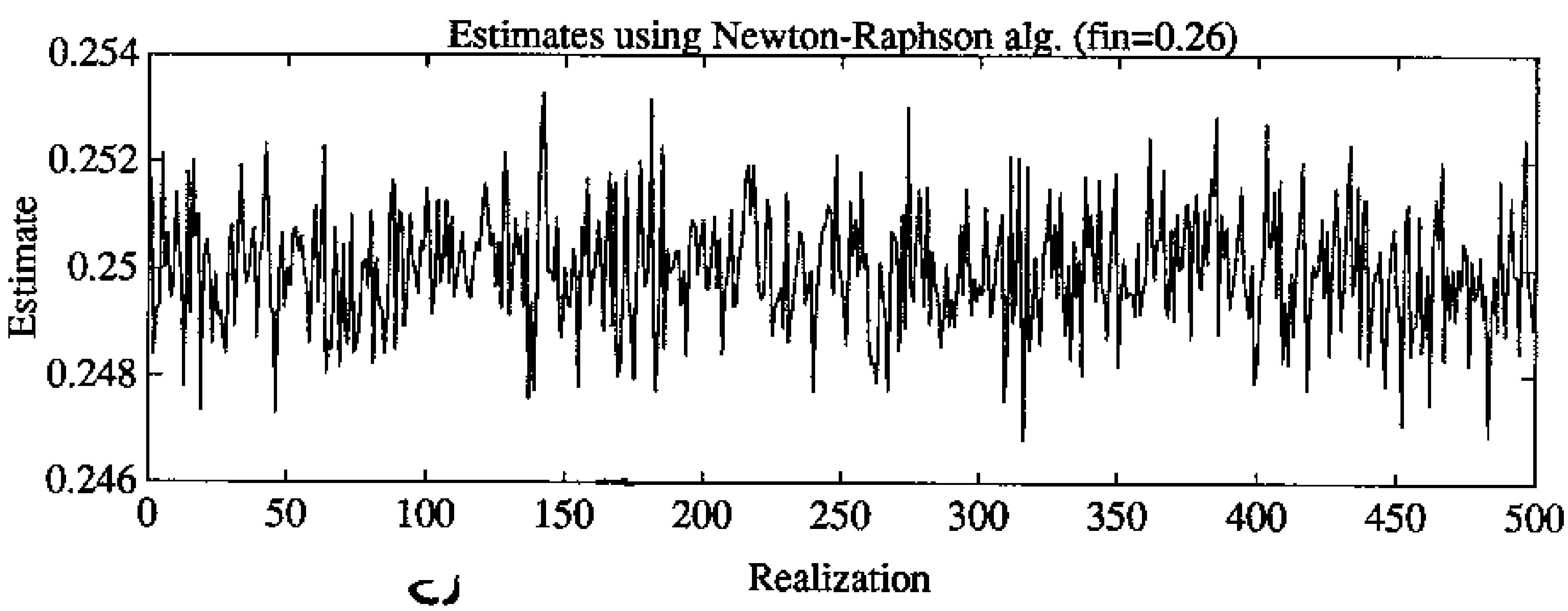



Estimates using grid search
(1024 points over [0, ½))
Estimate
0.254
0.252
0.25
0.248
0.246
0
50
100
150
200
250
300
350
400
450
500
Realization

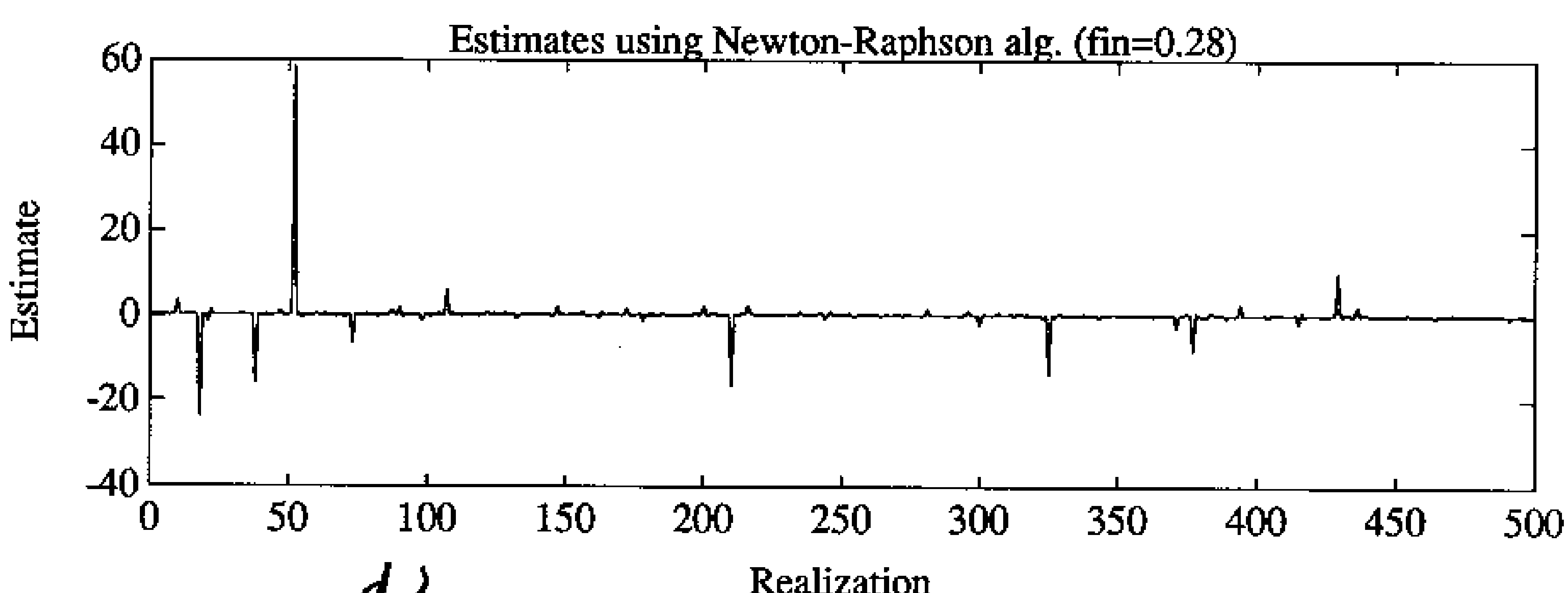
a)


Estimates using Newton-Raphson alg. (fin=0.24)
10 iterations
Estimate
0.254
0.252
0.25
0.248
0.246
0
50
100
150
200
250
300
350
400
450
500
Realization

b)


Estimates using Newton-Raphson alg. (fin=0.26)
Estimate
0.254
0.252
0.25
0.248
0.246
0
50
100
150
200
250
300
350
400
450
500
Realization

c)


Estimates using Newton-Raphson alg. (fin=0.28)
Estimate
60
40
20
0
-20
-40
0
50
100
150
200
250
300
350
400
450
500
Realization

d)

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

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

Hence, $\hat{\underline{s}}$ is efficient. However, it is not consistent since as $N \to \infty$, $\mathrm{var}(\hat{s}[n]) = \sigma^2 \not\to 0$.

21) From Example 7.12 $\hat{A} = \bar{x}$

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

Hence, by the invariance property $\hat{\alpha} = \frac{\hat{A}^2}{\hat{\sigma}^2} = \frac{\bar{x}^2}{\frac{1}{N}\sum_n (x[n]-\bar{x})^2}$

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

$$= 2\,\mathrm{Re} \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln A(f)\, df$$

Now let

$$z = e^{j2\pi f}, \quad dz = j2\pi e^{j2\pi f} df = j2\pi z\, df$$

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

$$= 2\,\mathrm{Re}\left\{ \frac{1}{2\pi j}\oint \ln A(z) \frac{dz}{z} \right\}$$

$$= 2\,\mathrm{Re}\left\{ \mathcal{Z}^{-1}\{\ln A(z)\}\Big|_{n=0} \right\}$$

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 $n=0$ is found from the initial value theorem or

$$\mathcal{Z}^{-1}\{\ln A(z)\}_{n=0} = \lim_{z\to\infty} \ln A(z)$$

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

$$= 0$$

23) We use (7.60) which when differentiated produces

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

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

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

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

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

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

$$= \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] \quad \text{from Prob 7.25}$$

24) See plots on next page. For $f_0 = 0.25$ the peak of the periodogram and that of the exact function $\underline{x}^T \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x}$ 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 $x'[n] \overset{\mathcal{F}}{\leftrightarrow} X'(f)$

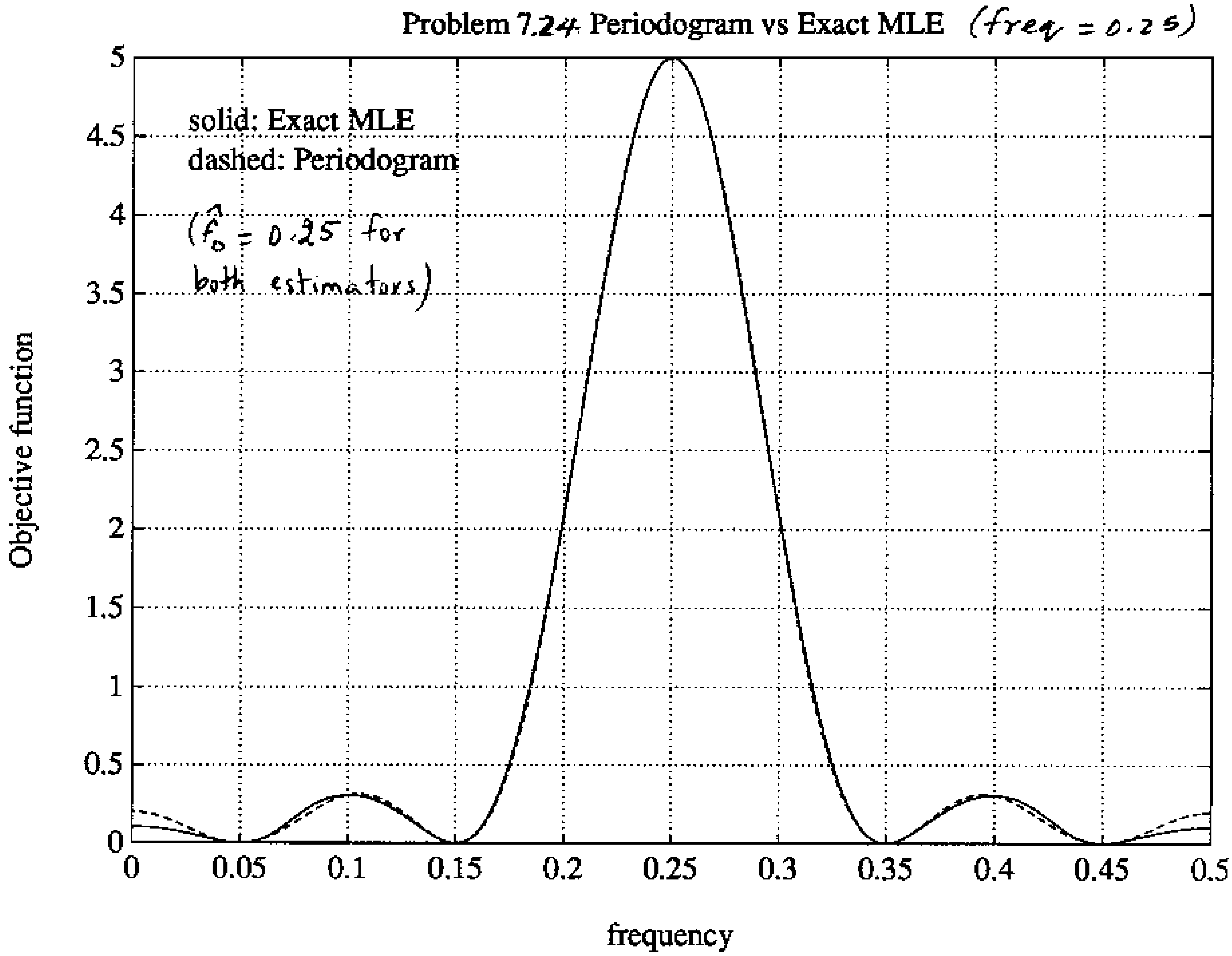
$x'[-n] \overset{\mathcal{F}}{\leftrightarrow} X'(f)^*$ for $x[n]$ real

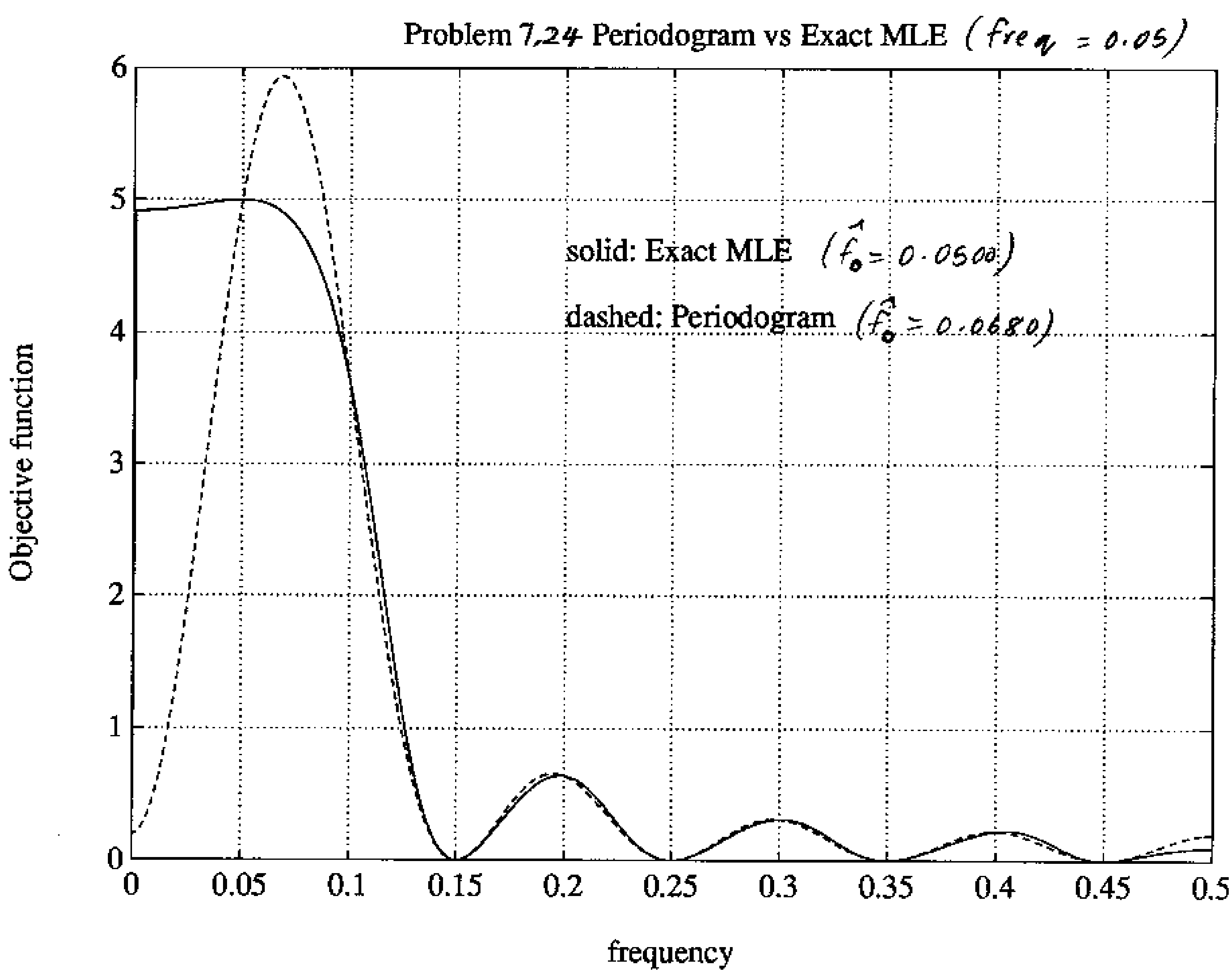
$$\Rightarrow \mathcal{F}^{-1}\{I(f)\} = \frac{1}{N} x'[n] \star x'[-n]$$
$$= \frac{1}{N} \sum_{n=-\infty}^{\infty} x'[-n]\, x'[k-n]$$
$$= \frac{1}{N} \sum_{n=-\infty}^{\infty} x'[n]\, x'[n+k]$$

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 (freq = 0.25)
solid: Exact MLE
dashed: Periodogram
($\hat{f}_0$ = 0.25 for both estimators)
Objective function
frequency
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5


Problem 7.24 Periodogram vs Exact MLE (freq = 0.05)
solid: Exact MLE ($\hat{f}_0$ = 0.0500)
dashed: Periodogram ($\hat{f}_0$ = 0.0680)
Objective function
frequency
0
1
2
3
4
5
6
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5

For $k < 0$

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

Let $l = n+k$

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

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

Combining the results we have our solution.

2b) $\hat{a}(1) = -\hat{r}_{xx}[1]/\hat{r}_{xx}[0]$

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

By invariance the MLE of $P_{xx}(f_0)$ is

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

But

$$C_{\hat{\alpha}} - \frac{\partial g}{\partial \underline{\theta}} \underline{I}^{-1}(\underline{\theta}) \frac{\partial g}{\partial \underline{\theta}}^T \geq 0$$

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

and $g(a[1], \sigma_u^2) = \dfrac{\sigma_u^2}{|1+a[1]e^{-j2\pi f_0}|^2}$

so that

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

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

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

$$\mathrm{var}\left(\hat{P}_{xx}(f_0)\right) \approx \frac{\partial g}{\partial \underline{\theta}} \, I^{-1}(\underline{\theta}) \, \frac{\partial g}{\partial \underline{\theta}}^T$$

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

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

$$= \frac{4 P_{xx}^4(f_0)}{N r_{xx}[0]\, \sigma_u^2} \, \mathrm{Re}^2\left(A(f)\, e^{j2\pi f_0}\right)$$

$$+ \frac{2}{N} P_{xx}^2(f_0)$$

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 $0 \le f_0 \le \frac{1}{2}$.

2) $$J_{MIN} = \sum_{n=0}^{N-1} x^2(n) - N\bar{x}^2 \le \sum_{n=0}^{N-1} x^2(n)$$

also, $$J_{MIN} = \sum_{n=0}^{N-1} (x[n] - \bar{x})^2 \ge 0$$

3) $$J = \sum_{n=0}^{M-1} (x[n] - A)^2 + \sum_{n=M}^{N-1} (x[n] + A)^2$$

$$\frac{\partial J}{\partial A} = -2 \sum_{0}^{M-1} (x[n] - A) + 2 \sum_{M}^{N-1} (x[n] + A) = 0$$

$$-2 \sum_{0}^{M-1} x[n] + 2MA + 2 \sum_{M}^{N-1} x[n] + 2(N-M)A = 0$$

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

$$J_{MIN} = \sum_{0}^{M-1} (x[n] - \hat{A})(x[n] - \hat{A}) + \sum_{M}^{N-1} (x[n] + \hat{A})(x[n] + \hat{A})$$

$$= \sum_{0}^{M-1} x[n](x[n] - \hat{A}) + \sum_{M}^{N-1} x[n](x[n] + \hat{A})$$

using the $\partial J/\partial A = 0$ equation.

$$J_{MIN} = \sum_0^{N-1} x^2[n] - \hat{A}\left(\sum_0^{M-1} x[n] - \sum_M^{N-1} x[n]\right)$$

$$= \sum_0^{N-1} x^2[n] - N\hat{A}^2$$

For $w[n]$ WGN we have

$$E(\hat{A}) = 1/N\,[\,MA - (N-M)A\,] = A$$

$$\text{var}(\hat{A}) = \frac{1}{N^2}\left(\text{var}\left(\sum_0^{M-1} x[n]\right) + \text{var}\left(\sum_M^{N-1} x[n]\right)\right)$$

$$= \frac{1}{N^2}\left(\sum_0^{M-1}\text{var}(x[n]) + \sum_M^{N-1}\text{var}(x[n])\right)$$

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

$\Rightarrow$ $\hat{A} \sim N(A, \sigma^2/N)$ since $\hat{A}$ is a linear function of the $x[n]$'s.

4) See solution for Prob 5,19

5) $$\underline{s} = \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \cos 2\pi f_1 & \cos 2\pi f_2 & \cdots & \cos 2\pi f_p \\ \vdots & & & \\ \cos 2\pi f_1(N-1) & \cos 2\pi f_2(N-1) & \cdots & \cos 2\pi f_p(N-1) \end{bmatrix}}_{\underline{H}} \underbrace{\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_p \end{bmatrix}}_{\underline{\theta}}$$

$$\underline{H}^T\underline{H}\,\hat{\underline{\theta}} = \underline{H}^T\underline{x}$$ are normal equations

If $f_i = i/N$, the column vectors of $\underline{H}$ are orthogonal (see (4.13)). Thus,

$$\underline{H}^T\underline{H} = \frac{N}{2}\,\underline{I}$$

$$\Rightarrow \quad \hat{\underline{\theta}} = \frac{2}{N}\,\underline{H}^T\underline{x} \Rightarrow \hat{A}_i = \frac{2}{N}\sum_{n=0}^{N-1} x(n)\cos 2\pi f_i n$$

$$J_{MIN} = \underline{x}^T(\underline{I} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T)\underline{x}$$

$$= \underline{x}^T(\underline{I} - \frac{2}{N}\,\underline{H}\,\underline{H}^T)\,\underline{x}$$

$$= \underline{x}^T\underline{x} - 2/N\,\|\underline{H}^T\underline{x}\|^2$$

$$= \underline{x}^T\underline{x} - \frac{2}{N}\left(\frac{N}{2}\right)^2\|\hat{\underline{\theta}}\|^2$$

$$= \sum_{n=0}^{N-1} x^2(n) - \frac{N}{2}\sum_{i=1}^{p}\hat{A}_i^2$$

For $w(n)$ WGN the PDF is

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

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

$$C_{\hat{\theta}} = E\left[(\hat{\underline{\theta}} - \underline{\theta})(\hat{\underline{\theta}} - \underline{\theta})^T\right]$$

$$= E\left[(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underbrace{(\underline{x} - \underline{H}\,\underline{\theta})}_{\underline{w}}\underbrace{(\underline{x} - \underline{H}\,\underline{\theta})^T}_{\underline{w}^T}\underline{H}(\underline{H}^T\underline{H})^{-1}\right]$$

$$= (\underline{H}^T\underline{H})^{-1}\underline{H}^T \sigma^2 \underline{I}\, \underline{H}(\underline{H}^T\underline{H})^{-1} = \sigma^2(\underline{H}^T\underline{H})^{-1}$$

or $\underline{C}_{\hat{\theta}} = 2\sigma^2/N\ \underline{I}$

and since $\hat{\underline{\theta}}$ is a linear function of $\underline{x}$, we have a Gaussian PDF or

$$\hat{\underline{\theta}} \sim N(\underline{\theta}, 2\sigma^2/N\ \underline{I})$$

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)
$$E(\hat{\sigma}^2) = \frac{1}{N} E\left[\underline{x}^T(\underline{I} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T)\underline{x}\right]$$
$$= \frac{1}{N}\, \mathrm{tr}\left[(\underline{I} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T)\, E(\underline{x}\,\underline{x}^T)\right]$$

since $E(\underline{x}^T\underline{y}) = \mathrm{tr}(E(\underline{y}\,\underline{x}^T))$

$$E(\hat{\sigma}^2) = \frac{1}{N}\, \mathrm{tr}\left[(\underline{I} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T)\, E((H\theta + w)(H\theta + w)^T)\right]$$

$$= \frac{1}{N}\, \mathrm{tr}\left[(\underline{I} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T)(\underline{H}\underline{\theta}\,\underline{\theta}^T\underline{H}^T + \sigma^2\underline{I})\right]$$

$$= \frac{1}{N}\, \mathrm{tr}\left[\underline{H}\underline{\theta}\,\underline{\theta}^T\underline{H}^T + \sigma^2\underline{I} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{H}\underline{\theta}\,\underline{\theta}^T\underline{H}^T - \sigma^2\underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T\right]$$

$$= \frac{\sigma^2}{N} \, tr \left[\underline{I} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T \right]$$

$$= \sigma^2 - \frac{\sigma^2}{N} \, tr \left[\underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T \right]$$

$$= \sigma^2 - \sigma^2/N \; tr \left[\underbrace{(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{H}}_{\underline{I}\ (p \times p)} \right]$$

$$= \sigma^2 - \sigma^2 p/N = \sigma^2 \frac{N-p}{N}$$

$\hat{\sigma}^2$ is biased. To make it unbiased use

$$\hat{\sigma}^2 = \frac{1}{N-p} J_{MIN}.$$

It is said that we lose p degrees of freedom in estimating $\underline{\theta}$.

8) $$J(A) = \sum_{n=0}^{N-1} \frac{1}{\sigma_n^2} (x(n) - A)^2$$

$$\frac{\partial J}{\partial A} = -2 \sum_n \frac{1}{\sigma_n^2} (x(n) - A) = 0$$

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

$$E(\hat{A}) = \frac{\sum_n E(x(n))/\sigma_n^2}{\sum_n 1/\sigma_n^2} = A$$

$$\mathrm{var}(\hat{A}) = \frac{1}{\left(\sum_n 1/\sigma_n^2\right)^2} \underbrace{\sum_n \mathrm{var}\left(x(n)/\sigma_n^2\right)}$$

$$= \sum_n 1/\sigma_n^4 \, \mathrm{var}(x(n))$$

$$= \sum_n 1/\sigma_n^2$$

$$= \frac{1}{\sum_{n=0}^{N-1} 1/\sigma_n^2}$$

9) $$J = (\underline{x} - \underline{H}\underline{\theta})^T \underline{W} (\underline{x} - \underline{H}\underline{\theta})$$

$$= (\underline{x} - \underline{H}\underline{\theta})^T \underline{D}^T \underline{D} (\underline{x} - \underline{H}\underline{\theta})$$

$$= (\underbrace{\underline{D}\underline{x}}_{\underline{x}'} - \underbrace{\underline{D}\underline{H}}_{\underline{H}'}\underline{\theta})^T (\underline{D}\underline{x} - \underline{D}\underline{H}\underline{\theta})$$

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

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

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

$$J_{MIN} = \underline{x}'^T \left(\underline{I} - \underline{H}' (\underline{H}'^T \underline{H}')^{-1} \underline{H}'^T \right) \underline{x}'$$

$$= \underline{x}^T \underline{D}^T \left(\underline{I} - \underline{D}\underline{H} (\underline{H}^T \underline{D}^T \underline{D} \underline{H})^{-1} \underline{H}^T \underline{D}^T \right) \underline{D}\underline{x}$$

$$= \underline{x}^T \left(\underline{W} - \underline{W}\underline{H} (\underline{H}^T \underline{W} \underline{H})^{-1} \underline{H}^T \underline{W} \right) \underline{x}$$

10) $$\hat{\underline{s}} = \underline{H}\hat{\underline{\theta}} = \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{x} = \underline{P}\underline{x}$$

$$\underline{x} - \hat{\underline{s}} = \underline{x} - \underline{P}\,\underline{x} = (\underline{I} - \underline{P})\,\underline{x}$$

$$\|\hat{\underline{s}}\|^2 + \|\underline{x} - \hat{\underline{s}}\|^2 = \underline{x}^T \underline{P}^T \underline{P}\,\underline{x} + \underline{x}^T (\underline{I} - \underline{P})^T (\underline{I} - \underline{P})\,\underline{x}$$

$$= \underline{x}^T \underline{P}\,\underline{x} + \underline{x}^T (\underline{I} - \underline{P})\,\underline{x} = \underline{x}^T \underline{x} = \|\underline{x}\|^2$$

since $\underline{P}$ and $(\underline{I} - \underline{P})$ are symmetric and idempotent.

11) $\underline{x}_1 = \underline{z}_1 + \underline{z}_1^{\perp} \quad \underline{x}_2 = \underline{z}_2 + \underline{z}_2^{\perp}$

$$l = \underline{x}_1^T \underline{P}\,\underline{x}_2 - \underline{x}_2^T \underline{P}\,\underline{x}_1 = (\underline{z}_1 + \underline{z}_1^{\perp})^T \underline{P}\,(\underline{z}_2 + \underline{z}_2^{\perp}) - (\underline{z}_2 + \underline{z}_2^{\perp})^T \underline{P}\,(\underline{z}_1 + \underline{z}_1^{\perp})$$

$$= (\underline{z}_1 + \underline{z}_1^{\perp})^T \underline{P}\,\underline{z}_2 - (\underline{z}_2 + \underline{z}_2^{\perp})^T \underline{P}\,\underline{z}_1$$

since $\underline{P}\,\underline{z}^{\perp} = \underline{0}$

$$l = (\underline{z}_1 + \underline{z}_1^{\perp})^T \underline{z}_2 - (\underline{z}_2 + \underline{z}_2^{\perp})^T \underline{z}_1$$

since $\underline{P}\,\underline{z} = \underline{z}$

$$l = \underline{z}_1^T \underline{z}_2 - \underline{z}_2^T \underline{z}_1 = 0 \quad \text{since } \underline{z}_i^{\perp T} \underline{z}_j = 0$$

Now let $\underline{x}_1 = \underline{e}_i = [\,0\,0 \ldots 0\,1\,0 \ldots 0\,]^T$

$\underline{x}_2 = \underline{e}_j$ $\qquad\qquad \uparrow$ i^{th} place

$$\underline{e}_i^T \underline{P} \underline{e}_j = \underline{e}_j^T \underline{P} \underline{e}_i$$ from previous result

$$\Rightarrow \quad [\underline{P}]_{ij} = [\underline{P}]_{ji}$$

12) a) $\underline{P}^2 = \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T$

$$= \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T = \underline{P}$$

b) $\underline{x}^T\underline{P}\underline{x} = \underline{x}^T\underline{P}\underline{P}\underline{x} = \underline{x}^T\underline{P}^T\underline{P}\underline{x}$

$$= (\underline{P}\underline{x})^T\underline{P}\underline{x} = \|\underline{P}\underline{x}\|^2 \geq 0$$

c) $\underline{P}\underline{x} = \lambda\underline{x} \Rightarrow \underline{P}\underline{x} = \underline{P}\underline{P}\underline{x} = \underline{P}\lambda\underline{x}$

$$= \lambda(\underline{P}\underline{x}) = \lambda^2\underline{x}$$

$\Rightarrow$ If λ is an eigenvalue, so is λ^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 $tr(\underline{P}) = \sum_{i=1}^{N} \lambda_i$

$$tr(\underline{P}) = tr(\underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T)$$

$$= tr((\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{H})$$

$$= tr(\underline{I}) = p$$

$\Rightarrow \sum_{i=1}^{p} \lambda_i = p \Rightarrow p$ eigenvalues $= 1$

$\Rightarrow$ rank $= p$

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

where $\underline{H} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & N-1 \end{bmatrix}$

$$\underline{H}^T\underline{H} = \begin{bmatrix} N & \sum_n n \\ \sum_n n & \sum_n n^2 \end{bmatrix} = \begin{bmatrix} N & N(N-1)/2 \\ N(N-1)/2 & \frac{N(N-1)(2N-1)}{6} \end{bmatrix}$$

using (3.22)

$$(\underline{H}^T\underline{H})^{-1} = \frac{\begin{bmatrix} \frac{N(N-1)(2N-1)}{6} & -\frac{N(N-1)}{2} \\ -\frac{N(N-1)}{2} & N \end{bmatrix}}{\frac{N^2(N-1)(2N-1)}{6} - \frac{N^2(N-1)^2}{4}}$$

$$\hat{\underline{\theta}} = \begin{bmatrix} \frac{2(2N-1)}{N(N+1)} & \frac{-6}{N(N+1)} \\ \frac{-6}{N(N+1)} & \frac{12}{N(N^2-1)} \end{bmatrix} \begin{bmatrix} \sum_n x(n) \\ \sum_n n\,x(n) \end{bmatrix}$$

which produces (8.23).

14) $\underline{h}'_{k+1} \perp \{\underline{h}_1, \underline{h}_2, \ldots, \underline{h}_k\}$

Let $\underline{P}_k = \underline{H}_k(\underline{H}_k^T\underline{H}_k)^{-1}\underline{H}_k^T$

where $\underline{H}_k = [\underline{h}_1 \ldots \underline{h}_k]$

Then, $\underline{h}'_{k+1} = \underline{h}_{k+1} - \underline{P}_k\underline{h}_{k+1}$ (See Fig 8.8)

$= \underline{P}_k^{\perp}\underline{h}_{k+1}$

Since $\underline{h}'_{k+1} \perp \{\underline{h}_1, \dots, \underline{h}_k\}$ it is also $\perp$ to $\hat{\underline{s}}_k \Rightarrow$ we can project $\underline{x}$ onto $\underline{h}'_{k+1}$ and add the result to $\hat{\underline{s}}_k$.

$$\alpha \underline{h}'_{k+1} = \frac{\underline{x}^T \underline{h}'_{k+1}}{\| \underline{h}'_{k+1} \|} \frac{\underline{h}'_{k+1}}{\| \underline{h}'_{k+1} \|}$$

Since $\dfrac{\underline{h}'_{k+1}}{\| \underline{h}'_{k+1} \|}$ is a unit vector in the $\underline{h}'_{k+1}$ direction

$$\text{or } \alpha = \frac{\underline{x}^T \underline{h}'_{k+1}}{\| \underline{h}'_{k+1} \|^2} = \frac{\underline{x}^T \underline{P}_k^\perp \underline{h}_{k+1}}{\| \underline{P}_k^\perp \underline{h}_{k+1} \|^2}$$

$$\hat{\underline{s}}_{k+1} = \underline{H}_k \hat{\underline{\theta}}_k + \frac{\underline{x}^T \underline{P}_k^\perp \underline{h}_{k+1}}{\underline{h}_{k+1}^T \underline{P}_k^\perp \underline{h}_{k+1}} \underline{P}_k^\perp \underline{h}_{k+1}$$

$$= \underline{H}_k \hat{\underline{\theta}}_k + \frac{(\underline{I} - \underline{P}_k) \underline{h}_{k+1} \underline{h}_{k+1}^T \underline{P}_k^\perp \underline{x}}{\underline{h}_{k+1}^T \underline{P}_k^\perp \underline{h}_{k+1}}$$

$$= \underline{H}_k \hat{\underline{\theta}}_k + \frac{\underline{h}_{k+1} \underline{h}_{k+1}^T \underline{P}_k^\perp \underline{x}}{\underline{h}_{k+1}^T \underline{P}_k^\perp \underline{h}_{k+1}} - \frac{\underline{H}_k (\underline{H}_k^T \underline{H}_k)^{-1} \underline{H}_k^T \cdot \underline{h}_{k+1} \underline{h}_{k+1} \underline{P}_k^\perp \underline{x}}{\underline{h}_{k+1}^T \underline{P}_k^\perp \underline{h}_{k+1}}$$

$$= [\underline{H}_k \ \underline{h}_{k+1}] \left[\begin{array}{c} \hat{\underline{\theta}}_k - \dfrac{(\underline{H}_k^T \underline{H}_k)^{-1} \underline{H}_k^T \cdot \underline{h}_{k+1} \underline{h}_{k+1}^T \underline{P}_k^\perp \underline{x}}{\underline{h}_{k+1}^T \underline{P}_k^\perp \underline{h}_{k+1}} \\ \underline{h}_{k+1}^T \underline{P}_k^\perp \underline{x} \end{array} \right.$$

$$= \underline{H}_{n+1} \hat{\underline{\theta}}_{k+1}$$

15) Clearly, $\hat{A}_1 = \bar{x}$. From (8.28) with
$\underline{H}_1 = \underline{1} \quad \underline{h}_2 = [1 \; r \; \ldots \; r^{N-1}]^T = \underline{h}$

$$\hat{A}_2 = \hat{A}_1 - \frac{\frac{1}{N} \underline{1}^T \underline{h} \underline{h}^T \underline{P}_1^{\perp} \underline{x}}{\underline{h}^T \underline{P}_1^{\perp} \underline{h}}$$

$$\hat{B}_2 = \frac{\underline{h}^T \underline{P}_1^{\perp} \underline{x}}{\underline{h}^T \underline{P}_1^{\perp} \underline{h}}$$

$$\underline{P}_1^{\perp} = \underline{I} - \underline{1}(\underline{1}^T \underline{1})^{-1} \underline{1}^T = \underline{I} - 1/N \, \underline{1}\underline{1}^T$$

$$\underline{h}^T \underline{P}_1^{\perp} \underline{x} = \underline{h}^T (\underline{x} - \bar{x} \underline{1})$$

$$= \sum_n x[n] r^n - \bar{x} \sum_n r^n$$

$$\underline{h}^T \underline{P}_1^{\perp} \underline{h} = \underline{h}^T (\underline{I} - 1/N \, \underline{1}\underline{1}^T) \underline{h}$$

$$= \underline{h}^T \underline{h} - 1/N (\underline{1}^T \underline{h})^2$$

$$= \sum_n r^{2n} - 1/N \left(\sum_n r^n\right)^2$$

$$\hat{A}_2 = \bar{x} - \frac{\frac{1}{N} \sum_n r^n \left[\sum_n x[n] r^n - \bar{x} \sum_n r^n \right)}{\sum_n r^{2n} - \frac{1}{N} \left(\sum_n r^n\right)^2}$$

$$\hat{B}_2 = \frac{\sum_n x(n) r^n - \bar{x} \sum_n r^n}{\sum_n r^{2n} - \frac{1}{N} \left(\sum_n r^n\right)^2}$$

16) $\underline{H}_1 = [1\,1\,\ldots\,1]^T \Rightarrow \hat{A}_1 = (\underline{H}_1^T \underline{H}_1)^{-1} \underline{H}_1^T \underline{x} = \bar{x}$

$J_{MIN_1} = \sum_{n=0}^{N-1} x^2[n] - N\bar{x}^2$

$$\underline{H}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ \vdots & 0 \\ \vdots & 1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}0\\0\\0\\0\end{matrix}}\right\} M\times 1 \\ \left.\vphantom{\begin{matrix}0\\0\\0\end{matrix}}\right\} (N-M)\times 1 \end{matrix} \quad \text{for} \quad \underline{\theta}_2 = \begin{bmatrix} A \\ B-A \end{bmatrix}$$

$$\begin{matrix} \uparrow & \uparrow \\ \underline{H}_1 & \underline{h}_2 \end{matrix}$$

$\underline{P}_1^\perp = \underline{I} - 1/N\, \underline{1}\,\underline{1}^T$

$$\begin{aligned} \underline{h}_2^T \underline{P}_1^\perp \underline{x} &= \underline{h}_2^T (\underline{x} - 1/N\, \underline{1}\,\underline{1}^T \underline{x}) \\ &= \underline{h}_2^T (\underline{x} - \bar{x}\underline{1}) \\ &= \sum_{n=M}^{N-1} (x[n] - \bar{x}) = \sum_{M}^{N-1} x[n] - (N-M)\bar{x} \end{aligned}$$

$$\begin{aligned} \underline{h}_2^T \underline{P}_1^\perp \underline{h}_2 &= \underline{h}_2^T \underline{h}_2 - \underline{h}_2^T \underline{1}\, \underline{1}^T \underline{h}_2 / N \\ &= (N-M) - (N-M)^2/N \\ &= (N-M)M/N \end{aligned}$$

$$\Rightarrow \hat{A}_2 = \hat{A}_1 - \frac{1/N\, \underline{1}^T \underline{h}_2 \left[\sum_{M}^{N-1} x[n] - (N-M)\bar{x} \right]}{(N-M)M/N}$$

$$= \bar{x} - \frac{\frac{N-M}{N} \left[\sum_{M}^{N-1} x[n] - (N-M)\bar{x} \right]}{(N-M)M/N}$$

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

$$\widehat{B-A} = \frac{\sum_{M}^{N-1} x[n] - (N-M)\bar{x}}{(N-M)M/N}$$

Let $\bar{x}_1 = \frac{1}{N-M} \sum_{M}^{N-1} x[n]$

$$\widehat{B-A} = \frac{N}{M} (\bar{x}_1 - \bar{x})$$

$$\hat{\underline{\theta}}_2 = \begin{bmatrix} \frac{1}{M} \sum_0^{M-1} x[n] \\ \frac{N}{M} (\bar{x}_1 - \bar{x}) \end{bmatrix}$$

The decrease in J_{MIN} is from (8.31)

$$\frac{(\underline{h}_2^T \underline{P}_1^{\perp} \underline{x})^2}{\underline{h}_2^T \underline{P}_1^{\perp} \underline{h}_2} = \underline{h}_2^T \underline{P}_1^{\perp} \underline{x} \, [\hat{\underline{\theta}}_2]_2$$

$$= \left(\sum_{M}^{N-1} x[n] - (N-M)\bar{x} \right) (\hat{\underline{\theta}}_2)_2$$

$$= (N-M)(\bar{x}_1 - \bar{x}) \frac{N}{M} (\bar{x}_1 - \bar{x})$$

$$= \frac{N}{M} (N-M) (\bar{x}_1 - \bar{x})^2$$

To detect a jump (positive or negative) we test to see if the LS 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 J_{MIN} is just $(N-M) \frac{M}{N} (\widehat{B-A})^2$.

17) Wish to show that $\underline{h}_i^T P_k^{\perp} \underline{x} = 0$ for $i = 1, 2, \ldots, k$

$$\begin{bmatrix} \underline{h}_1^T \underline{P}_k^\perp \underline{x} \\ \vdots \\ \underline{h}_k^T \underline{P}_k^\perp \underline{x} \end{bmatrix} = \begin{bmatrix} \underline{h}_1^T \\ \vdots \\ \underline{h}_k^T \end{bmatrix} \underline{P}_k^\perp \underline{x} = \underline{H}_k^T \underline{P}_k^\perp \underline{x}$$

$$= \underline{H}_k^T (\underline{I} - \underline{H}_k (\underline{H}_k^T \underline{H}_k)^{-1} \underline{H}_k^T) \underline{x}$$
$$= (\underline{H}_k^T - \underline{H}_k^T) \underline{x} = \underline{0}$$

18) $$J_{MIN_{k+1}} = \underline{x}^T \underline{P}_{k+1}^\perp \underline{x}$$
$$= \underline{x}^T \underline{x} - \underline{x}^T \underline{P}_{k+1} \underline{x}$$
$$= \underline{x}^T \underline{x} - \underline{x}^T \underline{P}_k \underline{x} - \underline{x}^T \frac{\underline{P}_k^\perp \underline{h}_{k+1} \underline{h}_{k+1}^T \underline{P}_k^\perp \underline{x}}{\underline{h}_{k+1}^T \underline{P}_k^\perp \underline{h}_{k+1}}$$
$$= J_{MIN_k} - \frac{(\underline{h}_{k+1}^T \underline{P}_k^\perp \underline{x})^2}{\underline{h}_{k+1}^T \underline{P}_k^\perp \underline{h}_{k+1}}$$

19) $$J_{MIN}[N] = \sum_0^N \frac{1}{\sigma_n^2} (x[n] - \hat{A}[N])^2$$

$$= \sum_0^{N-1} \frac{1}{\sigma_n^2} \left[x[n] - \hat{A}[N-1] - K[N] (x[N] - \hat{A}[N-1]) \right]^2$$
$$+ \frac{1}{\sigma_N^2} (x[N] - \hat{A}[N])^2$$

$$= \sum_0^{N-1} \frac{1}{\sigma_n^2} (x[n] - \hat{A}[N-1])^2 - 2K[N] (x[N] - \hat{A}[N-1]) \cdot \sum_0^{N-1} 1/\sigma_n^2 (x[n] - \hat{A}[N-1])$$
$$+ K^2[N] (x[N] - \hat{A}[N-1])^2 \sum_0^{N-1} 1/\sigma_n^2$$
$$+ \frac{1}{\sigma_N^2} \left[x[N] - \hat{A}[N-1] - K[N] (x[N] - \hat{A}[N-1]) \right]^2$$

But $\sum_0^{N-1} 1/\sigma_n^2 (x[n] - \hat{A}[N-1]) = 0$

due to definition of $\hat{A}[N-1]$.

$$J_{MIN}[N] = J_{MIN}[N-1] + \frac{K^2[N]}{var(\hat{A}[N-1])}(x[N]-\hat{A}[N-1])^2$$

$$+ \frac{1}{\sigma_N^2}(x[N]-\hat{A}[N-1])^2 - 2\frac{K[N]}{\sigma_N^2}(x[N]-\hat{A}[N-1])^2$$

$$+ \frac{1}{\sigma_N^2}K^2[N](x[N]-\hat{A}[N-1])^2$$

$$= J_{MIN}[N-1] + (x[N]-\hat{A}[N-1])^2 \, \mathcal{J}$$

where $\mathcal{J} = \dfrac{K^2[N]}{var(\hat{A}[N-1])} + \dfrac{1}{\sigma_N^2} - \dfrac{2K[N]}{\sigma_N^2} + \dfrac{K^2[N]}{\sigma_N^2}$

But from (8.38), $1-K[N] = \dfrac{\sigma_N^2}{var(\hat{A}[N-1])+\sigma_N^2}$

$$\mathcal{J} = \frac{var(\hat{A}[N-1])}{(var(\hat{A}[N-1])+\sigma_N^2)^2} + \frac{1}{\sigma_N^2}\frac{\sigma_N^4}{(var(\hat{A}[N-1])+\sigma_N^2)^2}$$

$$= \frac{1}{var(\hat{A}[N-1])+\sigma_N^2}$$

20) $\underline{H}[n] = \begin{bmatrix} 1 \\ r \\ \vdots \\ r^n \end{bmatrix} \Rightarrow \underline{h}[n] = r^n$

$$\hat{A}[n] = \hat{A}[n-1] + K[n](x[n] - r^n \hat{A}[n-1])$$

from (8.46)

Now $\sigma_n^2 = \sigma^2 = 1$ and $\Sigma[n] = \mathrm{var}(\hat{A}[n])$

$$\Rightarrow \quad K[n] = \frac{\mathrm{var}(\hat{A}[n-1])\, r^n}{1 + r^{2n}\,\mathrm{var}(\hat{A}[n-1])} \qquad \text{from } (8.47)$$

$$\mathrm{var}(\hat{A}[n]) = (1 - K[n] r^n)\,\mathrm{var}(\hat{A}[n-1]) \qquad \text{from } (8.48)$$

To find the variance explicitly let
$v_n = \mathrm{var}(\hat{A}[n])$

$$v_n = \left[1 - \frac{v_{n-1} r^{2n}}{1 + r^{2n} v_{n-1}} \right] v_{n-1}$$

$$= \frac{v_{n-1}}{1 + r^{2n} v_{n-1}}$$

Now let $v_0 = 1$ as was given

$$v_1 = \frac{1}{1+r^2} \qquad v_2 = \frac{\frac{1}{1+r^2}}{1 + r^4\left(\frac{1}{1+r^2}\right)}$$

$$= \frac{1}{1 + r^2 + r^4}$$

or in general

$$v_n = \mathrm{var}(\hat{A}[n]) = \frac{1}{\sum_{k=0}^{n} r^{2k}}$$

21) $$K[N] = \frac{\mathrm{var}(\hat{A}[N-1])}{\mathrm{var}(\hat{A}[N-1]) + \sigma_N^2}$$

$$\mathrm{var}(\hat{A}[N]) = (1-K[N])\,\mathrm{var}(\hat{A}[N-1])$$

Let $v_N = \mathrm{var}(\hat{A}[N])$

$$v_N = \left(1 - \frac{v_{N-1}}{v_{N-1}+\sigma_N^2}\right) v_{N-1}$$

$$= \frac{\sigma_N^2\, v_{N-1}}{v_{N-1}+\sigma_N^2}$$

$$\frac{1}{v_N} = \frac{v_{N-1}+\sigma_N^2}{\sigma_N^2\, v_{N-1}} = \frac{1}{\sigma_N^2} + \frac{1}{v_{N-1}}$$

$$= \frac{1}{r^N} + \frac{1}{v_{N-1}}$$

Since $\frac{1}{v_0} = 1$, we have

$$\frac{1}{v_N} = \sum_{n=0}^{N} 1/r^n \quad \text{or} \quad v_N = \frac{1}{\sum_{n=0}^{N} 1/r^n}$$

$$K[N] = \frac{1/\sum_{n=0}^{N} 1/r^n}{1/\sum_{n=0}^{N} 1/r^n + r^N} = \frac{1}{1 + r^N \sum_{n=0}^{N} 1/r^n}$$

$$= \frac{1}{1+\sum_{0}^{N} r^{N-n}} = \frac{1}{1+\sum_{0}^{N} r^n}$$

If $r=1$, $\mathrm{var}(\hat{A}[N]) \to 0$ as $N\to\infty$

$K[N] \to 0$ as $N \to \infty$

If $0 < r < 1$, $var(\hat{A}[N]) \to 0$ as $N \to \infty$
$K[N] \to$ Constant as $N \to \infty$

If $r > 1$, $var(\hat{A}[N]) \to \frac{r-1}{r}$ as $N \to \infty$

since $\sum_{n=0}^{\infty} 1/r^n = \frac{1}{1-1/r} = \frac{r}{r-1}$

and $K[N] \to 0$ as $N \to \infty$. In this case the data become so noisy that the gain $\to 0$ (we do not use the data) and hence the variance does not go to zero.

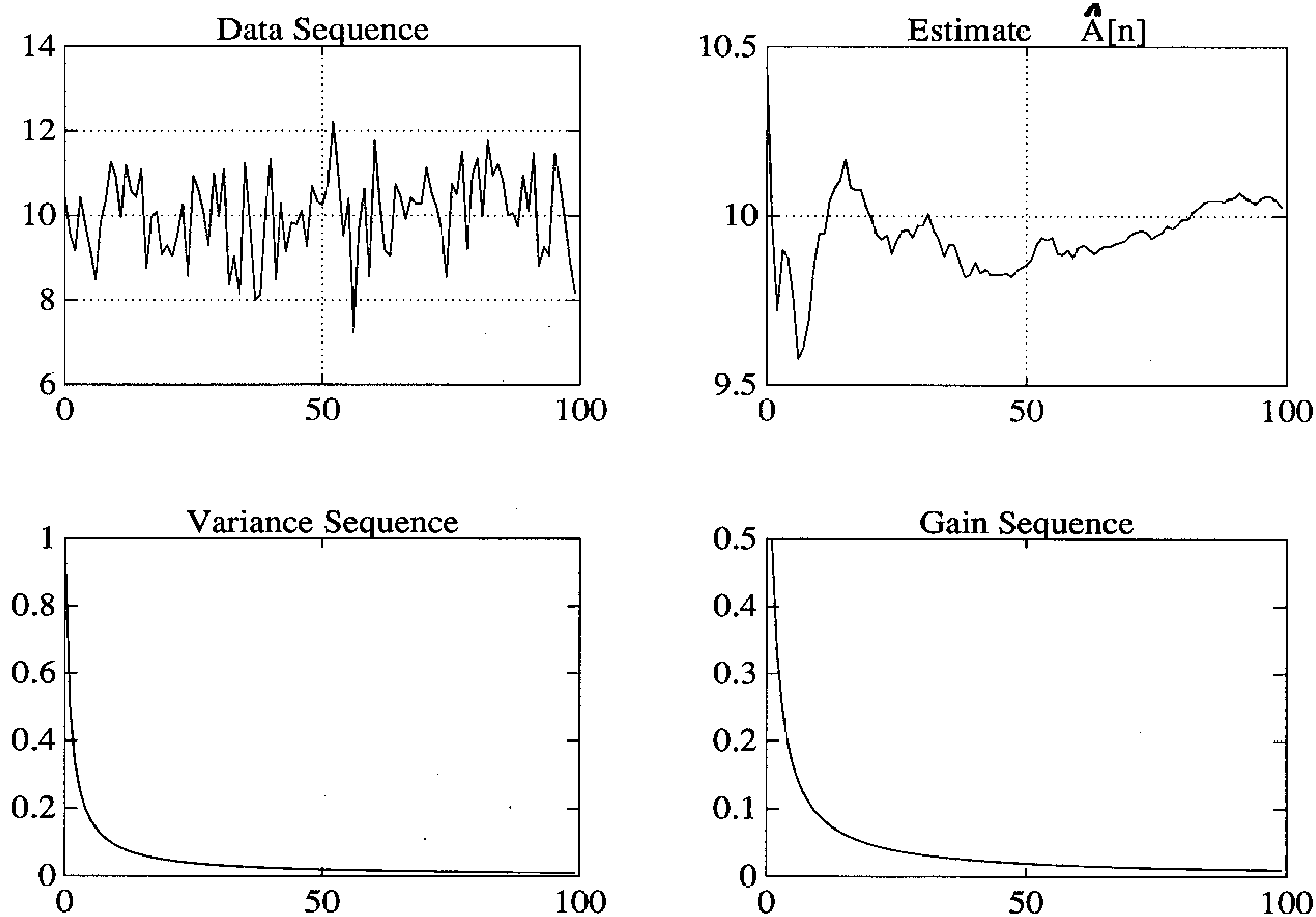
22) See plots on next few pages. Compare these results to those of the previous problem.

23) $$\hat{\underline{\theta}}_s[n] = \left(\sum_{k=-p}^{-1} 1/\sigma_k^2 \, \underline{h}[k] \underline{h}^T[k] + \sum_{k=0}^{n} \frac{1}{\sigma_k^2} \underline{h}[k] \underline{h}^T[k] \right)^{-1}$$
$$\cdot \left(\sum_{k=-p}^{-1} \frac{1}{\sigma_k^2} x[k] \underline{h}^T[k] + \sum_{k=0}^{n} \frac{1}{\sigma_k^2} x[k] \underline{h}^T[k] \right)$$

But $$\hat{\underline{\theta}}[-1] = \left(\sum_{k=-p}^{-1} \frac{1}{\sigma_k^2} \underline{h}[k] \underline{h}^T[k] \right)^{-1}$$
$$\cdot \left(\sum_{k=-p}^{-1} \frac{1}{\sigma_k^2} x[k] \underline{h}^T[k] \right)$$

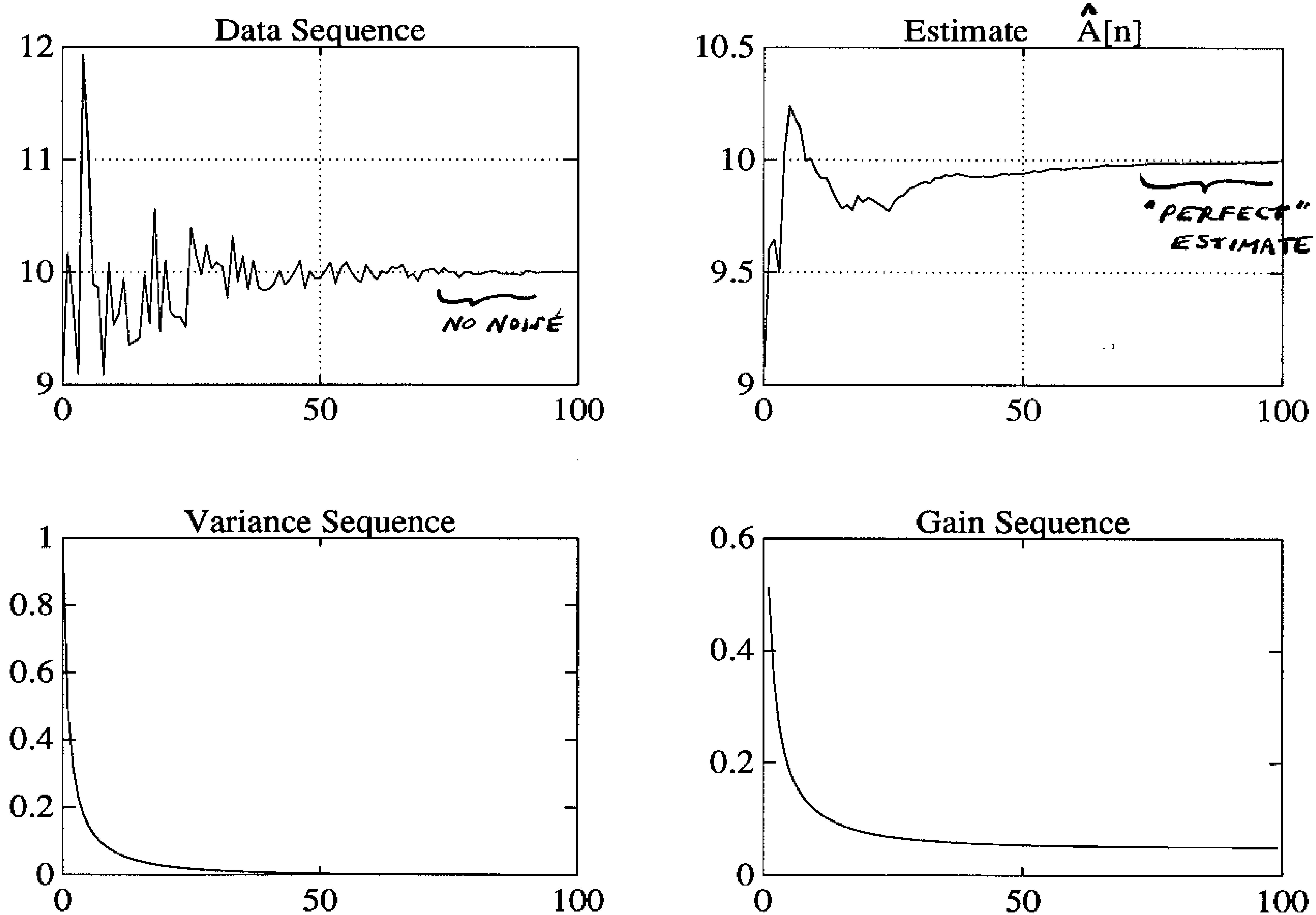
$$\underline{\Sigma}[-1] = \left(\sum_{k=-p}^{-1} \frac{1}{\sigma_k^2} \underline{h}[k] \underline{h}^T[k] \right)^{-1}$$

Prob. 8.22



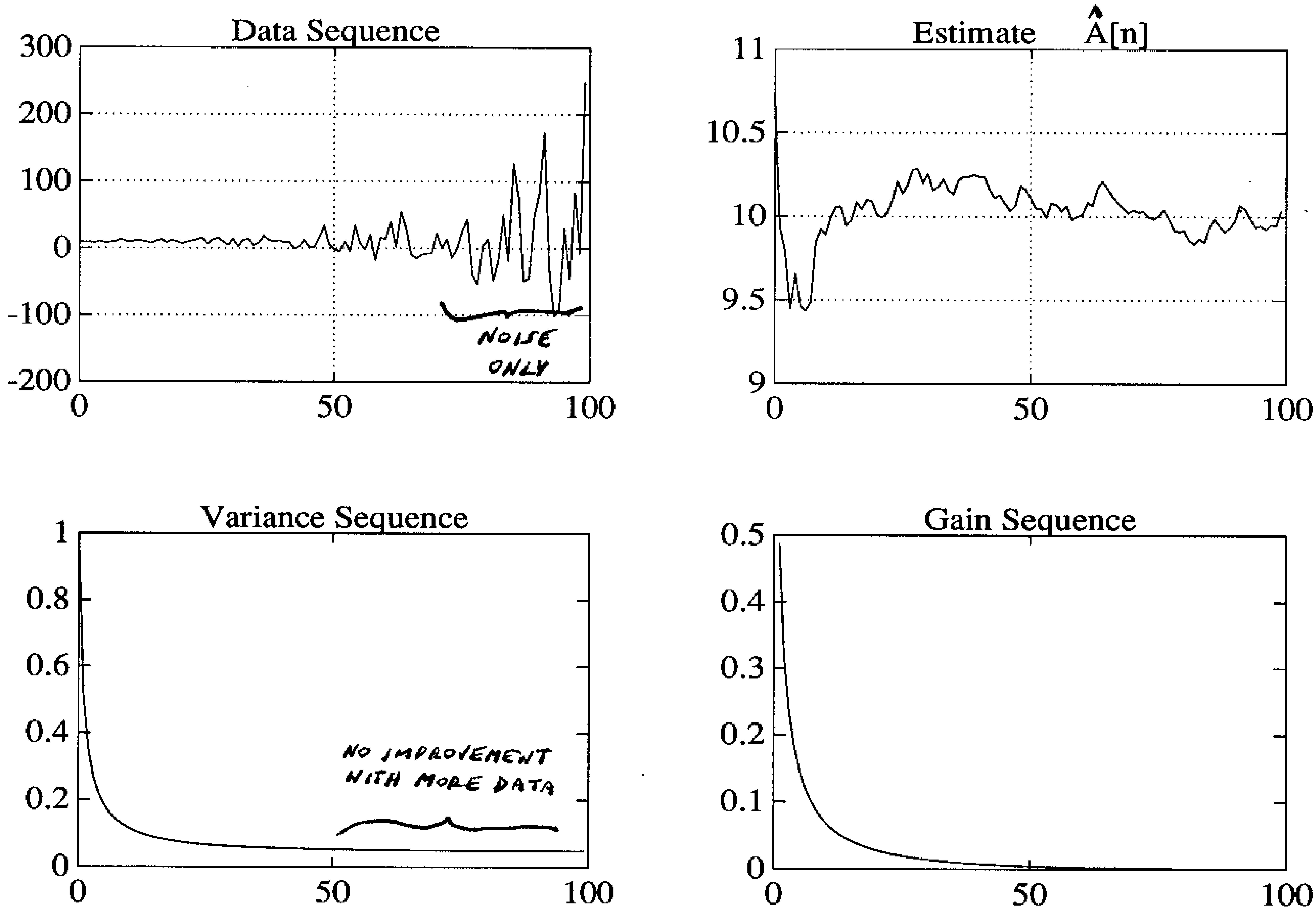
r = 1.00

Prob. 8.22



$r = 0.95$

Prob. 8.22



$r = 1.05$

$$\Rightarrow \quad \hat{\underline{\theta}}_S[n] = \left(\underline{\Sigma}^{-1}[-1] + \sum_{k=0}^{n} \frac{1}{\sigma_k^2} \underline{h}[k] \underline{h}^T[k] \right)^{-1}$$

$$\cdot \left(\underline{\Sigma}^{-1}[-1] \, \hat{\underline{\theta}}[-1] + \sum_{k=0}^{n} \frac{1}{\sigma_k^2} x[k] \underline{h}^T[k] \right)$$

Now for $n \geq p$, $\sum_{k=0}^{n} 1/\sigma_k^2 \, \underline{h}[k] \underline{h}^T[k]$ will be invertible so we can let $\alpha \to \infty$. Then, $\underline{\Sigma}^{-1}[-1] \to \underline{0}$ to yield

$$\hat{\underline{\theta}}_S[n] = \left(\sum_{k=0}^{n} \frac{1}{\sigma_k^2} \underline{h}[k] \underline{h}^T[k] \right)^{-1}$$

$$\cdot \left(\sum_{k=0}^{n} \frac{1}{\sigma_k^2} x[k] \underline{h}^T[k] \right)$$

$$= \hat{\theta}_B[n]$$

Note that if $\underline{\Sigma}[-1] \to \infty$, the choice of $\hat{\underline{\theta}}[-1]$ is immaterial.

24) Projecting $\underline{x}$ onto subspace spanned by $\underline{h}_1$ and $\underline{h}_2$ produces $\hat{\underline{s}} = \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{x}$ where $\underline{H} = [\underline{h}_1 \; \underline{h}_2]$. Now the constraint subspace is spanned by $[1 \; 1 \; 0]^T$. The projection onto this subspace is

$$\hat{\underline{s}}_c = \underline{e}_c (\underline{e}_c^T \underline{e}_c)^{-1} \underline{e}_c^T \hat{\underline{s}} \qquad \underline{e}_c = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

$$= \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} (1 \; 1 \; 0) \begin{bmatrix} x[0] \\ x[1] \\ 0 \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}(x[0]+x[1]) \\ \frac{1}{2}(x[0]+x[1]) \\ 0 \end{bmatrix}$$

25) $\underline{H} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ \vdots & \vdots \\ 1 & -1 \\ 1 & 1 \end{bmatrix}$ $\Rightarrow$ columns of $\underline{H}$ are orthogonal

$$\hat{\underline{\theta}} = (\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{x} = \begin{bmatrix} N & 0 \\ 0 & N \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix}^T \underline{x}$$

$$= \begin{bmatrix} \frac{1}{N} \sum_0^{N-1} x[n] \\ \frac{1}{N} \sum_0^{N-1} (-1)^n x[n] \end{bmatrix}$$

Now, if $A = B$ we have $(1 \ -1)\underline{\theta} = 0$
or $\underline{A} = [1 \ -1]$, $b = 0$ and thus from (8.52)

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

$$= \hat{\underline{\theta}} - \frac{1}{N}\underline{A}^T\left(\frac{1}{N}\underline{A}\underline{A}^T\right)^{-1}\underline{A}\hat{\underline{\theta}}$$

$$= (\underline{I} - \underline{A}^T(\underline{A}\underline{A}^T)^{-1}\underline{A})\hat{\underline{\theta}}$$

$$\underline{A}^T(\underline{A}\underline{A}^T)^{-1}\underline{A} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}\left([1 \ -1]\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)^{-1}(1 \ -1)$$

$$= \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

$$\underline{I} - \underline{A}^T(\underline{A}\underline{A}^T)^{-1}\underline{A} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$\hat{\underline{\theta}}_c = \begin{bmatrix} \frac{1}{2}\left(\bar{x} + \frac{1}{N}\sum_{n=0}^{N-1}(-1)^n x[n]\right) \\ \frac{1}{2}\left(\bar{x} + \frac{1}{N}\sum_{n=0}^{N-1}(-1)^n x[n]\right) \end{bmatrix}$$

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 $A = B$

$s[n] = A \quad n$ even

$\qquad 0 \quad n$ odd

26) By assumption

$h(g^{-1}(\hat{\alpha})) \leq h(g^{-1}(\alpha))$ for all α

$\Rightarrow \quad h(\theta_0) \leq h(\theta)$ for all θ

where $\theta_0 = g^{-1}(\hat{\alpha})$

But then θ_0 minimizes $h(\theta)$, i.e. $\theta_0 = \hat{\theta}$

$\Rightarrow \quad \hat{\theta} = g^{-1}(\hat{\alpha})$

27) From (8.61)

$$\theta_{k+1} = \theta_k + \left(\underline{H}^T(\theta_k)\underline{H}(\theta_k) - \sum_{n=0}^{N-1} G_n(\theta_k)\,(x[n] - e^{\theta_k}) \right)^{-1} \cdot \underline{H}^T(\theta_k)\,(\underline{x} - e^{\theta_k}\underline{1})$$

$$[\underline{H}(\theta)]_i = \frac{\partial s[i]}{\partial \theta} = e^{\theta} \qquad i = 0, 1, \ldots, N-1$$

$$[\underline{G}_n(\theta)]_{11} = \frac{\partial^2 s[n]}{\partial \theta^2} = e^{\theta}$$

$$\Rightarrow \quad \underline{H}(\theta) = e^{\theta}\underline{1} \qquad G_n(\theta) = e^{\theta}$$

$$\theta_{k+1} = \theta_k + \left(N e^{2\theta_k} - \sum_{n=0}^{N-1} e^{\theta_k} (x[n] - e^{\theta_k}) \right)^{-1}$$

$$\cdot \; e^{\theta_k} \underline{1}^T (\underline{x} - e^{\theta_k} \underline{1})$$

$$= \theta_k + \frac{e^{\theta_k}(N\bar{x} - N e^{\theta_k})}{N e^{2\theta_k} - N\bar{x} e^{\theta_k} + N e^{2\theta_k}}$$

$$= \theta_k + \frac{\bar{x} - e^{\theta_k}}{2 e^{\theta_k} - \bar{x}}$$

To find analytically let $\alpha = e^{\theta} \Rightarrow \hat{\alpha} = \bar{x}$ and from Prob 8.26 $\hat{\theta} = \ln \bar{x}$.

28) Since $H(z) = B(z)/A(z)$

$$= B(z) \frac{1}{A(z)}$$

$$h[n] = b[n] \star g[n]$$
$$= \sum_{k=0}^{n} b[k] g[n-k]$$

since $g[n]$ is causal. In matrix form for $n = 0, 1, \ldots, N-1$ we have $\underline{s} = \underline{G}\underline{b}$.

$$J(\underline{a}, \underline{b}) = (\underline{x} - \underline{s})^T (\underline{x} - \underline{s})$$
$$= (\underline{x} - \underline{G}\underline{b})^T (\underline{x} - \underline{G}\underline{b})$$
$$\Rightarrow \hat{\underline{b}} = (\underline{G}^T \underline{G})^{-1} \underline{G}^T \underline{x}$$

and $J(\underline{a}, \hat{\underline{b}}) = \underline{x}^T (\underline{I} - \underline{G}(\underline{G}^T\underline{G})^{-1}\underline{G}^T) \underline{x}$
from (8.11)

Now $G(z) A(z) = 1 \Rightarrow g[n] * a[n] = \delta[n]$

$$[\underline{A}^T \underline{G}]_{ij} = \sum_{k=1}^{N} (\underline{A}^T)_{ik} (\underline{G})_{kj} \qquad \begin{array}{l} i = 1, 2, \ldots, N-p \\ j = 1, 2, \ldots, q+1 \\ \;\; = 1, 2, \ldots, p \end{array}$$

$$= \sum_{k=1}^{N} a[p+i-k]\, g[k-j]$$

$$= \sum_{k=-\infty}^{\infty} a[p+i-k]\, g[k-j]$$

where $a[n] = 0$ for $n < 0$, $n > p$
$g[n] = 0$ for $n < 0$

$$= \sum_{l=-\infty}^{\infty} g[l]\, a[p+i-j-l]$$

$$= g[n] * a[n] \Big|_{n = p+i-j}$$

$$= \delta[p+i-j]$$

Since $1 \le i \le N-p$, $1 \le j \le p$,
$p+i-j > 0$ for all i, j

$\Rightarrow \underline{A}^T \underline{G} = \underline{0}$

$$\underline{L}(\underline{L}^T\underline{L})^{-1}\underline{L}^T = \underline{I}$$

$$[\underline{A}\ \underline{G}]\left(\begin{bmatrix}\underline{A}^T\\ \underline{G}^T\end{bmatrix}[\underline{A}\ \underline{G}]\right)^{-1}\begin{bmatrix}\underline{A}^T\\ \underline{G}^T\end{bmatrix} = \underline{I}$$

$$[\underline{A}\ \underline{G}]\begin{bmatrix}\underline{A}^T\underline{A} & \underline{0}\\ \underline{0} & \underline{G}^T\underline{G}\end{bmatrix}^{-1}\begin{bmatrix}\underline{A}^T\\ \underline{G}^T\end{bmatrix} = \underline{I}$$

$$\Rightarrow \quad \underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T + \underline{G}(\underline{G}^T\underline{G})^{-1}\underline{G}^T = \underline{I}$$

29) If $f_{0_{k+1}} = f_{0_k}$, $\phi_{k+1} = \phi_k$, we have

$$\sum_{n=-M}^{M} n\, x[n] \sin(2\pi \hat{f}_0 n + \hat{\phi}) = 0$$

$$\sum_{n=-M}^{M} x[n] \sin(2\pi \hat{f}_0 n + \hat{\phi}) = 0$$

at high SNR we have

$$\sum_n n \cos(2\pi f_0 n + \phi) \sin(2\pi \hat{f}_0 n + \hat{\phi}) = 0$$

$$\sum_n \cos(2\pi f_0 n + \phi) \sin(2\pi \hat{f}_0 n + \hat{\phi}) = 0$$

$$\Rightarrow$$

$$\frac{1}{2}\sum_n n\Big[\sin(2\pi(f_0+\hat{f}_0)n + \phi + \hat{\phi}) + \sin(2\pi(\hat{f}_0 - f_0)n + (\hat{\phi}-\phi))\Big] = 0$$

$$\frac{1}{2}\sum_n \sin(2\pi(f_0+\hat{f}_0)n + \phi + \hat{\phi}) + \sin(2\pi(\hat{f}_0 - f_0)n + (\hat{\phi}-\phi)) = 0$$

Neglecting the high frequency term we have

$$\sum_n n \sin\left(2\pi(\hat{f}_0 - f_0)n + \hat{\phi} - \phi\right) = 0$$

$$\sum_n \sin\left(2\pi(\hat{f}_0 - f_0)n + \hat{\phi} - \phi\right) = 0$$

for which the solution is $\hat{f}_0 = f_0$, $\hat{\phi} = \phi$.

Chapter 9

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

$$\Rightarrow \quad \sigma^2 = \frac{2}{\pi} E^2(x)$$

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

2) $E(x) = 0$ since the PDF is even.
Try $E(x^2)$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2}\sigma} e^{-\sqrt{2}|x|/\sigma} dx$$

$$= \frac{2}{\sqrt{2}\sigma} \int_0^\infty x^2 e^{-\sqrt{2}x/\sigma} dx = \sigma^2$$

$$\Rightarrow \quad \sigma = \sqrt{E(x^2)} \quad \text{and} \quad \hat{\sigma} = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} x^2[n]}$$

3) $\rho = \mathrm{cov}(u, v)$ where $\underline{x} = \begin{bmatrix} u \\ v \end{bmatrix}$
$= E(uv)$

$$\Rightarrow \quad \hat{\rho} = 1/N \sum_{n=0}^{N-1} u_n v_n \quad \text{where} \quad \underline{x}[n] = \begin{bmatrix} u_n \\ v_n \end{bmatrix}$$

The cubic equation was found in Prob. 7.11. Clearly, the method of moments estimator is much simpler to find and implement.

4) Since $\mu = E(x)$, $\sigma^2 = \mathrm{var}(x)$

$$\hat{\mu} = \frac{1}{N}\sum_{n=0}^{N-1} x[n] \qquad \hat{\sigma}^2 = \frac{1}{N}\sum_{n=0}^{N-1}(x[n]-\hat{\mu})^2$$

5) Replace the theoretical ACF by $\hat{r}_{xx}[k]$ where

$$\hat{r}_{xx}[n] = \frac{1}{N}\sum_{n=0}^{N-1-|k|} x[n]\,x[n+|k|]$$

$$\Rightarrow \quad \hat{r}_{xx}[n] = -\sum_{k=1}^{p} a[k]\,\hat{r}_{xx}[n-k] \qquad n > q$$

Choosing $n = q+1, \ldots, q+p$ yields the linear equations

$$\begin{bmatrix} \hat{r}_{xx}[q] & \ldots & \hat{r}_{xx}[q-p+1] \\ & & \\ \hat{r}_{xx}[q+p-1] & \ldots & \hat{r}_{xx}[q] \end{bmatrix} \begin{bmatrix} a[1] \\ \vdots \\ a[p] \end{bmatrix} = -\begin{bmatrix} \hat{r}_{xx}[q+1] \\ \vdots \\ \hat{r}_{xx}[q+p] \end{bmatrix}$$

which can be solved for the $a[k]$'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)

$$\mathrm{var}(\hat{\theta}) = \left(\frac{\partial g}{\partial T}\bigg|_{T=\mu}\right)^2 \mathrm{var}(T)$$

$$= \left(2T\big|_{T=\mu}\right)^2 \mathrm{var}(T)$$

$$= (2A)^2\sigma^2/N = 4A^2\sigma^2/N$$

7) $$E(x[n]) = \cos\phi$$
$$\Rightarrow \quad \phi = \arccos E(x[n])$$
or
$$\hat{\phi} = \arccos\left(\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right)$$

$$\hat{\phi} = h(\underline{w}) \quad \text{where}$$
$$h(\underline{w}) = \arccos\left[\frac{1}{N}\sum_{n=0}^{N-1}(\cos\phi + w[n])\right]$$

From (9.18)
$$E(\hat{\theta}) = h(\underline{0}) = \arccos\left(\frac{1}{N}\sum_{n=0}^{N-1}\cos\phi\right) = \phi$$
From (9.19)
$$\operatorname{var}(\hat{\phi}) = \left.\frac{\partial h}{\partial \underline{w}}^T\right|_{\underline{w}=\underline{0}} \sigma^2 \underline{I} \left.\frac{\partial h}{\partial \underline{w}}\right|_{\underline{w}=\underline{0}}$$

$$\frac{\partial h}{\partial w[i]} = -\frac{1}{\sqrt{1-u^2}}\frac{\partial u}{\partial w[i]}$$
$$\text{where} \quad u = \frac{1}{N}\sum_{n=0}^{N-1}(\cos\phi + w[n])$$

$$\left.\frac{\partial h}{\partial w[i]}\right|_{\underline{w}=\underline{0}} = -\frac{1}{\sqrt{1-\cos^2\phi}}\frac{1}{N} = -\frac{1}{N\sin\phi}$$

$$\operatorname{var}(\hat{\phi}) = \sigma^2 \sum_{n=0}^{N-1}\left(\left.\frac{\partial h}{\partial w[n]}\right|_{\underline{w}=\underline{0}}\right)^2$$
$$= \sigma^2 \sum_{n=0}^{N-1}\frac{1}{N^2\sin^2\phi} = \frac{\sigma^2}{N\sin^2\phi}$$

8) $$\hat{\theta} = g(\underline{T})$$

$$\simeq g(\underline{\mu}) + \sum_{k=1}^{r} \frac{\partial g}{\partial T_k}\Big|_{\underline{T}=\underline{\mu}} (T_k - \mu_k)$$

$$+ \frac{1}{2}(\underline{T}-\underline{\mu})^T \frac{\partial^2 g}{\partial \underline{T} \partial \underline{T}^T}\Big|_{\underline{T}=\underline{\mu}} (\underline{T}-\underline{\mu})$$

where $\frac{\partial^2 g}{\partial \underline{T} \partial \underline{T}^T}$ is the Hessian

$$\begin{aligned}
E(\hat{\theta}) &= g(\underline{\mu}) + \tfrac{1}{2} E\left[(\underline{T}-\underline{\mu})^T \underline{G}(\mu)(\underline{T}-\underline{\mu})\right] \\
&= g(\underline{\mu}) + \tfrac{1}{2} E\left[tr\left((\underline{T}-\underline{\mu})^T \underline{G}(\mu)(\underline{T}-\underline{\mu})\right)\right] \\
&= g(\underline{\mu}) + \tfrac{1}{2} tr\, E\left[G(\underline{\mu})(\underline{T}-\underline{\mu})(\underline{T}-\underline{\mu})^T\right] \\
&= g(\underline{\mu}) + \tfrac{1}{2} tr(\underline{G}(\mu)\underline{C}_T)
\end{aligned}$$

9) $$\hat{\theta} \simeq g(\underline{\mu}) + \frac{\partial g}{\partial \underline{T}}\Big|_{\underline{T}=\underline{\mu}}^T (\underline{T}-\underline{\mu})$$

$$+ \tfrac{1}{2}(\underline{T}-\underline{\mu})\underline{G}(\underline{\mu})(\underline{T}-\underline{\mu})$$

$$\begin{aligned}
var(\hat{\theta}) &= E\left[(\hat{\theta}-E(\hat{\theta}))^2\right] \\
&= E\Big[\Big(\frac{\partial g}{\partial T}\Big|_{\underline{T}=\underline{\mu}}^T (\underline{T}-\underline{\mu}) + \tfrac{1}{2}(\underline{T}-\underline{\mu})^T \underline{G}(\underline{\mu})(\underline{T}-\underline{\mu}) \\
&\qquad - \tfrac{1}{2} tr\left(\underline{G}(\underline{\mu})\underline{C}_T\right)\Big)^2\Big]
\end{aligned}$$

using the results from Prob 9.8

But $\underline{T}-\underline{\mu} \sim N(\underline{0}, \underline{C}_T)$ so that all odd-order moments are zero. Let $\underline{x} = \underline{T}-\underline{\mu}$, $\underline{b} = \frac{\partial g}{\partial \underline{T}}\Big|_{T=\underline{\mu}}$ and $\underline{A} = \underline{G}(\underline{\mu})$.

$$\mathrm{var}(\hat{\theta}) = E\left[\left(\underline{b}^T\underline{x} + \tfrac{1}{2}\underline{x}^T\underline{A}\underline{x} - \tfrac{1}{2}\,\mathrm{tr}[\underline{A}\underline{C}_T]\right)^2\right]$$
$$= \underline{b}^T\underline{C}_T\underline{b} + \tfrac{1}{4}E[(\underline{x}^T\underline{A}\underline{x})^2] - \tfrac{1}{4}E(\underline{x}^T\underline{A}\underline{x})\,\mathrm{tr}(\underline{A}\underline{C}_T)$$
$$- \tfrac{1}{4}E(\underline{x}^T\underline{A}\underline{x})\,\mathrm{tr}(\underline{A}\underline{C}_T) + \tfrac{1}{4}\mathrm{tr}^2(\underline{A}\underline{C}_T)$$

But
$$E(\underline{x}^T\underline{A}\underline{x}) = E[\mathrm{tr}(\underline{A}\underline{x}\underline{x}^T)]$$
$$= \mathrm{tr}(\underline{A}E(\underline{x}\underline{x}^T))$$
$$= \mathrm{tr}(\underline{A}\underline{C}_T)$$
$$E[(\underline{x}^T\underline{A}\underline{x})^2] = \mathrm{var}(\underline{x}^T\underline{A}\underline{x}) + E^2(\underline{x}^T\underline{A}\underline{x})$$
$$= 2\,\mathrm{tr}[(\underline{A}\underline{C}_T)^2] + \mathrm{tr}^2(\underline{A}\underline{C}_T)$$

$$\mathrm{var}(\hat{\theta}) = \underline{b}^T\underline{C}_T\underline{b} + \tfrac{1}{2}\mathrm{tr}[(\underline{A}\underline{C}_T)^2] + \tfrac{1}{4}\mathrm{tr}^2(\underline{A}\underline{C}_T)$$
$$- \tfrac{1}{2}\mathrm{tr}^2(\underline{A}\underline{C}_T) + \tfrac{1}{4}\mathrm{tr}^2(\underline{A}\underline{C}_T)$$

$$= \underline{b}^T\underline{C}_T\underline{b} + \tfrac{1}{2}\mathrm{tr}[(\underline{A}\underline{C}_T)^2]$$

10) $g(T_1) = 1/T_1$ where $T_1 = \frac{1}{N}\sum_{n=0}^{N-1} x[n] = \bar{x}$

$\bar{x}$ will be approximately Gaussian due to the central limit theorem

$$E(\hat{\lambda}) = g(\mu) + \tfrac{1}{2}\mathrm{tr}[G(\mu)C_T]$$

But $\mu = E(T_1) = 1/\lambda$

$$C_T = \mathrm{var}(T_1) = \mathrm{var}(x(n))/N = \frac{1}{N\lambda^2}$$
$$G(\mu) = \left.\frac{\partial^2 g}{\partial T^2}\right|_{T=\mu}$$
$$= \left.\frac{2}{T^3}\right|_{T=\mu} = \frac{2}{\mu^3} = 2\lambda^3$$

$$E(\hat{\lambda}) = \lambda + \frac{1}{2}(2\lambda^3)\frac{1}{N\lambda^2} = \lambda + \lambda/N$$

$$= \lambda(1 + 1/N)$$

$$\operatorname{var}(\hat{\lambda}) = \left(\frac{\partial g}{\partial T}\Big|_{T=\mu}\right)^2 \operatorname{var}(T) + \frac{1}{2}\operatorname{tr}\left[(G(\mu)C_T)^2\right]$$

$$= \lambda^2/N + \frac{1}{2}\left(2\lambda^3 \, 1/N\lambda^2\right)^2$$

$$= \lambda^2/N + 2\lambda^2/N^2 = \frac{\lambda^2}{N}(1 + 2/N)$$

The estimator displays a bias of λ/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.

11) $$\frac{1}{N-1}\sum_{n=0}^{N-2} A\cos(2\pi f_0 n + \phi) A\cos(2\pi f_0(n+1) + \phi)$$

$$= \frac{A^2}{2(N-1)}\sum_{n=0}^{N-2}\left[\cos(4\pi f_0 n + 2\pi f_0 + 2\phi) + \cos(2\pi f_0)\right]$$

as $N \to \infty$, the double-frequency term $\to 0$ for f_0 not near 0 or 1/2. Hence, the overall expression $\to \frac{A^2}{2}\cos 2\pi f_0$.
Method of moments is valid since ensemble mean equals temporal mean as $N \to \infty$ (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 $A^2/2 \cos 2\pi f_0$. We need to normalize out the $A^2/2$ factor.

Now at high SNR

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

where $s(n) = A\cos(2\pi f_0 n + \phi)$

and as $N \to \infty$, this becomes $A^2/2 \cos 2\pi f_0$ from Prob. 9.11. Similarly,

$$\frac{1}{N}\sum_{n=0}^{N-1} x^2(n) \to A^2/2$$

so that $\hat{f}_0 \to f_0$. Also, from (9.18) we have $E(\hat{f}_0) = f_0$ for large N. Note that

$$\hat{f}_0 = \frac{1}{2\pi}\arccos\left[\frac{\hat{r}_{xx}(1)}{\hat{r}_{xx}[0]}\right]$$ so that

it is a method of moments estimator.

At lower SNR, as $N \to \infty$ we have

$$\hat{r}_{xx}[1] \to r_{xx}[1] = A^2/2 \cos 2\pi f_0$$

$$\hat{r}_{xx}(0) \to r_{xx}(0) = \frac{A^2}{2} + \sigma^2$$

(assuming ϕ is $U[0,2\pi]$). Thus, as $N \to \infty$

$$\hat{f}_0 \to \frac{1}{2\pi}\arccos\left[\frac{A^2\cos 2\pi f_0}{A^2+2\sigma^2}\right] \neq f_0$$

$\hat{f}_0$ will be severely biased.

13) $$\text{var}(\hat{f}_0) = \frac{\sigma^2}{(2\pi)^2 (N-1)^2 \sin^2 2\pi f_0}$$

$$\cdot \left[s^2[1] + 4\cos^2 2\pi f_0 \sum_{n=1}^{N-2} s^2[n] + s^2[N-2] \right]$$

$$f_0 = 0.25 \Rightarrow \cos 2\pi f_0 = 0$$

$$s[1] = \sqrt{2} \cos 2\pi f_0 = 0$$

$$s[N-2] = \sqrt{2} \cos 2\pi \tfrac{1}{4}(N-2)$$

$$= \sqrt{2} \cos \tfrac{\pi}{2}(N-2) = 0$$

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

1) $\hat{\theta} = E(\theta \mid \underline{x}) = \int \theta\, p(\theta \mid \underline{x})\, d\theta$

$$= \int \theta \, \frac{p(\underline{x} \mid \theta)\, p(\theta)}{p(\underline{x})}\, d\theta$$

$$= \frac{\int \theta\, p(\underline{x} \mid \theta)\, p(\theta)\, d\theta}{\int p(\underline{x} \mid \theta)\, p(\theta)\, d\theta} = \frac{\int \theta\, p(\underline{x} \mid \theta)\, \delta(\theta - \theta_0)\, d\theta}{\int p(\underline{x} \mid \theta)\, \delta(\theta - \theta_0)\, d\theta}$$

$$= \frac{\theta_0\, p(\underline{x} \mid \theta_0)}{p(\underline{x} \mid \theta_0)} = \theta_0$$

The MMSE estimator is just the true value of θ since our prior knowledge is perfect. Of course, this is not a valid estimator.

2) $p_{\underline{x}}(x[0], x[1] \mid A) = p_{\underline{w}}(x[0] - A, x[1] - A \mid A)$

But $\underline{w}$ is independent of A

$\Rightarrow \quad = p_{\underline{w}}(x[0] - A, x[1] - A)$

and $w[0]$ is independent of $w[1]$

$\Rightarrow \quad = p_w(x[0] - A)\, p_w(x[1] - A)$

$\quad = p_{x[0]}(x[0] \mid A)\, p_{x[1]}(x[1] \mid A)$

$$\underline{x} = \begin{bmatrix} A + w[0] \\ A + w[1] \end{bmatrix} = A\underline{1} + \underline{w} \sim \mathcal{N}(\underline{0}, \underline{1}\underline{1}^T + \underline{I})$$

since $A, w[0], w[1]$ are IID and $\sim \mathcal{N}(0, 1)$

Now consider the exponent of $p(\underline{x})$ or $\underline{x}^T \underline{C}^{-1} \underline{x}$

$$\underline{C} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \Rightarrow \underline{C}^{-1} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\underline{x}^T \underline{C}^{-1} \underline{x} = \frac{1}{3}\underline{x}^T \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \underline{x} = \frac{1}{3}\left(2x^2[0] + 2x^2[1] - 2x[0]x[1]\right)$$

which does not factor $\Rightarrow$ $x[0], x[1]$ are not independent

3) $$p(\underline{x}|\theta) = e^{-\sum_n (x[n]-\theta)} \qquad \text{all } x[n]\text{'s} > \theta$$

$$= e^{-N(\bar{x}-\theta)} \qquad \xi = \min x[n] > \theta$$

$$p(\theta|\underline{x}) = \frac{p(\underline{x}|\theta)p(\theta)}{\int p(\underline{x}|\theta)p(\theta)d\theta}$$

$$= \frac{e^{-N(\bar{x}-\theta)}e^{-\theta}}{\int_0^{\xi} e^{-N(\bar{x}-\theta)}e^{-\theta}d\theta}$$

$$= \frac{e^{\theta(N-1)}}{\left.\frac{e^{\theta(N-1)}}{N-1}\right|_0^{\xi}} = \begin{cases} \frac{(N-1)e^{\theta(N-1)}}{e^{(N-1)\xi}-1} & 0 < \theta < \xi \\ 0 & \text{otherwise} \end{cases}$$

$$\hat{\theta} = E(\theta|\underline{x}) = \frac{\int_0^{\xi} \theta(N-1)e^{\theta(N-1)}d\theta}{e^{(N-1)\xi}-1}$$

$$= \frac{N-1}{e^{(N-1)\xi}-1}\left[\left.\frac{\theta(N-1)-1}{(N-1)^2}e^{\theta(N-1)}\right|_0^{\xi}\right)$$

$$= \frac{1}{(N-1)(e^{(N-1)\mathfrak{z}}-1)}\left[(\mathfrak{z}(N-1)-1)\, e^{(N-1)\mathfrak{z}} + 1 \right]$$

$$= \frac{1}{(N-1)(e^{(N-1)\mathfrak{z}}-1)}\left[1 - e^{(N-1)\mathfrak{z}} + (N-1)\mathfrak{z}\, e^{(N-1)\mathfrak{z}} \right]$$

$$= \frac{\mathfrak{z}\, e^{(N-1)\mathfrak{z}}}{e^{(N-1)\mathfrak{z}}-1} - \frac{1}{N-1}$$

$$\hat{\theta} = \frac{\min x[n]}{1 - e^{-(N-1)\min x[n]}} - \frac{1}{N-1}$$

4) $p(\underline{x} \mid \theta) = \dfrac{1}{\theta^N}$ all $x[n] \le \theta$ or $\max x[n] \le \theta$

$p(\theta) = 1/\beta \qquad 0 \le \theta \le \beta$

$p(\theta \mid \underline{x})$ will be nonzero for $\max x[n] \le \theta \le \beta$
or $\mathfrak{z} \le \theta \le \beta$

$$p(\theta \mid \underline{x}) = \frac{p(\underline{x} \mid \theta)\, p(\theta)}{\int p(\underline{x} \mid \theta)\, p(\theta)\, d\theta}$$

$$= \frac{\frac{1}{\theta^N}\frac{1}{\beta}}{\int_{\mathfrak{z}}^{\beta} \frac{1}{\theta^N}\frac{1}{\beta}\, d\theta} = \frac{\frac{1}{\theta^N}}{-\frac{\theta^{-(N-1)}}{N-1}\Big|_{\mathfrak{z}}^{\beta}}$$

$$= \frac{-(N-1)\, 1/\theta^N}{\beta^{-(N-1)} - \mathfrak{z}^{-(N-1)}}$$

$$= \frac{1/\theta^N}{\frac{1}{(N-1)}(\mathfrak{z}^{-(N-1)} - \beta^{-(N-1)})}$$

$$\hat{\theta} = E(\theta|\underline{x}) = \int \theta p(\theta|\underline{x}) d\theta$$

$$= c\int_{\xi}^{\beta} \theta \frac{d\theta}{\theta^N} \qquad c = \frac{1}{\frac{1}{N-1}\left(\xi^{-(N-1)} - \beta^{-(N-1)}\right)}$$

$$= c \left. \frac{\theta^{-(N-2)}}{-(N-2)} \right|_{\xi}^{\beta} = -\frac{c}{N-2}\left(\beta^{-(N-2)} - \xi^{-(N-2)}\right)$$

$$= \frac{(\max x[n])^{-(N-2)} - \beta^{-(N-2)}}{(\max x[n])^{-(N-1)} - \beta^{-(N-1)}} \quad \frac{N-1}{N-2}$$

Note that for β large (no prior knowledge) and N large

$$\hat{\theta} \approx \frac{N-1}{N-2} \max x[n] \approx \max x[n]$$

which agrees with the MLE.

57) $$Bmse(\hat{\theta}) = E_{x,\theta}\left\{\left[(\theta - E(\theta|\underline{x})) + (E(\theta|\underline{x}) - \hat{\theta})\right]^2\right\}$$

$$= E_x\left[E_{\theta|x}\left\{\left[(\theta - E(\theta|\underline{x})) + (E(\theta|\underline{x}) - \hat{\theta})\right]^2\right\}\right]$$

$$= E_x \Big\{E_{\theta|x}\left[(\theta - E(\theta|\underline{x}))^2\right]$$
$$+ 2E_{\theta|x}\left[(\theta - E(\theta|\underline{x}))(E(\theta|\underline{x}) - \hat{\theta})\right]$$
$$+ E_{\theta|x}\left[(E(\theta|\underline{x}) - \hat{\theta})^2\right]\Big\}$$

The middle term is zero since conditioned on $\underline{x}$, $\hat{\theta}$ is a constant so that

$$E_{\theta|x}\left[(\theta - E(\theta|\underline{x}))(E(\theta|\underline{x}) - \hat{\theta})\right]$$
$$= E_{\theta|x}\left[\theta - E(\theta|\underline{x})\right]\left[E(\theta|\underline{x}) - \hat{\theta}\right]$$
$$= \left[\underbrace{E_{\theta|x}(\theta)}_{E(\theta|\underline{x})} - E(\theta|\underline{x})\right]\left[E(\theta|\underline{x}) - \hat{\theta}\right] = 0$$

$$Bmse(\hat{\theta}) = E_x\left\{E_{\theta|x}\left[(\theta - E(\theta|\underline{x}))^2\right] + E_{\theta|x}\left[(E(\theta|\underline{x}) - \hat{\theta})^2\right]\right\}$$

Clearly, to minimize $Bmse(\hat{\theta})$ we choose $\hat{\theta} = E(\theta|\underline{x})$ so that the last term (which is nonnegative) is zero.

b) Now A and $w[n]$ are not independent.

$$p(w[n]|A) = \begin{cases} \dfrac{1}{\sqrt{2\pi\sigma_+^2}} e^{-\frac{1}{2\sigma_+^2} w^2[n]} & A \geq 0 \\ \dfrac{1}{\sqrt{2\pi\sigma_-^2}} e^{-\frac{1}{2\sigma_-^2} w^2[n]} & A < 0 \end{cases}$$

$$\Rightarrow p(x[n]|A) = \begin{cases} \dfrac{1}{\sqrt{2\pi\sigma_+^2}} e^{-\frac{1}{2\sigma_+^2}(x[n]-A)^2} & A \geq 0 \\ \dfrac{1}{\sqrt{2\pi\sigma_-^2}} e^{-\frac{1}{2\sigma_-^2}(x[n]-A)^2} & A < 0 \end{cases}$$

or

$$p(\underline{x}|A) = \begin{cases} \dfrac{1}{(2\pi\sigma_+^2)^{N/2}} e^{-\frac{1}{2\sigma_+^2}\sum_n (x[n]-A)^2} & A \geq 0 \\ \dfrac{1}{(2\pi\sigma_-^2)^{N/2}} e^{-\frac{1}{2\sigma_-^2}\sum_n (x[n]-A)^2} & A < 0 \end{cases}$$

since conditional on A the $w[n]$'s and hence the $x[n]$'s are independent. Clearly,

for $\sigma_+^2 \neq \sigma_-^2$, $p(\underline{x}|A) \neq p(\underline{x};A)$.
For $\sigma_+^2 = \sigma_-^2$ they are identical.

7) $$\hat{A} = \frac{\int_{-A_0}^{A_0} A \frac{1}{\sqrt{2\pi\sigma^2/N}} e^{-\frac{1}{2\sigma^2/N}(A-\bar{x})^2} dA}{\int_{-A_0}^{A_0} \frac{1}{\sqrt{2\pi\sigma^2/N}} e^{-\frac{1}{2\sigma^2/N}(A-\bar{x})^2} dA}$$

$$I_1 = \int_{-A_0}^{A_0} A e^{-a(A-\bar{x})^2} dA \qquad a = \frac{1}{2\sigma^2/N}$$

$$= \int_{-A_0}^{A_0} (A-\bar{x}) e^{-a(A-\bar{x})^2} dA + \bar{x} \int_{-A_0}^{A_0} e^{-a(A-\bar{x})^2} dA$$

$$= \underbrace{\int_{-A_0-\bar{x}}^{A_0-\bar{x}} y e^{-ay^2} dy}_{I_2} + \bar{x} \int_{-A_0}^{A_0} e^{-a(A-\bar{x})^2} dA$$

$$I_2 = \left. \frac{e^{-ay^2}}{-2a} \right|_{-A_0-\bar{x}}^{A_0-\bar{x}} = -\frac{1}{2a}\left(e^{-a(A_0-\bar{x})^2} - e^{-a(A_0+\bar{x})^2}\right)$$

The numerator becomes

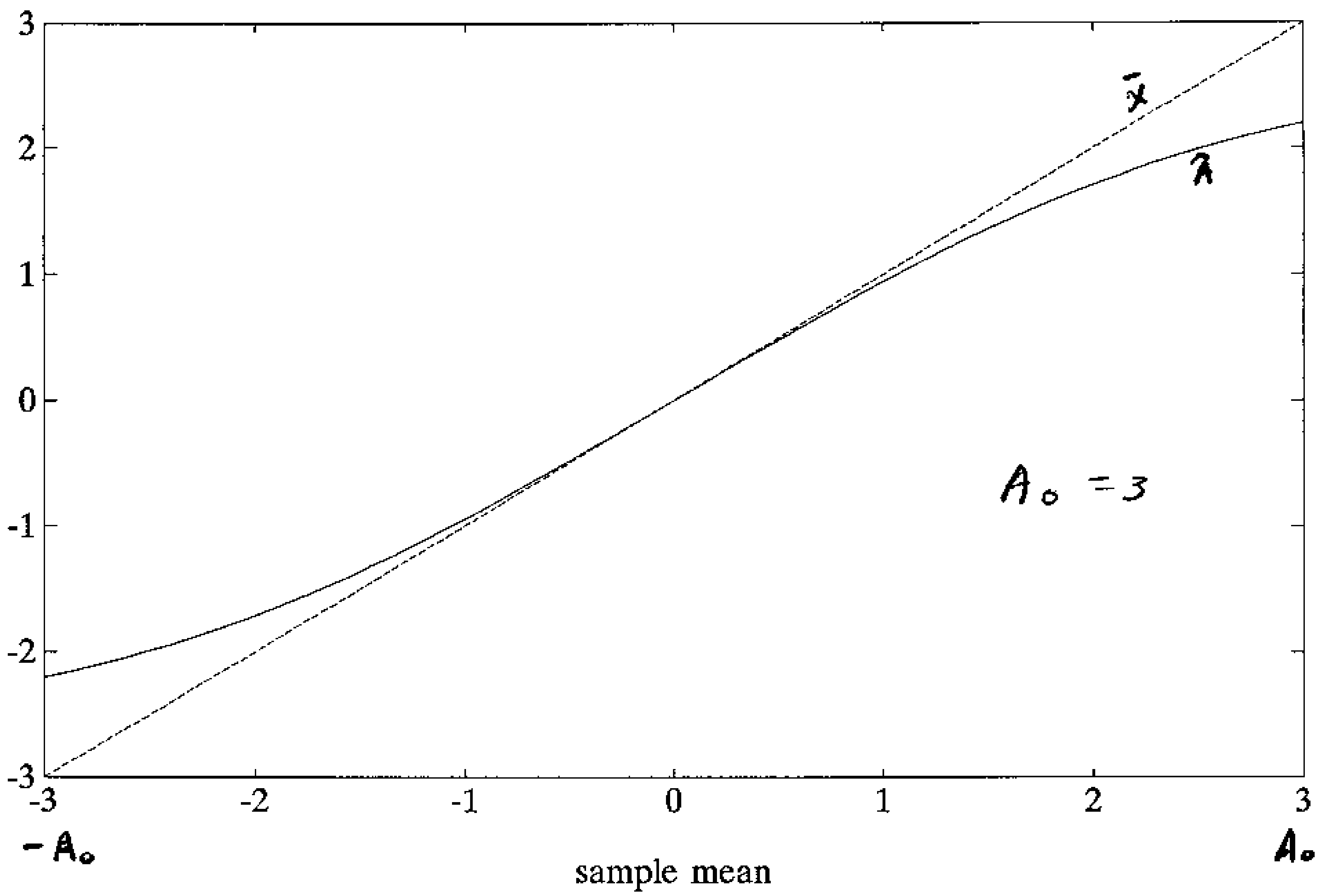
$$-\frac{1}{\sqrt{2\pi\sigma^2/N}} \frac{1}{2\frac{1}{2\sigma^2/N}} \left(e^{-\frac{1}{2\sigma^2/N}(A_0-\bar{x})^2} - e^{-\frac{1}{2\sigma^2/N}(A_0+\bar{x})^2}\right)$$
$$+ \frac{\bar{x}}{\sqrt{2\pi\sigma^2/N}} \int_{-A_0}^{A_0} e^{-\frac{1}{2\sigma^2/N}(A-\bar{x})^2} dA$$

or letting $\Phi(x)$ be the CDF for a $N(0,1)$ random variable

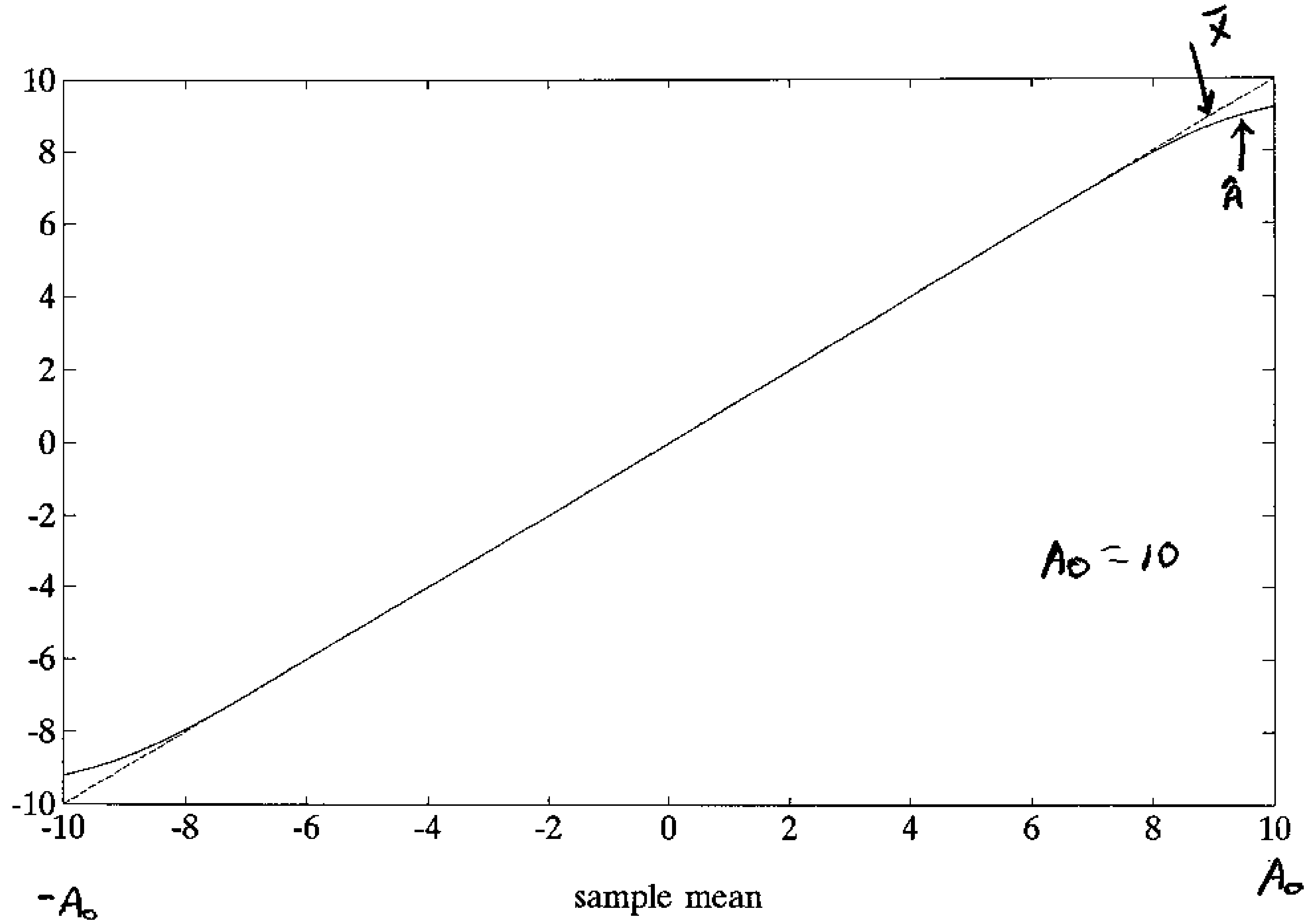
$$\hat{A} = \bar{x} + \frac{\sqrt{\frac{\sigma^2/N}{2\pi}}\left[e^{-\frac{1}{2\sigma^2/N}(A_0+\bar{x})^2} - e^{-\frac{1}{2\sigma^2/N}(A_0-\bar{x})^2}\right)}{\Phi\left(\frac{A_0-\bar{x}}{\sqrt{\sigma^2/N}}\right) - \Phi\left(\frac{-A_0-\bar{x}}{\sqrt{\sigma^2/N}}\right)}$$

For $\sqrt{\sigma^2/N} = 1$, $A_0 = 3$ we have

$$\hat{A} = \bar{x} + \frac{e^{-\frac{1}{2}(3+\bar{x})^2} - e^{-\frac{1}{2}(3-\bar{x})^2}}{\sqrt{2\pi}\left[\Phi(3-\bar{x}) - \Phi(-3-\bar{x})\right)}$$



and for $A_0 = 10$ (see the plot on next page) the curves are nearly identical or $\hat{A} \simeq \bar{x}$ for $|\bar{x}| \le A_0$. Also, as $\bar{x} \to \infty$, we will always have $\hat{A} \to A_0$.



8) Let $J = E[(\theta - \hat{\theta})^2]$

$$J = E[((\theta - E(\theta)) + (E(\theta) - \hat{\theta}))^2]$$

$$= E[(\theta - E(\theta))^2] + 2E[(\theta - E(\theta))(E(\theta) - \hat{\theta})]$$

$$+ E[(E(\theta) - \hat{\theta})^2]$$

Since $\hat{\theta}$ is a constant, the middle term is zero,

$$E[(\theta - E(\theta))(E(\theta) - \hat{\theta})] = E[(\theta - E(\theta))](E(\theta) - \hat{\theta})$$

$$= (E(\theta) - E(\theta))(E(\theta) - \hat{\theta}) = 0$$

$$J = E[(\theta - E(\theta))^2] + (E(\theta) - \hat{\theta})^2$$
$$\geq E[(\theta - E(\theta))^2]$$

$\Rightarrow \quad \hat{\theta} = E(\theta)$. The minimum MSE is just $E[(\theta - E(\theta))^2] = \text{var}(\theta)$.

From Example 10.1 with no data

$$\text{Bmse}(\hat{A}) = \text{var}(\hat{A}) = \sigma_A^2$$

and with data

$$\text{Bmse}(\hat{A}) = \sigma_A^2 \frac{\sigma^2/N}{\sigma_A^2 + \sigma^2/N} < \sigma_A^2.$$

9) Prior PDF: $\mathcal{N}(\underset{\mu_R}{100}, \underset{\sigma_R^2}{0.011})$

Data model: $x[n] = R + w[n] \quad n = 0, 1, \ldots, N-1$
where $w[n] \sim \mathcal{N}(0, \underset{\sigma^2}{1})$ and $w[n]$'s are independent

This is just Example 10.1.

$$\Rightarrow \quad \text{Bmse}(\hat{R}) = \frac{\sigma_R^2 \sigma^2/N}{\sigma_R^2 + \sigma^2/N}$$

For the error to be 0.1 on the average

$\Rightarrow \quad \text{Bmse}(\hat{R}) = 0.01$

$$0.01 = \frac{0.011/N}{0.011 + 1/N} \Rightarrow N = 9.09$$

or $N = 10$

Without prior knowledge or as $\sigma_R^2 \to \infty$

$\text{Bmse}(\hat{R}) \to \sigma^2/N$

$0.01 = 1/N \Rightarrow$ Require $N = 100$.

10) $$p(\theta|\underline{x}) = \frac{p(\underline{x}|\theta)p(\theta)}{\int p(\underline{x}|\theta)p(\theta)d\theta}$$

$$= \frac{\theta^N e^{-N\theta\bar{x}} \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\lambda\theta}}{\frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty \theta^{N+\alpha-1} e^{-(N\bar{x}+\lambda)\theta} d\theta}$$

But $\int_0^\infty p(\theta)d\theta = 1 \Rightarrow \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\lambda\theta} d\theta = 1$

$$\int_0^\infty \theta^{\underbrace{N+\alpha}_{\alpha'}-1} e^{-\underbrace{(N\bar{x}+\lambda)}_{\lambda'}\theta} d\theta = \frac{\Gamma(\alpha')}{\lambda'^{\alpha'}}$$

$$= \frac{\Gamma(N+\alpha)}{(N\bar{x}+\lambda)^{N+\alpha}}$$

$$p(\theta|\underline{x}) = \begin{cases} \frac{(N\bar{x}+\lambda)^{N+\alpha}}{\Gamma(N+\alpha)} \theta^{N+\alpha-1} e^{-(N\bar{x}+\lambda)\theta} & \theta > 0 \\ 0 & \theta < 0 \end{cases}$$

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 $p(\underline{x}|\theta)$ 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)

$$\hat{W} = E(W) + \frac{\operatorname{cov}(h, W)}{\operatorname{var}(h)}(h - E(h))$$

From Fig 10.9b there does not appear to be any correlation between height and weight $\Rightarrow \operatorname{Cov}(h, W) = 0$ or $\hat{W} = E(W) = 150$ for any height.

12) $$p(x, y) = \frac{1}{2\pi \det^{1/2}(\underline{C})} e^{-\frac{1}{2}\underline{x}^T \underline{C}^{-1}\underline{x}}$$

$$g(y) = p(x_0, y) = \frac{1}{2\pi \det^{1/2}(\underline{C})} e^{-\frac{1}{2}h(y)}$$

where $$h(y) = \begin{bmatrix} x_0 \\ y \end{bmatrix}^T \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{-1} \begin{bmatrix} x_0 \\ y \end{bmatrix}$$

$$= \frac{\begin{bmatrix} x_0 \\ y \end{bmatrix}^T \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y \end{bmatrix}}{1-\rho^2}$$

$$= (x_0^2 + y^2 - 2\rho x_0 y) / (1-\rho^2)$$

$g(y)$ 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 $E(x) = E(y) = 0$, $\mathrm{Cov}(x,y) = \rho$, $\mathrm{var}(x) = 1$

$$\Rightarrow E(y|x) = \rho x$$

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.

If $\rho = 0$, $E(y|x_0) = 0 = E(y)$

Contours of $p(x,y)$

13) Since

$$(\underline{A}+\underline{B}\underline{C}\underline{D})^{-1} = \underline{A}^{-1} - \underline{A}^{-1}\underline{B}(\underline{D}\underline{A}^{-1}\underline{B}+\underline{C}^{-1})^{-1}\underline{D}\underline{A}^{-1}$$

$$(\underline{C}_\theta^{-1} + \underline{H}^T\underline{C}_w^{-1}\underline{H})^{-1} = \underline{C}_\theta - \underline{C}_\theta \underline{H}^T(\underline{H}\underline{C}_\theta\underline{H}^T + \underline{C}_w)^{-1}\underline{H}\underline{C}_\theta$$

$\Rightarrow$ (10.33)

Now, using the inversion lemma again

$$(\underline{C}_\theta^{-1} + \underline{H}^T\underline{C}_w^{-1}\underline{H})^{-1}\underline{H}^T\underline{C}_w^{-1} = \underline{C}_\theta\underline{H}^T\underline{C}_w^{-1} - \underline{C}_\theta\underline{H}^T(\underline{H}\underline{C}_\theta\underline{H}^T+\underline{C}_w)^{-1}\underline{H}\underline{C}_\theta\underline{H}^T\underline{C}_w^{-1}$$

$$= \underline{C}_\theta\underline{H}^T \underbrace{\left[\underline{C}_w^{-1} - (\underline{H}\underline{C}_\theta\underline{H}^T+\underline{C}_w)^{-1}\underline{H}\underline{C}_\theta\underline{H}^T\underline{C}_w^{-1}\right]}_{\underline{E} = \underline{A}^{-1} - (\underline{B}+\underline{A})^{-1}\underline{B}\underline{A}^{-1}}$$

$$\begin{aligned}
\underline{E} &= \underline{A}^{-1} - \underline{A}^{-1}\underline{B}(\underline{B}^{-1}\underline{A})(\underline{B}+\underline{A})^{-1}\underline{B}\underline{A}^{-1} \\
&= \underline{A}^{-1} - \underline{A}^{-1}\underline{B}\left[\underline{B}^{-1}(\underline{B}+\underline{A})(\underline{B}^{-1}\underline{A})^{-1}\right]^{-1}\underline{A}^{-1} \\
&= \underline{A}^{-1} - \underline{A}^{-1}\underline{B}\left[\underline{B}^{-1}(\underline{B}+\underline{A})(\underline{A}^{-1}\underline{B})\right]^{-1}\underline{A}^{-1} \\
&= \underline{A}^{-1} - \underline{A}^{-1}\underline{B}\left[\underline{I} + \underline{A}^{-1}\underline{B}\right]^{-1}\underline{A}^{-1} \\
&= (\underline{A}+\underline{B})^{-1} \quad \text{by inversion lemma}
\end{aligned}$$

$$\Rightarrow \quad \underline{E} = (\underline{H}\underline{C}_\theta\underline{H}^T + \underline{C}_w)^{-1}$$

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)

$$\hat{A} = 0 + \left(1/\sigma_A^2 + \frac{\underline{H}^T \underline{H}}{\sigma^2} \right)^{-1} \underline{H}^T \frac{1}{\sigma^2} (\underline{x} - \underline{0})$$

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

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

From (10.13)

$$Bmse(\hat{A}) = \int var(A|\underline{x}) p(\underline{x}) d\underline{x}$$

But $var(\hat{A}) = \dfrac{1}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n}}$ from (10.33)

which does not depend on $\underline{x}$

$$\Rightarrow \quad Bmse(\hat{A}) = \frac{1}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n}}$$

15) $\ln p(\theta) = \ln \dfrac{1}{\sqrt{2\pi\sigma_\theta^2}} - \dfrac{1}{2\sigma_\theta^2} (\theta - \mu_\theta)^2$

$$E[\ln p(\theta)] = -\frac{1}{2} \ln 2\pi\sigma_\theta^2 - \frac{1}{2\sigma_\theta^2} \underbrace{E((\theta - \mu_\theta)^2)}_{\sigma_\theta^2}$$

$$= -\frac{1}{2} [1 + \ln 2\pi \sigma_\theta^2]$$

$$H(\theta) = \frac{1}{2}(1 + \ln 2\pi\sigma_\theta^2)$$

The more concentrated the PDF or for σ_θ^2 small,

the smaller will be $H(\theta)$. Higher entropy
$\Rightarrow$ larger σ_θ^2 or a more random θ.

$$
\begin{aligned}
I &= H(\theta) - H(\theta|\underline{x}) \\
&= -\int p(\theta)\ln p(\theta)\,d\theta + \iint p(\underline{x},\theta)\ln p(\theta|\underline{x})\,d\underline{x}\,d\theta \\
&= -\iint p(\underline{x},\theta)\,d\underline{x}\,\ln p(\theta)\,d\theta + \quad ''
\end{aligned}
$$

$$
\begin{aligned}
&= -\iint p(\underline{x},\theta)\ln p(\theta)\,d\underline{x}\,d\theta \\
&\quad + \iint p(\underline{x},\theta)\ln p(\theta|\underline{x})\,d\underline{x}\,d\theta
\end{aligned}
$$

$$
= \iint p(\underline{x},\theta)\ln\frac{p(\theta|\underline{x})}{p(\theta)}\,d\underline{x}\,d\theta
$$

$$
= \int \underbrace{\int p(\theta|\underline{x})\ln\frac{p(\theta|\underline{x})}{p(\theta)}\,d\theta}_{\geq 0 \text{ by given inequality}}\, p(\underline{x})\,d\underline{x}
$$

≥ 0 since $p(\underline{x}) \geq 0$.

$= 0$ if and only if $p(\theta|\underline{x}) = p(\theta)$

Makes sense since if posterior PDF is the same as prior PDF $\Rightarrow$ no information.

16) $H(\theta) = \frac{1}{2}(1 + \ln 2\pi\sigma_\theta^2)$ from Prob 10.15.
Similarly, since $p(\theta|\underline{x})$ is Gaussian,
$H(\theta|\underline{x}) = \frac{1}{2}(1 + \ln 2\pi\sigma_{\theta|x}^2)$
$\Rightarrow \quad I = H(\theta) - H(\theta|\underline{x}) = \frac{1}{2}\ln \sigma_\theta^2/\sigma_{\theta|x}^2$

But $\sigma_\theta^2 = \sigma_A^2$, $\sigma_{\theta|x}^2 = \sigma_{A|x}^2$

$$= \frac{\sigma_A^2 \sigma^2/N}{\sigma_A^2 + \sigma^2/N}$$

$$I = \frac{1}{2} \ln \frac{\sigma_A^2 + \sigma^2/N}{\sigma^2/N}$$

$$= \frac{1}{2} \ln \left(1 + \frac{\sigma_A^2}{\sigma^2/N} \right)$$

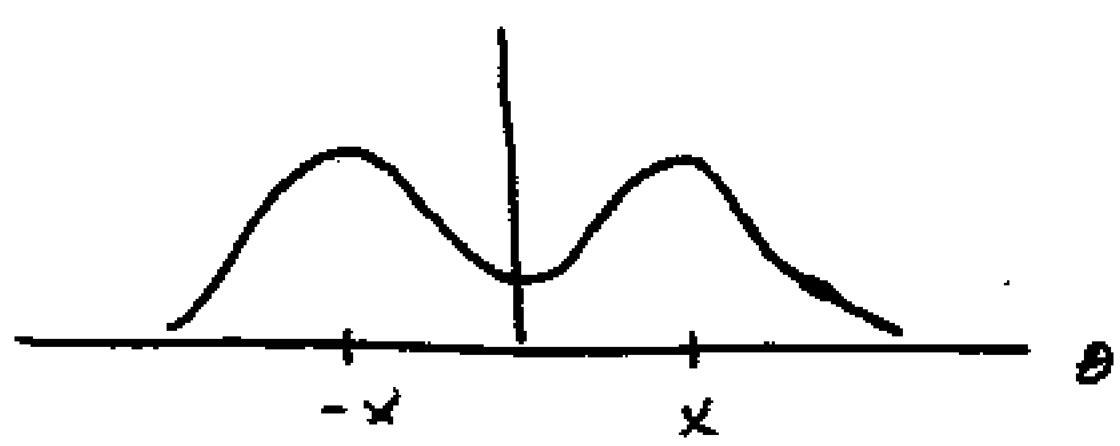
17) From Prob 10.16 we want to maximize I. We do so by letting $\sigma_A^2 \to \infty$. It doesn't matter what we choose for μ_A. This choice swamps out the prior as we have already observed.

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 $\sigma_0^2 \to 0$, $\hat{\mu} \to \mu_0$ prior knowledge dominates
As $\sigma_0^2 \to \infty$, $\hat{\mu} \to \bar{x}$ data " "

2) 

$\epsilon = \frac{1}{2}$

MMSE estimator is $E(\theta|x) = 0$ due to even symmetry of PDF. MAP estimator is mode or $\pm x$ (not unique)

$\epsilon = 3/4$

To find MMSE estimator

$$\hat{\theta} = E(\theta|x) = \int_{-\infty}^{\infty} \theta \frac{\epsilon}{\sqrt{2\pi}} e^{-\frac{1}{2}(\theta-x)^2} d\theta$$

$$+ \int_{-\infty}^{\infty} \theta \frac{1-\epsilon}{\sqrt{2\pi}} e^{-\frac{1}{2}(\theta+x)^2} d\theta$$

$$= \epsilon x + (1-\epsilon)(-x) = x(2\epsilon - 1) = x/2$$

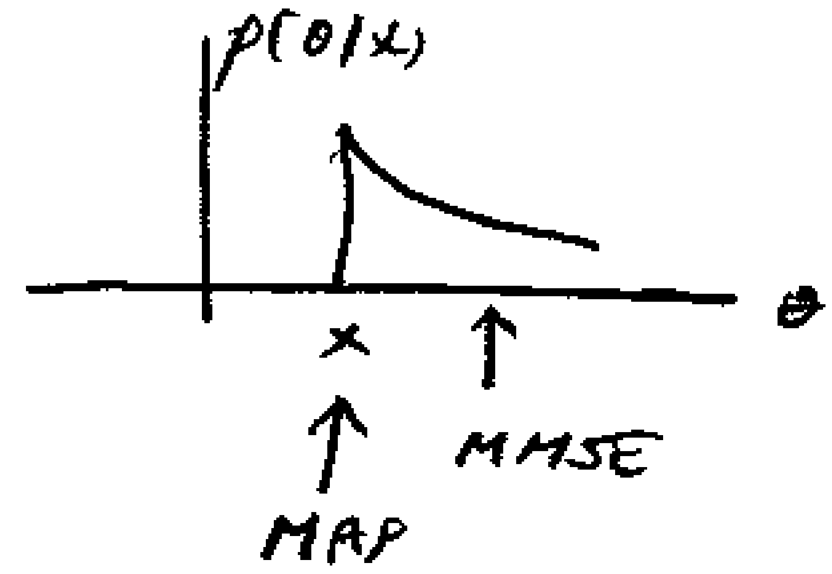
The MAP estimator is $\hat{\theta} = x$. Note that MAP $\neq$ MMSE.

3) MMSE :

$$\hat{\theta} = \int_x^\infty \theta\, e^{-(\theta - x)} d\theta = e^x \left[-\theta e^{-\theta} - e^{-\theta}\right]\Big|_x^\infty$$

$$= e^x (x e^{-x} + e^{-x}) = x + 1$$

MAP est. is just $\hat{\theta} = x$



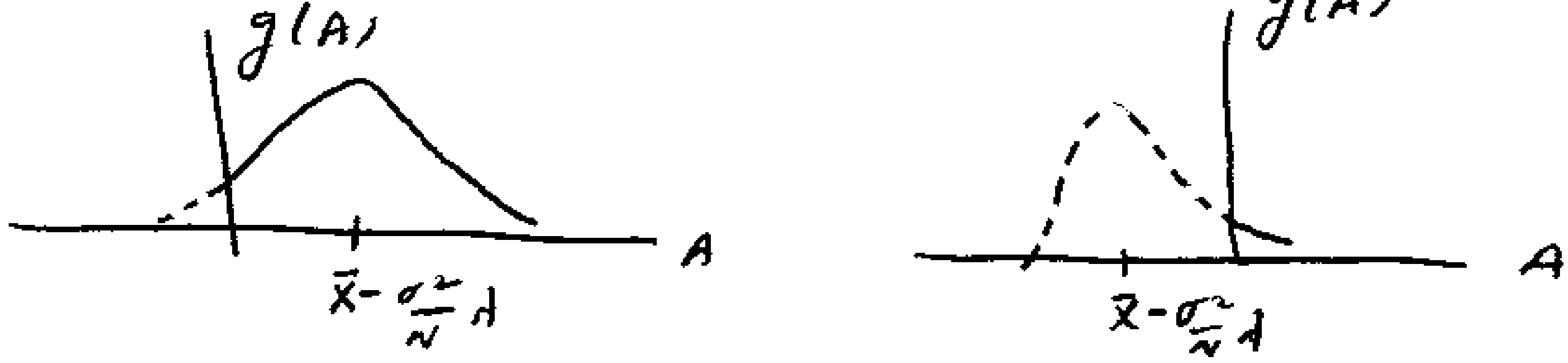
4) $$g(A) = p(\underline{x} \mid A)\, p(A)$$

$$= \begin{cases} \dfrac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_n (x(n) - A)^2} \lambda e^{-\lambda A} & A > 0 \\ 0 & A < 0 \end{cases}$$

$$\frac{\partial \ln g}{\partial A} = \frac{1}{\sigma^2} \sum_n (x(n) - A) - \lambda = 0$$

$$\Rightarrow \hat{A} = \bar{x} - \frac{\sigma^2}{N}\lambda$$

Note that if $\hat{A}$ is negative we use $\hat{A} = 0$ since



$$\Rightarrow \hat{A} = \max\left(0, \bar{x} - \frac{\sigma^2}{N}\lambda\right)$$

5) From (11.17)

$$\hat{\underline{\theta}} = E(\underline{\theta}|\underline{x})$$

$$= E(\underline{\theta}) + \underline{C}_{\theta x}\underline{C}_{xx}^{-1}(\underline{x} - E(\underline{x}))$$

From (11.27)

$$Bmse(\hat{\theta}_i) = (\underline{M}_{\hat{\theta}})_{ii}$$

$$= [\underline{C}_{\theta\theta} - \underline{C}_{\theta x}\underline{C}_{xx}^{-1}\underline{C}_{x\theta}]_{ii}$$

If $\underline{C}_{\theta x} = \underline{0}$, $\hat{\underline{\theta}} = E(\underline{\theta})$

$$Bmse(\hat{\theta}_i) = [\underline{C}_{\theta\theta}]_{ii} = var(\theta_i)$$

Data is irrelevant to estimator since $\underline{\theta}$ and $\underline{x}$ are independent (due to Gaussian assumption)

6) The Jacobian is

$$\underline{J} = \frac{\partial [A\ \phi]^T}{\partial [a\ b]^T} = \begin{bmatrix} \partial A/\partial a & \partial A/\partial b \\ \partial \phi/\partial a & \partial \phi/\partial b \end{bmatrix}$$

$$= \begin{bmatrix} \frac{a}{\sqrt{a^2+b^2}} & \frac{b}{\sqrt{a^2+b^2}} \\ \frac{b/a^2}{1+(b/a)^2} & \frac{-1/a}{1+(b/a)^2} \end{bmatrix}$$

$$= \begin{bmatrix} a/A & b/A \\ b/A^2 & -a/A^2 \end{bmatrix}$$

$$|\det \underline{J}| = |-a^2/A^3 - b^2/A^3| = 1/A$$

$$p(A, \phi) = \frac{p(a, b)}{|\det \underline{J}|}$$ Note: transformation is 1-1

$$= \frac{1}{2\pi\sigma_\theta^2} e^{-\frac{1}{2\sigma_\theta^2}(a^2+b^2)} \frac{1}{A}$$

$$= \frac{A}{\sigma_\theta^2} e^{-\frac{1}{2\sigma_\theta^2}A^2} \cdot \frac{1}{2\pi} \qquad \begin{array}{l} A > 0 \\ 0 \le \phi \le 2\pi \end{array}$$

$$p(A) = \int_0^{2\pi} p(A,\phi)\,d\phi = \frac{A}{\sigma_\theta^2} e^{-\frac{1}{2\sigma_\theta^2}A^2}$$

$$p(\phi) = \int_0^{\infty} p(A,\phi)\,dA = \frac{1}{2\pi}$$

Since $p(A,\phi)$ factors, A and ϕ are independent.

7) For Bayesian linear model, MMSE estimator = MAP estimator since $p(\underline{\theta}|\underline{x})$ is Gaussian. But MAP estimator maximizes $p(\underline{x}|\underline{\theta})p(\underline{\theta})$. With no prior information this is equivalent to maximizing $p(\underline{x}|\underline{\theta})$. In the Bayesian model $p(\underline{x}|\underline{\theta}) = p(\underline{x};\underline{\theta})$. Thus, maximizing $p(\underline{x};\underline{\theta})$, which yields the MLE or MVU estimator, also yields the MMSE estimator.

8) $$\hat{\underline{\theta}}[n] = E(\underline{\theta}[n]|\underline{x}) = E(\underline{A}\,\underline{\theta}[n-1]|\underline{x})$$

$$= \underline{A}\,E(\underline{\theta}[n-1]|\underline{x}) = \underline{A}\,\hat{\underline{\theta}}[n-1]$$

due to linearity of the expectation

Let $n=1 \Rightarrow \hat{\underline{\theta}}[1] = \underline{A}\,\hat{\underline{\theta}}[0]$

$n=2 \Rightarrow \hat{\underline{\theta}}[2] = \underline{A}\,\hat{\underline{\theta}}[1] = \underline{A}^2\,\hat{\underline{\theta}}[0]$

etc.

9) $$\underline{\theta}(n) = \begin{bmatrix} x[n] \\ y[n] \\ v_x \\ v_y \end{bmatrix} = \begin{bmatrix} x[n-1] + v_x \\ y[n-1] + v_y \\ v_x \\ v_y \end{bmatrix}$$

Since $x[n] - x[n-1] = x[0] + v_x n$
$- x[0] - v_x(n-1)$
$= v_x$

and similarly for the y component.

$$\underline{\theta}[n] = \underbrace{\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\underline{A}} \underbrace{\begin{bmatrix} x[n-1] \\ y[n-1] \\ v_x \\ v_y \end{bmatrix}}_{\underline{\theta}[n-1]}$$

$\Rightarrow$ from Prob. 11.8 $\quad \hat{\underline{\theta}}[n] = \underline{A}^n \hat{\underline{\theta}}[0]$

10) $\hat{\theta} = \dfrac{1}{\bar{x} + A/N}$ as $N \to \infty$, $\hat{\theta} \to 1/\bar{x}$ and thus as $N \to \infty$ for a given realization of θ, $\bar{x} \to E(x) = 1/\theta \Rightarrow \hat{\theta} \to \theta$. In general, as $N \to \infty$ the MAP estimator is just the value that maximizes $p(\underline{x}|\theta)$ or the Bayesian MLE. For a given realization of θ, 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.

11) $R = E[c(\underline{\epsilon})]$

$$= \iint c(\underline{\epsilon})\, p(\underline{x}, \underline{\theta})\, d\underline{x}\, d\underline{\theta}$$

$$= \int \underbrace{\int c(\underline{\epsilon})\, p(\underline{\theta} | \underline{x})\, d\underline{\theta}}_{\text{minimize over } \hat{\underline{\theta}}}\, p(\underline{x})\, d\underline{x}$$

$$\int c(\underline{\epsilon})\, p(\underline{\theta} | \underline{x})\, d\underline{\theta} = \int_{\{\underline{\theta} : \|\underline{\epsilon}\| > \delta\}} p(\underline{\theta} | \underline{x})\, d\underline{\theta}$$

$$= \int_{\{\underline{\theta} : \|\underline{\theta} - \hat{\underline{\theta}}\| > \delta\}} p(\underline{\theta} | \underline{x})\, d\underline{\theta} = \int_{\{\underline{\theta} : \|\underline{\theta} - \hat{\underline{\theta}}\| \le \delta\}^c} p(\underline{\theta} | \underline{x})\, d\underline{\theta}$$

c denotes Complement

$$= 1 - \int_{\{\underline{\theta} : \|\underline{\theta} - \hat{\underline{\theta}}\| < \delta\}} p(\underline{\theta} | \underline{x})\, d\underline{\theta}$$

As $\delta \to 0$, 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

$$p(\underline{x} | \underline{\theta})\, p(\underline{\theta}) = p(\underline{x}, \underline{\theta})$$

If $\underline{\alpha} = \underline{A}\,\underline{\theta}$, $\quad \dfrac{\partial \underline{\alpha}}{\partial \underline{\theta}} = \underline{A}$

$$p(\underline{x}, \underline{\alpha}) = \frac{p(\underline{x}, \underline{\theta})}{\left|\det \dfrac{\partial \underline{\alpha}}{\partial \underline{\theta}}\right|} = \frac{p(\underline{x}, \underline{\theta})}{|\det \underline{A}|}$$

But $\underline{A}$ does not depend on $\underline{\alpha}$ and $\underline{\theta} = \underline{A}^{-1}\underline{\alpha}$

so that $\quad p(\underline{x}, \underline{\alpha}) = \dfrac{p_{x,\theta}(\underline{x}, \underline{A}^{-1}\underline{\alpha})}{|\det \underline{A}|}$

The MAP estimator of $\underline{\alpha}$ maximizes $p_{x,\theta}(\underline{x}, \underline{A}^{-1}\underline{\alpha})$ or because $\underline{\theta} = \underline{A}^{-1}\underline{\alpha}$ is invertible we can maximize $p(\underline{x}, \underline{\theta}) \Rightarrow$ maximizing value is $\hat{\underline{\theta}}$ and since $\underline{\alpha} = \underline{A}\underline{\theta}$ $\Rightarrow$ $\hat{\underline{\alpha}} = \underline{A}\hat{\underline{\theta}}$.

13) $\underline{x}^T\underline{C}^{-1}\underline{x} = \underline{x}^T\underline{D}^T\underline{D}\underline{x} = \underline{y}^T\underline{y}$ where $\underline{y} = \underline{D}\underline{x}$

But $\underline{C}_y = E(\underline{y}\,\underline{y}^T) = E(\underline{D}\underline{x}\,\underline{x}^T\underline{D}^T) = \underline{D}\underline{C}\underline{D}^T$

$= \underline{D}(\underline{D}^T\underline{D})^{-1}\underline{D}^T = \underline{I}$

$\Rightarrow$ $\underline{y} \sim N(\underline{0}, \underline{I})$ and

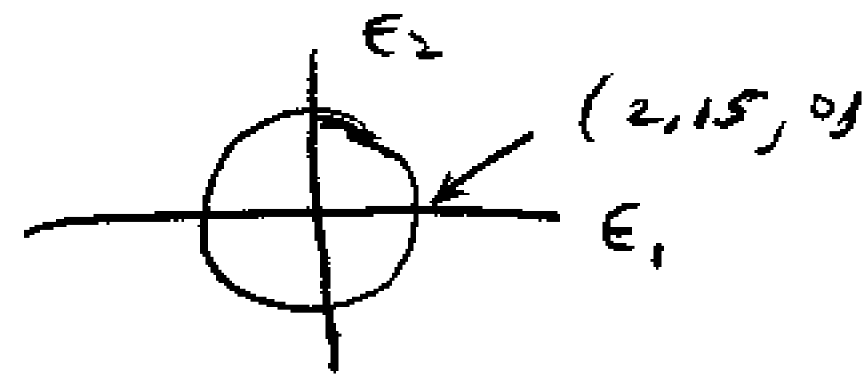
$\underline{y}^T\underline{y} = y_1^2 + y_2^2 \sim \chi_2^2$

since $\left.\begin{array}{l} y_1 \sim N(0,1) \\ y_2 \sim N(0,1) \end{array}\right\}$ independent

14) $\underline{\epsilon}^T \underline{M}_{\hat{\theta}}^{-1} \underline{\epsilon} = 2 \ln \dfrac{1}{1-P}$

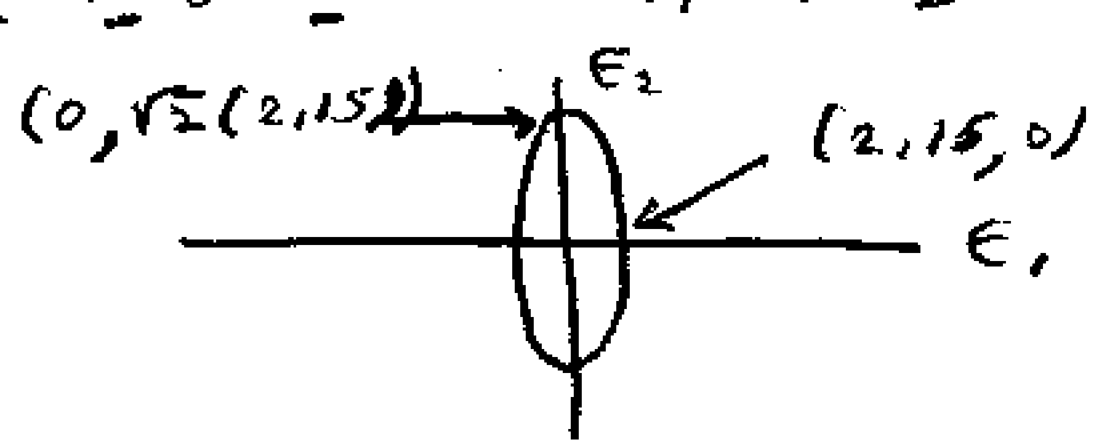
$= (2.15)^2$

a) $\underline{\epsilon}^T\underline{\epsilon} = (2.15)^2$



b) $\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}$

$\underline{\epsilon}^T \underline{M}_{\hat{\theta}}^{-1} \underline{\epsilon} = \epsilon_1^2 + \frac{1}{2}\epsilon_2^2 = (2.15)^2$



c) $\underline{M}_{\hat{\theta}}^{-1} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$

$$[\epsilon_1\ \epsilon_2]\ \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} = (2.15)^2$$

$$2\epsilon_1^2 + 2\epsilon_2^2 - 2\epsilon_1\epsilon_2 = 3(2.15)^2$$

$$\epsilon_1^2 - \epsilon_1\epsilon_2 + \epsilon_2^2 = \frac{3}{2}(2.15)^2$$

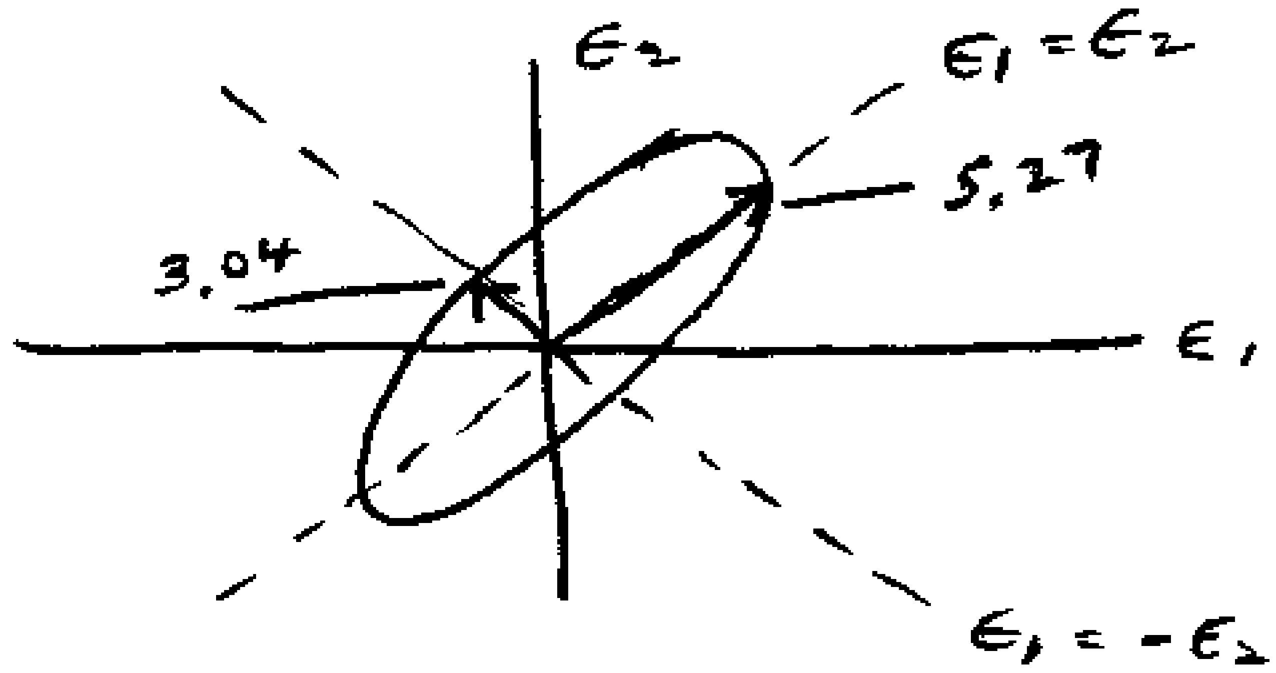
This is a rotated ellipse. It can be rewritten in standard form as

$$\frac{(\epsilon_1+\epsilon_2)^2}{4} + \frac{(\epsilon_1-\epsilon_2)^2}{4/3} = \frac{3}{2}(2.15)^2$$

or $$\frac{(\epsilon_1+\epsilon_2)^2}{a^2} + \frac{(\epsilon_1-\epsilon_2)^2}{b^2} = 1$$

where $a = 5.27$

$b = 3.04$



15) $\underline{M}_{\hat{\theta}} = (\underline{C}_\theta^{-1} + \underline{H}^T\underline{C}_W^{-1}\underline{H})^{-1} = (\underline{C}_\theta^{-1} + \underline{C}_W^{-1})^{-1}$

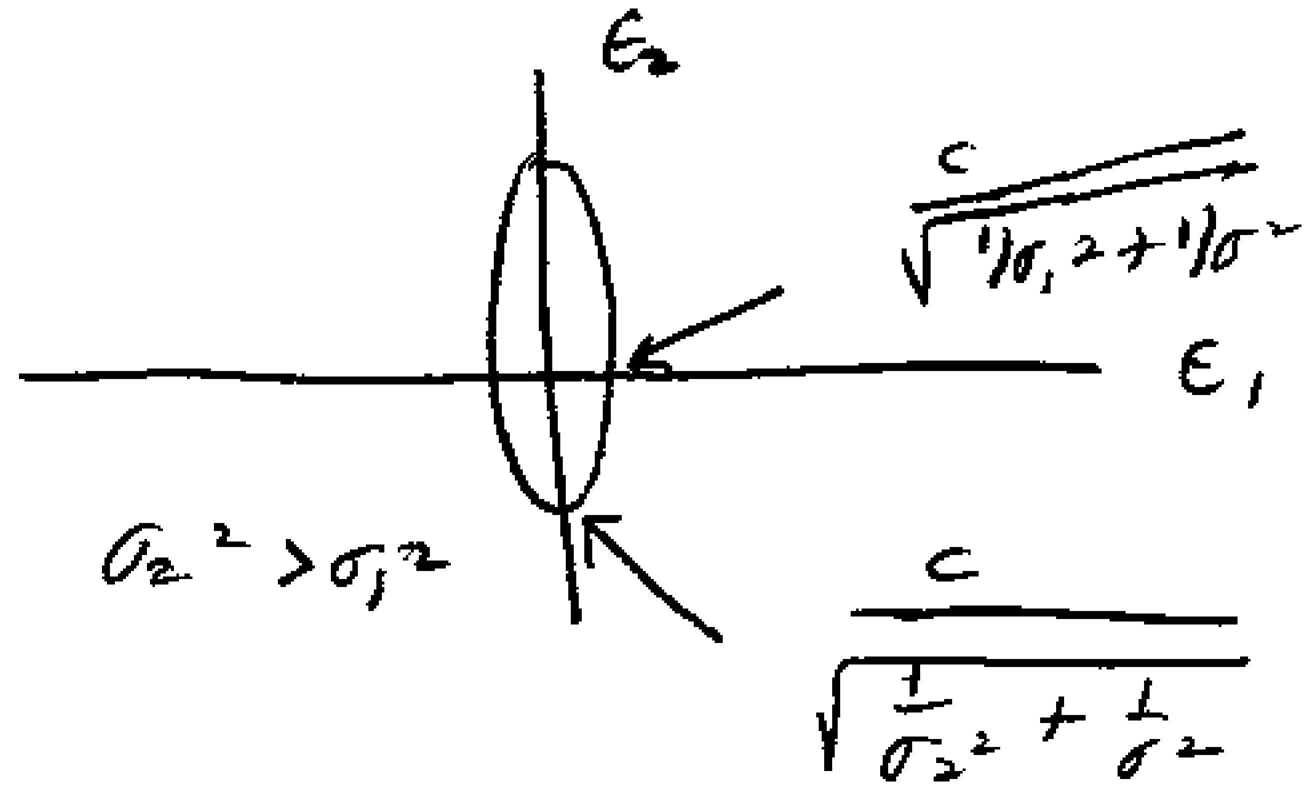
$$\underline{M}_{\hat{\theta}}^{-1} = \underline{C}_\theta^{-1} + \underline{C}_W^{-1} = \begin{bmatrix} 1/\sigma_1^2 + 1/\sigma^2 & 0 \\ 0 & 1/\sigma_2^2 + 1/\sigma^2 \end{bmatrix}$$

$$\underline{\epsilon}^T\underline{M}_{\hat{\theta}}^{-1}\underline{\epsilon} = 2\ln\frac{1}{1-P}$$

$$\epsilon_1^2\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma^2}\right) + \epsilon_2^2\left(\frac{1}{\sigma_2^2}+\frac{1}{\sigma^2}\right) = 2\ln\frac{1}{1-P}$$

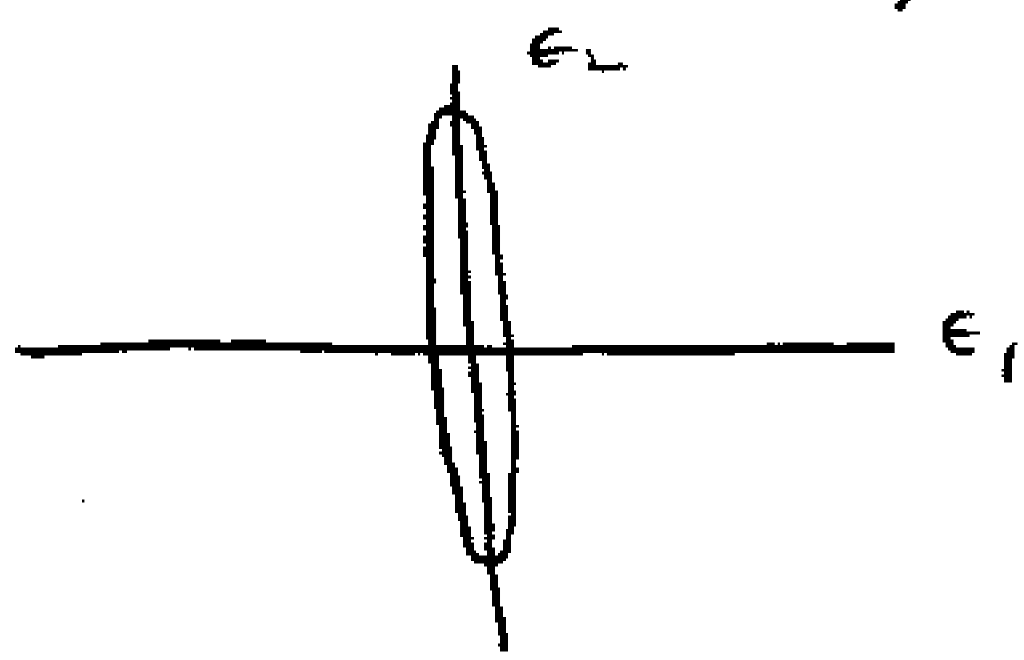
If $\sigma_1^2 = \sigma_2^2$, we have a circle.

If $\sigma_2^2 > \sigma_1^2$, we have an ellipse



$c = \sqrt{2 \ln \frac{1}{1-P}}$

For $\sigma_2^2 >> \sigma_1^2$ the ellipse appears as



Most of uncertainty is along ϵ_2 direction, as expected.

16) This is the Bayesian linear model.
From (11.38)

$$\hat{\underline{\theta}} = \underline{\mu}_\theta + (\underline{C}_\theta^{-1} + \underline{H}^T \underline{C}_w^{-1} \underline{H})^{-1} \underline{H}^T \underline{C}_w^{-1} (\underline{x} - \underline{H}\underline{\mu}_\theta)$$

where $\underline{\mu}_\theta = (A_0 \; B_0)^T$

$$\underline{C}_\theta = \begin{pmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{pmatrix}$$

$$\underline{H} = \begin{bmatrix} 1 & -M \\ 1 & -M+1 \\ \vdots & \vdots \\ 1 & M \end{bmatrix}$$

$$\underline{C}_w = \sigma^2 \underline{I}$$

$$\Rightarrow \underline{H}^T \underline{C}_w^{-1} \underline{H} = \frac{1}{\sigma^2} \underline{H}^T \underline{H} = \frac{1}{\sigma^2} \begin{bmatrix} N & 0 \\ 0 & \sum n^2 \end{bmatrix}$$

where $N = 2M+1$

$$(\underline{C}_\theta^{-1} + \underline{H}^T \underline{C}_w^{-1} \underline{H})^{-1} = \begin{bmatrix} \frac{1}{\sigma_A^2} + \frac{N}{\sigma^2} & 0 \\ 0 & \frac{1}{\sigma_B^2} + \frac{\sum n^2}{\sigma^2} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \dfrac{1}{\frac{1}{\sigma_A^2}+\frac{N}{\sigma^2}} & 0 \\ 0 & \dfrac{1}{\frac{1}{\sigma_B^2}+\frac{\sum n^2}{\sigma^2}} \end{bmatrix}$$

$$\underline{H}^T \underline{C}_W^{-1}(\underline{x}-\underline{H}\underline{\mu}_\theta) = \frac{1}{\sigma^2}(\underline{H}^T\underline{x} - \underline{H}^T\underline{H}\underline{\mu}_\theta)$$

$$\underline{H}^T\underline{H} = \begin{bmatrix} N & 0 \\ 0 & \sum n^2 \end{bmatrix}$$

$$\underline{H}^T\underline{C}_W^{-1}(\underline{x}-\underline{H}\underline{\mu}_\theta) = \frac{1}{\sigma^2}\begin{bmatrix} \sum x(n) - NA_0 \\ \sum n\,x(n) - \sum n^2 B_0 \end{bmatrix}$$

$$\hat{A} = A_0 + \frac{\frac{1}{\sigma^2}\left[\sum x(n) - NA_0\right]}{\frac{1}{\sigma_A^2}+\frac{N}{\sigma^2}}$$

$$= A_0 + \frac{N/\sigma^2}{\frac{1}{\sigma_A^2}+\frac{N}{\sigma^2}}(\bar{x}-A_0)$$

$$\hat{B} = B_0 + \frac{\frac{1}{\sigma^2}\left(\sum n\,x(n) - \sum n^2 B_0\right)}{\frac{1}{\sigma_B^2}+\frac{\sum n^2}{\sigma^2}}$$

$$= B_0 + \frac{\sum_{n=-M}^{M} n^2/\sigma^2}{\frac{1}{\sigma_B^2}+\sum_{n=-M}^{M} n^2/\sigma^2}\left[\frac{\sum_{n=-M}^{M} n\,x(n)}{\sum_{n=-M}^{M} n^2} - B_0\right]$$

$$\mathrm{Bmse}(\hat{A}) = \left[(\underline{C}_\theta^{-1} + \underline{H}^T\underline{C}_W^{-1}\underline{H})^{-1}\right]_{11}$$

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

$$\text{Bmse}(\hat{B}) = \left[\left(\underline{C}_\theta^{-1} + \underline{H}^T \underline{C}_W^{-1} \underline{H}\right)^{-1}\right]_{22}$$

$$= \left(\frac{1}{\sigma_B^2} + \sum_{n=-M}^{M} n^2 / \sigma^2 \right)^{-1}$$

The intercept $(\hat{A})$ will benefit most from prior knowledge since the reduction in Bmse due to the data is much larger for the slope $(\sum n^2/\sigma^2)$ than for the intercept (N/σ^2).

17) $$\underline{s} = \begin{bmatrix} 1 & -M \\ 1 & -M+1 \\ \vdots & \vdots \\ 1 & M \end{bmatrix} \underline{\theta} = \underline{H}\underline{\theta}$$

$$\hat{\underline{s}} = E(\underline{s} \mid \underline{x}) = E(\underline{H}\underline{\theta} \mid \underline{x}) = \underline{H}\hat{\underline{\theta}}$$

where $\hat{\underline{\theta}}$ is given in Prob 10.16.

$$\underline{\epsilon} = \underline{s} - \hat{\underline{s}} = \underline{s} - \underline{H}\hat{\underline{\theta}} = \underline{H}(\underline{\theta} - \hat{\underline{\theta}})$$

$$E(\underline{\epsilon}) = \underline{H}\left(E(\underline{\theta}) - E(\hat{\underline{\theta}})\right)$$

But $E(\hat{\underline{\theta}}) = \underline{\mu}_\theta \Rightarrow E(\underline{\epsilon}) = \underline{H}(\underline{\mu}_\theta - \underline{\mu}_\theta) = \underline{0}$

The covariance is

$$E(\underline{\epsilon}\underline{\epsilon}^T) = \underline{H}\, E\left[(\underline{\theta} - \hat{\underline{\theta}})(\underline{\theta} - \hat{\underline{\theta}})^T\right] \underline{H}^T$$

$$= \underline{H}\, \underline{M}_{\hat{\theta}}\, \underline{H}^T$$

$$= \begin{bmatrix} 1 & -M \\ \vdots & \vdots \\ 1 & M \end{bmatrix} \begin{bmatrix} \frac{1}{1/\sigma_A^2 + N/\sigma^2} & 0 \\ 0 & \frac{1}{\frac{1}{\sigma_B^2} + \frac{\sum n^2}{\sigma^2}} \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 \\ -M & \cdots & M \end{bmatrix}$$

$$= \frac{1}{\frac{1}{\sigma_A^2} + \frac{N}{\sigma^2}} \underline{1}\,\underline{1}^T + \frac{1}{\frac{1}{\sigma_B^2} + \frac{\sum n^2}{\sigma^2}} \underline{m}\,\underline{m}^T$$

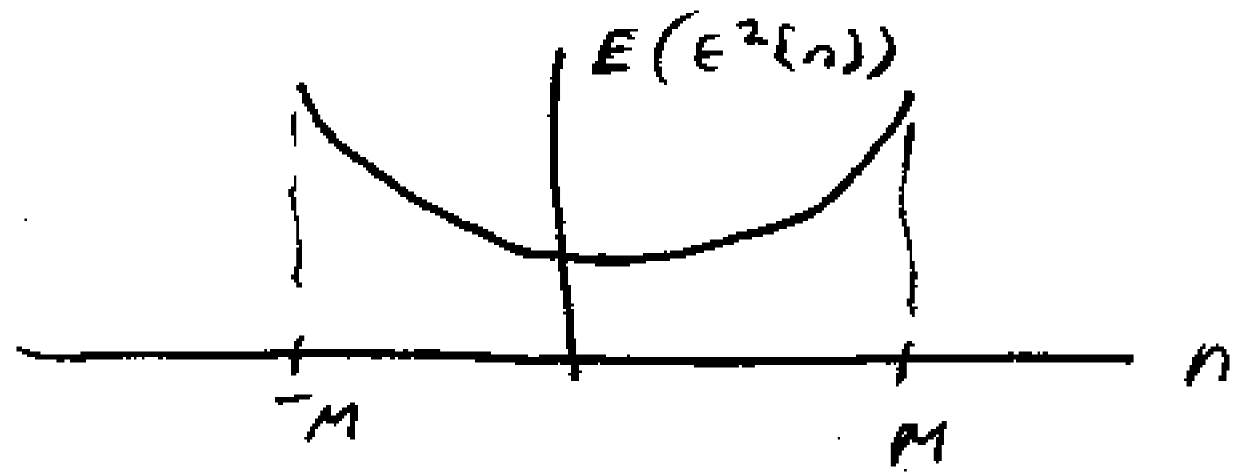
where $\underline{m} = [-M \dots M]^T$

Since $\underline{x}, \underline{\theta}$ are jointly Gaussian, $\underline{\epsilon}$ is also Gaussian. Note, however, that C_ϵ is singular, being of rank 2.

The mean squared error is for $\hat{s}(n)$:

$$[\underline{C}_\epsilon]_{n+M+1,\, n+M+1}$$

$$= \frac{1}{\frac{1}{\sigma_A^2} + \frac{N}{\sigma^2}} + \frac{1}{\frac{1}{\sigma_B^2} + \frac{\sum n^2}{\sigma^2}}\, n^2$$



As we depart from $n=0$ the signal estimation error increases. This is because any errors in B are magnified by n due to the signal dependence being Bn.

18) From (11.40)

$$\hat{\underline{s}} = C_s (C_s + \sigma^2 \underline{I})^{-1} \underline{x}$$
$$= [(\underline{C}_s + \sigma^2 \underline{I}) \underline{C}_s^{-1}]^{-1} \underline{x}$$
$$= (\underline{I} + \sigma^2 \underline{C}_s^{-1})^{-1} \underline{x}$$
$$(\underline{I} + \sigma^2 \underline{C}_s^{-1}) \hat{s} = \underline{x}$$
$$(C_s + \sigma^2 \underline{I}) \hat{\underline{s}} = \underline{C}_s \underline{x}$$

Let $\underline{x} = [\, x[-M] \ \dots \ x[M]\,]^T$

$\underline{s} = [\, s[-M] \ \dots \ s[M]\,]^T$ so that

$$[C_{\underline{s}}]_{ij} = E[\, s[i]\, s[j]\,] \qquad i,j = -M,\dots,0,\dots,M$$
$$= r_{ss}[i-j]$$

$$\Rightarrow \quad \sum_{j=-M}^{M} r_{xx}[i-j]\,\hat{s}[j] = \sum_{j=-M}^{M} r_{ss}[i-j]\,x[j]$$

for $i = -M, \dots, M$

As $M \to \infty \Rightarrow$

$$\sum_{j=-\infty}^{\infty} r_{xx}[i-j]\,\hat{s}[j] = \sum_{j=-\infty}^{\infty} r_{ss}[i-j]\,x[j]$$

$$-\infty < i < \infty$$

or $r_{xx}[n] \star \hat{s}[n] = r_{ss}[n] \star x[n]$

Taking Fourier transforms

$$P_{xx}(f)\,\hat{S}(f) = P_{ss}(f)\,X(f)$$

$$\Rightarrow \quad H(f) = \frac{P_{ss}(f)}{P_{xx}(f)} = \frac{P_{ss}(f)}{P_{ss}(f) + \sigma^2}$$

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), $H(f) \simeq 1$

If $\eta(f) \ll 1$ (low SNR), $H(f) \simeq 0$.

Chapter 12

1) $Bmse(\hat{\theta}) = E[(\theta - a x^2[0] - b x[0] - c)^2]$

$$\frac{\partial Bmse}{\partial a} = -2E[(\theta - a x^2[0] - b x[0] - c) x^2[0]] = 0$$

$$\frac{\partial Bmse}{\partial b} = -2E[(\quad '' \quad) x[0]] = 0$$

$$\frac{\partial Bmse}{\partial c} = -2E[(\quad '' \quad)] = 0$$

$$E(\theta x^2) = aE(x^4) + bE(x^3) + cE(x^2)$$
$$E(\theta x) = aE(x^3) + bE(x^2) + cE(x)$$
$$E(\theta) = aE(x^2) + bE(x) + c$$

$$\begin{bmatrix} E(x^4) & E(x^3) & E(x^2) \\ E(x^3) & E(x^2) & E(x) \\ E(x^2) & E(x) & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} E(\theta x^2) \\ E(\theta x) \\ E(\theta) \end{bmatrix}$$

The minimum MSE is

$$Bmse(\hat{\theta}) = E[\theta(\theta - \hat{a}x^2[0] - \hat{b}x[0] - \hat{c})]$$

$$= E(\theta^2) - \hat{a}E(\theta x^2) - \hat{b}E(\theta x) - \hat{c}E(\theta)$$

Now if $x[0] \sim U[-1/2, 1/2)$ and $\theta = \cos 2\pi x[0]$

$E(x) = 0, \quad E(\theta) = 0$

$$E(\theta x) = \int_{-\frac{1}{2}}^{\frac{1}{2}} x\cos 2\pi x \, dx = 0$$

$$E(\theta x^2) = \int_{-\frac{1}{2}}^{\frac{1}{2}} x^2 \cos 2\pi x \, dx = -\frac{1}{2\pi^2}$$

$$E(x^2) = \int_{-1/2}^{1/2} x^2 dx = \frac{1}{12}$$

$E(x^3) = 0$

$E(x^4) = \int_{-\frac{1}{2}}^{\frac{1}{2}} x^4 dx = \frac{1}{80}$

$$\begin{bmatrix} \frac{1}{80} & 0 & \frac{1}{12} \\ 0 & \frac{1}{12} & 0 \\ \frac{1}{12} & 0 & 1 \end{bmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{bmatrix} -\frac{1}{2\pi^2} \\ 0 \\ 0 \end{bmatrix}$$

$\Rightarrow \quad \hat{a} = -90/\pi^2$

$\hat{b} = 0$

$\hat{c} = 15/2\pi^2$

$Bmse(\hat{\theta}) = \frac{1}{2} - (-90/\pi^2)(-1/2\pi^2) - 0 - 0 = 0.04$

Using a linear estimator we have

$\hat{\theta} = b x(0) + c$

$$\begin{bmatrix} E(x^2) & E(x) \\ E(x) & 1 \end{bmatrix} \begin{pmatrix} b \\ c \end{pmatrix} = \begin{pmatrix} E(\theta x) \\ E(\theta) \end{pmatrix}$$

But $E(\theta x) = 0$, $E(\theta) = 0 \Rightarrow \hat{b} = \hat{c} = 0$ and $\hat{\theta} = 0$. The minimum MSE is $E(\theta^2) = 1/2$.

2) From (12.27)

$$\hat{A} = \mu_A + \left(1/\sigma_A^2 + \frac{\underline{h}^T \underline{h}}{\sigma^2} \right)^{-1} \frac{\underline{h}^T}{\sigma^2} (\underline{x} - \underline{h}\mu_A)$$

where $\underline{h} = [1\ r \ldots r^{N-1}]^T$

$$= \mu_A + \frac{\sum_{n=0}^{N-1} r^n (x[n] - r^n \mu_A)}{\frac{\sigma^2}{\sigma_A^2} + \sum_{n=0}^{N-1} r^{2n}}$$

$$Bmse(\hat{A}) = \frac{1}{1/\sigma_A^2 + \frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n}}$$

from (12.29), (12.30).

3) $\hat{x}_2 = a_1 x_1 + b$

Let $\theta = x_2$. Then from (12.6)

$$\hat{\theta} = E(\theta) + C_{\theta x_1} C_{x_1 x_1}^{-1} (x_1 - E(x_1))$$

$$= E(x_2) + \frac{E(x_1 x_2)}{E(x_1^2)} (x_1 - E(x_1))$$

$$Bmse(\hat{\theta}) = C_{\theta\theta} - C_{\theta x_1} C_{x_1 x_1}^{-1} C_{x_1 \theta} \quad \text{from (12.8)}$$

$$= E(x_2^2) - \frac{E^2(x_1 x_2)}{E(x_1^2)}$$

But $\underline{C}_{xx} = \begin{bmatrix} E(x_1^2) & E(x_1 x_2) \\ E(x_2 x_1) & E(x_2^2) \end{bmatrix}$

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

$$\hat{x}_1 = \sum_{i=2}^{N} a_i x_i$$

$$Bmse(\hat{x}_1) = E((x_1 - \hat{x}_1)^2)$$

$$= E\left[\left(x_1 - \sum_{i=2}^{N} a_i x_i\right)^2\right]$$

$$= E\left[(\underline{b}^T \underline{x})^2\right] \qquad \text{where} \quad b_1 = 1$$

$$b_i = -a_i$$

$$i = 2, \ldots, N$$

$$= \underline{b}^T \underline{C}_{xx} \underline{b}$$

But C_{xx} is positive semidefinite and for the BMSE to be zero we require $\underline{b}^T \underline{C}_{xx} \underline{b} = 0$ for some $\underline{b} \neq \underline{0}$. Hence, $\underline{C}_{xx}$ must be singular.

4) $(x, x) = E(x^2) \geq 0$ and $= 0$ if and only if $x = 0$

$(x, y) = (y, x)$ obvious

$$(c_1 x_1 + c_2 x_2, y) = E\left((c_1 x_1 + c_2 x_2) y\right)$$

$$= E(c_1 x_1 y + c_2 x_2 y)$$

$$= c_1 E(x_1 y) + c_2 E(x_2 y)$$

$$= c_1 (x_1, y) + c_2 (x_2, y)$$

5) $(x, x) = 0 \Rightarrow \mathrm{cov}(x, x) = 0 \Rightarrow \mathrm{var}(x) = 0$

$\not\Rightarrow x = 0$ but that x is a constant

6) Using (12.20)

$$\hat{\underline{s}} = \underline{C}_{sx} \underline{C}_{xx}^{-1} \underline{x}$$

$$\underline{C}_{xx} = E(\underline{x}\underline{x}^T) = \underline{R}_{ss} + \underline{R}_{ww} = (\sigma_s^2 + \sigma^2)\underline{I}$$

$$\underline{C}_{sx} = E(\underline{s}\underline{x}^T) = E(\underline{s}(\underline{s} + \underline{w})^T)$$

$$= E(\underline{s}\underline{s}^T) = \sigma_s^2 \underline{I}$$

$$\hat{\underline{s}} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma^2} \underline{x} \quad \text{or} \quad \hat{s}[n] = \frac{\sigma_s^2}{\sigma_s^2 + \sigma^2} x[n]$$

From (12.21)

$$\underline{M}_{\hat{s}} = \underline{C}_{ss} - \underline{C}_{sx}\underline{C}_{xx}^{-1}\underline{C}_{xs}$$

$$= \sigma_s^2 \underline{I} - \frac{(\sigma_s^2)^2}{\sigma_s^2+\sigma^2}\underline{I}$$

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

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

7) $$\hat{\underline{\theta}} = E(\underline{\theta}) + \underline{C}_{\theta x}\underline{C}_{xx}^{-1}(\underline{x}-E(\underline{x}))$$

$$\underline{M}_{\hat{\theta}} = E\left[(\underline{\theta}-\hat{\underline{\theta}})(\underline{\theta}-\hat{\underline{\theta}})^T\right]$$

$$= E\left[(\underline{\theta}-E(\underline{\theta})-\underline{C}_{\theta x}\underline{C}_{xx}^{-1}(\underline{x}-E(\underline{x})))(\underline{\theta}-E(\underline{\theta})-\underline{C}_{\theta x}\underline{C}_{xx}^{-1}(\underline{x}-E(\underline{x})))^T\right]$$

$$= \underline{C}_{\theta\theta} - E\left[(\underline{\theta}-E(\underline{\theta}))(\underline{x}-E(\underline{x}))^T\right]\underline{C}_{xx}^{-1}\underline{C}_{x\theta} - \underline{C}_{\theta x}\underline{C}_{xx}^{-1}E\left[(\underline{x}-E(\underline{x}))(\underline{\theta}-E(\underline{\theta}))^T\right] + \underline{C}_{\theta x}\underline{C}_{xx}^{-1}\underline{C}_{xx}\underline{C}_{xx}^{-1}\underline{C}_{x\theta}$$

$$= \underline{C}_{\theta\theta} - \underline{C}_{\theta x}\underline{C}_{xx}^{-1}\underline{C}_{x\theta} - \underline{C}_{\theta x}\underline{C}_{xx}^{-1}\underline{C}_{x\theta} + \underline{C}_{\theta x}\underline{C}_{xx}^{-1}\underline{C}_{x\theta}$$

$$= \underline{C}_{\theta\theta} - \underline{C}_{\theta x}\underline{C}_{xx}^{-1}\underline{C}_{x\theta}$$

$$Bmse(\hat{\theta}_i) = E\left[(\theta_i-\hat{\theta}_i)^2\right]$$

where the expectation is with respect to $p(\underline{x}, \theta_i)$

But $$Bmse(\hat{\theta}_i) = \int (\theta_i-\hat{\theta}_i)^2 p(\underline{x},\theta_i)\, d\theta_i\, d\underline{x}$$

$$= \int\cdots\int (\theta_i-\hat{\theta}_i)^2 p(\underline{x},\underline{\theta})\, d\underline{\theta}\, d\underline{x}$$

since $\hat{\theta}_i$ depends on $\underline{x}$ only

$$= \int \cdots \int \left[(\underline{\theta}-\hat{\underline{\theta}})(\underline{\theta}-\hat{\underline{\theta}})^T\right]_{ii} \, p(\underline{x},\underline{\theta}) \, d\underline{\theta} \, d\underline{x}$$

$$= \left[E\left((\underline{\theta}-\hat{\underline{\theta}})(\underline{\theta}-\hat{\underline{\theta}})^T\right)\right]_{ii} = [\underline{M}_{\hat{\theta}}]_{ii}$$

8) $$\hat{\underline{\alpha}} = E(\underline{\alpha}) + C_{\alpha x} C_{xx}^{-1}(\underline{x} - E(\underline{x}))$$

But $E(\underline{\alpha}) = \underline{A} E(\underline{\theta}) + \underline{b}$

$$\begin{aligned} C_{\alpha x} &= E\left[(\underline{\alpha} - E(\underline{\alpha}))(\underline{x} - E(\underline{x}))^T\right] \\ &= E\left[\underline{A}(\underline{\theta} - E(\underline{\theta}))(\underline{x} - E(\underline{x}))^T\right] \\ &= \underline{A}\, \underline{C}_{\theta x} \end{aligned}$$

$$\begin{aligned} \Rightarrow \hat{\underline{\alpha}} &= \underline{A}\, E(\underline{\theta}) + \underline{b} + \underline{A}\, \underline{C}_{\theta x} \underline{C}_{xx}^{-1}(\underline{x} - E(\underline{x})) \\ &= \underline{A}\,\hat{\underline{\theta}} + \underline{b} \end{aligned}$$

$$\hat{\underline{\alpha}} = E(\underline{\alpha}) + C_{\alpha x} C_{xx}^{-1}(\underline{x} - E(\underline{x}))$$

If $\underline{\alpha} = \underline{\theta}_1 + \underline{\theta}_2$,

$$E(\underline{\alpha}) = E(\underline{\theta}_1) + E(\underline{\theta}_2)$$

$$\begin{aligned} \underline{C}_{\alpha x} &= E\left[(\underline{\theta}_1 + \underline{\theta}_2 - E(\underline{\theta}_1) - E(\underline{\theta}_2))(\underline{x} - E(\underline{x}))^T\right] \\ &= E\left[(\underline{\theta}_1 - E(\underline{\theta}_1))(\underline{x} - E(\underline{x}))^T\right] \\ &\quad + E\left[(\underline{\theta}_2 - E(\underline{\theta}_2))(\underline{x} - E(\underline{x}))^T\right] \\ &= \underline{C}_{\theta_1 x} + \underline{C}_{\theta_2 x} \end{aligned}$$

$$\begin{aligned} \Rightarrow \hat{\underline{\alpha}} &= E(\underline{\theta}_1) + E(\underline{\theta}_2) + (\underline{C}_{\theta_1 x} + \underline{C}_{\theta_2 x}) \underline{C}_{xx}^{-1}(\underline{x} - E(\underline{x})) \\ &= \hat{\underline{\theta}}_1 + \hat{\underline{\theta}}_2 \end{aligned}$$

9) $$\hat{A}(N-1) = \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/N}\, \bar{x} + \frac{\sigma^2/N}{\sigma_A^2 + \sigma^2/N}\, \mu_A$$

$$\hat{A}[N] = \frac{N\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2}\, \frac{1}{N}\left(\sum_0^{N-1} x[n] + x[N]\right)$$

$$+ \quad \frac{\sigma^2}{(N+1)\sigma_A^2 + \sigma^2}\, \mu_A$$

$$= \frac{N\sigma_A^2}{(N+1)\sigma_A^2+\sigma^2} \quad \frac{\sigma_A^2+\sigma^2/N}{\sigma_A^2} \quad \underbrace{\frac{\sigma_A^2}{\sigma_A^2+\sigma^2/N} \quad \frac{1}{N}\sum_0^{N-1} x(n)}_{\hat{A}(N-1)}$$

$$+ \frac{\sigma_A^2}{(N+1)\sigma_A^2+\sigma^2}\, x[N] + \frac{N\sigma_A^2+\sigma^2}{(N+1)\sigma_A^2+\sigma^2} \quad \underbrace{\frac{\sigma^2}{N\sigma_A^2+\sigma^2}\,\mu_A}$$

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

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

$$+ \frac{\sigma_A^2}{(N+1)\sigma_A^2+\sigma^2}\, x(N)$$

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

$$+ \frac{\sigma_A^2}{(N+1)\sigma_A^2+\sigma^2}\, x[N]$$

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

Since $\mathrm{Bmse}(\hat{A}[N-1])$ for the case $\mu_A \neq 0$ 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.

$$e_1 = \frac{\begin{pmatrix}1\\1\\2\end{pmatrix}}{\sqrt{1+1+4}} = \begin{bmatrix}1/\sqrt{6}\\1/\sqrt{6}\\2/\sqrt{6}\end{bmatrix}$$

$$z_2 = x_2 - (x_2, e_1)e_1 = \begin{bmatrix}1\\0\\1\end{bmatrix} - (1\ 0\ 1)\begin{pmatrix}1/\sqrt{6}\\1/\sqrt{6}\\2/\sqrt{6}\end{pmatrix}\begin{pmatrix}1/\sqrt{6}\\1/\sqrt{6}\\2/\sqrt{6}\end{pmatrix}$$

$$= \begin{bmatrix}1\\0\\1\end{bmatrix} - \frac{3}{\sqrt{6}}\begin{pmatrix}1/\sqrt{6}\\1/\sqrt{6}\\2/\sqrt{6}\end{pmatrix} = \begin{bmatrix}1\\0\\1\end{bmatrix} - \begin{bmatrix}1/2\\1/2\\1\end{bmatrix} = \begin{bmatrix}1/2\\-1/2\\0\end{bmatrix}$$

$$e_2 = \frac{\begin{bmatrix}1/2\\-1/2\\0\end{bmatrix}}{\sqrt{1/4+1/4}} = \begin{pmatrix}1/\sqrt{2}\\-1/\sqrt{2}\\0\end{pmatrix}$$

$$z_3 = x_3 - (x_3, e_1)\, e_1 - (x_3, e_2)\, e_2$$

$$= \begin{bmatrix}1/3\\1/3\\-1/3\end{bmatrix}$$

$$e_3 = \begin{bmatrix}1/\sqrt{3}\\1/\sqrt{3}\\-1/\sqrt{3}\end{bmatrix}$$

(ii) $$y_1 = \frac{x_1}{\sqrt{\mathrm{var}(x_1)}} = x_1$$

$$z_2 = x_2 - (x_2, y_1)\, y_1 = x_2 - E(x_2 y_1)\, y_1$$

$$= x_2 - E(x_1 x_2)\, x_1 = x_2 - \rho x_1$$

$$y_2 = \frac{x_2 - \rho x_1}{\sqrt{E((x_2 - \rho x_1)^2)}} = \frac{x_2 - \rho x_1}{\sqrt{1 - 2\rho^2 + \rho^2}} = \frac{x_2 - \rho x_1}{\sqrt{1-\rho^2}}$$

$$z_3 = x_3 - (x_3, y_1)\, y_1 - (x_3, y_2)\, y_2$$

$$= x_3 - E(x_3 x_1)x_1 - E\left(x_3 \frac{x_2 - \rho x_1}{\sqrt{1-\rho^2}}\right) \frac{x_2 - \rho x_1}{\sqrt{1-\rho^2}}$$

$$= x_3 - \rho^2 x_1 - \frac{\rho - \rho^3}{\sqrt{1-\rho^2}} \frac{x_2 - \rho x_1}{\sqrt{1-\rho^2}}$$

$$= x_3 - \rho^2 x_1 - \rho(x_2 - \rho x_1) = x_3 - \rho x_2$$

$$y_3 = \frac{x_3 - \rho x_2}{\sqrt{E((x_3 - \rho x_2)^2)}} = \frac{x_3 - \rho x_2}{\sqrt{1-\rho^2}}$$

$$\underline{y} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ \frac{-\rho}{\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} & 0 \\ 0 & \frac{-\rho}{\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} \end{bmatrix}}_{\underline{A}} \underline{x}$$

$\underline{A}$ is lower triangular and will be in general.

Since $\underline{C}_{yy} = \underline{I} \Rightarrow \underline{C}_{yy} = \underline{A}\,\underline{C}_{xx}\underline{A}^T$

$\underline{C}_{xx} = \underline{A}^{-1}\underline{A}^{T^{-1}} \Rightarrow \underline{C}_{xx}^{-1} = \underline{A}^T\underline{A}$

Thus, $\underline{A}$ is a whitening transformation and $\underline{A}^T\underline{A}$ provides a Cholesky decomposition of $\underline{C}_{xx}^{-1}$.

12) From (12.47) for a scalar parameter so that $M[n-1]$ and $b[n]$ are scalars, as $\sigma_n^2 \to 0$

$$K[n] \to \frac{M[n-1]\,h[n]}{M[n-1]\,h^2(n)} = \frac{1}{h[n]}$$

$$\Rightarrow \quad \hat{\theta}[n] = \hat{\theta}[n-1] + \frac{1}{h[n]}\left(x[n] - h[n]\,\hat{\theta}[n-1]\right)$$
$$= x(n)/h[n]$$

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 $h[n] = 1$, $\sigma_n^2 = \sigma^2$

$$\Rightarrow \quad \hat{A}(n) = \hat{A}(n-1) + K(n)\left(x[n] - \hat{A}(n-1)\right)$$

$$K[n] = \frac{M[n-1]}{\sigma^2 + M[n-1]}$$

$$M[n] = (1 - K[n])\,M[n-1]$$

This is the same as the introductory example of Sect 12.6 except for the initialization since

$$\hat{A}[-1] = E(A) = 0$$

$$M[-1] = E\left[(A - \hat{A}[-1])^2\right] = E(A^2) = A_0^2/3$$

Solving for $K(n)$:

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

$$= \frac{\sigma^2 M(n-1)}{\sigma^2 + M(n-1)} = K[n]\,\sigma^2$$

$$\Rightarrow \quad K[n] = \frac{K[n-1]\,\sigma^2}{\sigma^2 + K[n-1]\,\sigma^2} = \frac{K[n-1]}{1+K[n-1]}$$

$$\frac{1}{K[n]} = \frac{1}{K[n-1]} + 1 \qquad n \geq 1$$

$$K[0] = \frac{M[-1]}{\sigma^2 + M[-1]} = \frac{A_0^2/3}{\sigma^2 + A_0^2/3}$$

$$\frac{1}{K[0]} = 1 + \frac{\sigma^2}{A_0^2/3}$$

$$\frac{1}{K[n]} = \frac{1}{K[0]} + n = (n+1) + \frac{\sigma^2}{A_0^2/3}$$

$$K[n] = \frac{A_0^2/3}{(n+1)A_0^2/3 + \sigma^2}$$

$$1 - K[n] = \frac{n A_0^2/3 + \sigma^2}{(n+1)A_0^2/3 + \sigma^2}$$

$$\hat{A}[n] = (1 - K[n])\,\hat{A}[n-1] + K[n]\,x[n]$$

$$= \frac{n A_0^2/3 + \sigma^2}{(n+1)A_0^2/3 + \sigma^2}\,\hat{A}[n-1] + \frac{A_0^2/3}{(n+1)A_0^2/3 + \sigma^2}\,x[n]$$

where $\hat{A}[-1] = 0$

$$\hat{A}[0] = \frac{A_0^2/3}{A_0^2/3 + \sigma^2}\,x[0]$$

$$\hat{A}[1] = \frac{A_0^2/3 + \sigma^2}{2A_0^2/3 + \sigma^2}\;\frac{A_0^2/3}{A_0^2/3 + \sigma^2}\,x[0]$$

$$+ \quad \frac{A_0^2/3}{2A_0^2/3+\sigma^2}\, x(1)$$

$$= \quad \frac{A_0^2/3}{2A_0^2/3+\sigma^2}\,(x(0)+x(1))$$

$$= \quad \frac{A_0^2/3}{A_0^2/3+\sigma^2/2}\,\underbrace{\tfrac{1}{2}(x(0)+x(1))}_{\bar{x}}$$

etc.

14) To minimize $E[(x(n)-\hat{x}(n))^2]$ we use the orthogonality principle or

$$E[(x(n)-\hat{x}[n])\,x[n-l]] = 0 \qquad l=-M,\dots,M \quad l\neq 0$$

$$r_{xx}[l] = E\Big[\sum_k a_k x[n-k]\,x[n-l]\Big] = \sum_k a_k\, r_{xx}[l-k]$$

To show that $a_{-k}=a_k$:

Let $k'=-k$

$$r_{xx}[l] = \sum_{\substack{k'=-M\\k'\neq 0}}^{M} a_{-k'}\, r_{xx}[l+k'] \qquad l=-M,\dots,M \quad l\neq 0$$

Let $l'=-l$

$$r_{xx}[-l'] = \sum_{\substack{k'=-M\\k'\neq 0}}^{M} a_{-k'}\, r_{xx}[-l'+k'] \qquad l'=-M,\dots,M \quad l'\neq 0$$

$$r_{xx}[l'] = \sum_{\substack{k'=-M\\k'\neq 0}}^{M} a_{-k'}\, r_{xx}[l'-k']$$

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) $$W = \frac{r_{ss}[0]}{r_{ss}[0] + r_{ww}[0]} = \eta/_{\eta+1}$$

$$M\hat{s} = \left(1 - \frac{\eta}{\eta+1}\right) r_{ss}[0]$$

But $$\rho = \frac{E(s[0](s[0] + w[0]))}{\sqrt{r_{ss}[0](r_{ss}[0] + r_{ww}[0])}}$$

$$= \frac{r_{ss}[0]}{\sqrt{r_{ss}[0](r_{ss}[0] + r_{ww}[0])}}$$

$$= \frac{1}{\sqrt{1 + 1/\eta}} = \sqrt{\eta/_{\eta+1}}$$

$$\Rightarrow \quad W = \rho^2 \quad \text{or} \quad \hat{s}[0] = \rho^2 x[0]$$

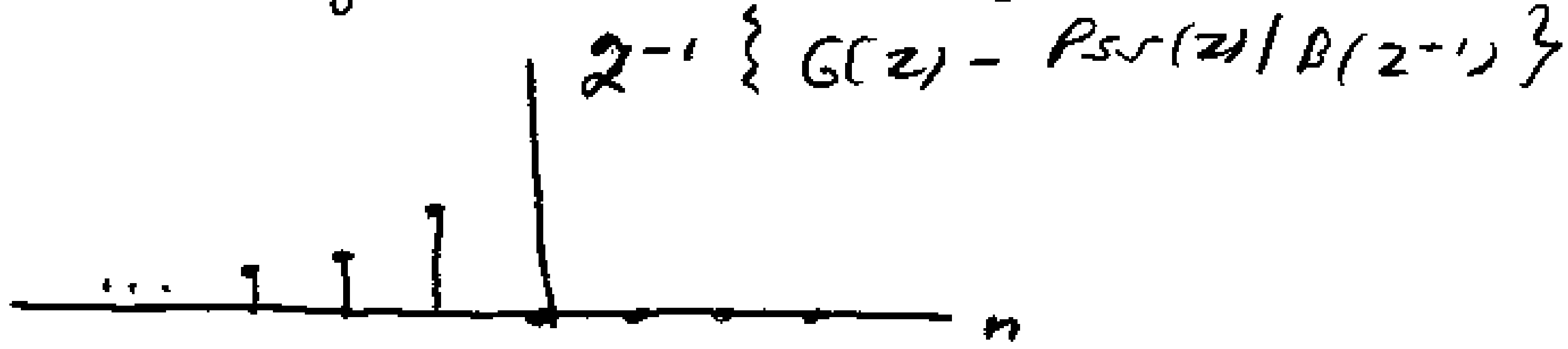
$$M\hat{s} = (1-\rho^2)\, r_{ss}[0]$$

The higher the SNR, η, the larger is ρ and hence the better is the estimator.

16) $\left[B(z^{-1}) \left(G(z) - \frac{P_{ss}(z)}{B(z^{-1})} \right) \right]_+ = 0$

But $B(z^{-1})$ is the z-transform of an anticausal sequence and thus if $G(z) - \frac{P_{ss}(z)}{B(z^{-1})}$ is also anticausal

with a sequence value of zero at $n = 0$ or



the convolution will be zero for $n \geq 0$ as required. Now, since $P_{ss}(z)/B(z^{-1})$ is a two-sided sequence we let

$$\mathcal{Z}^{-1}\{G(z)\} = \mathcal{Z}^{-1}\left\{ \frac{P_{ss}(z)}{B(z^{-1})} \right\} \text{ for } n \geq 0$$

or $\quad G(z) = \left[\frac{P_{ss}(z)}{B(z^{-1})} \right]_+$

or $\quad H(z) = \frac{1}{B(z)} \left[P_{ss}(z) / B(z^{-1}) \right]_+$

17) $E\{(s[n] - \hat{s}[n])\, x[n-\ell]\} = 0 \qquad -\infty < \ell < \infty$

$E\left(\left(s[n] - \sum_k h[k]\, x[n-k]\right) x[n-\ell]\right) = 0$

$$E(s[n]\,x[n-\ell]) = \sum_k h[k]\, E[x[n-k]\, x[n-\ell]]$$

$$r_{ss}[\ell] = \sum_{k=-\infty}^{\infty} h[k]\, r_{xx}[\ell - k]$$

18) $M\hat{s} = E\{(s[n]-\hat{s}[n])^2\}$

$$= E\left[(s[n]-\hat{s}[n])\left(s[n]-\sum_k h[k]x[n-k]\right)\right]$$

$$= E[(s[n]-\hat{s}[n])s[n]]$$

$$- \underbrace{E\left[(s[n]-\hat{s}[n])\sum_k h[k]x[n-k]\right]}_{=0 \text{ by orthogonality principle}}$$

$$= r_{ss}[0] - \sum_k h[k] \underbrace{E[x[n-k]s[n]]}_{r_{ss}[k]}$$

Now by Parseval's theorem

$$\sum_k h[k] r_{ss}[k] = \int P_{ss}(f) H(f) df$$

$$\Rightarrow \quad M\hat{s} = \int P_{ss}(f)df - \int P_{ss}(f)H(f)df$$

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} P_{ss}(f)(1-H(f))df$$

For the example given

$$H(f) = \begin{cases} \dfrac{P_0}{P_0+\sigma^2} & |f| \le 1/4 \\ 0 & \frac{1}{4} < |f| \le \frac{1}{2} \end{cases}$$

$$M\hat{s} = \int_{-1/4}^{1/4} P_0\left(1-\frac{P_0}{P_0+\sigma^2}\right)df = \frac{1}{2}\,\frac{P_0\sigma^2}{P_0+\sigma^2}$$

Since there is no signal power above $f = 1/4$, $H(f) = 0$. For $f \le 1/4$ 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) $$\hat{x}[n] = \sum_{k=1}^{N} h[k]\, x[n-k]$$

$$E\left[(x[n] - \hat{x}[n])\, x[n-l] \right] = 0 \qquad l = 1, 2, \ldots, N$$

$$r_{xx}[l] = \sum_{k=1}^{N} h[k]\, E\left(x[n-k] x[n-l] \right)$$

$$= \sum_{k=1}^{N} h[k]\, r_{xx}[l-k]$$

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.

$$M_Q = E\left[(x[n] - \hat{x}[n])\, x[n] \right] \underbrace{- E\left[(x[n] - \hat{x}[n])\, \hat{x}[n] \right]}_{= 0 \text{ by orthogonality principle}}$$

$$= E\left[x^2[n] \right] - \sum_{k=1}^{N} h[k]\, E\left[[n-k]\, x[n] \right]$$

$$= r_{xx}[0] - \sum_{k=1}^{N} h[k]\, r_{xx}[k]$$

20) From the previous problem we have

$$r_{xx}[l] = \sum_{k=1}^{N} h[k]\, r_{xx}[l-k] \qquad l = 1, 2, \ldots, N$$

must be solved for the optimal one-step predictor. But for an AR(N) process we know that (see Appendix 1)

$$r_{xx}[l] = -\sum_{k=1}^{N} a[k]\, r_{xx}[l-k] \qquad l \geq 1$$

which are the Yule-Walker equations. Since the solution for the $h[k]$'s is unique,

$$h[k] = -a[k]$$

so that $\hat{x}[n] = -\sum_{k=1}^{N} a(k)\, x[n-k]$ and the MMSE is

$$M_{\hat{x}} = r_{xx}[0] - \sum_{k=1}^{N} h[k]\, r_{xx}[k]$$

$$= r_{xx}[0] + \sum_{k=1}^{N} a[k]\, r_{xx}[k]$$

$$= \sigma_u^2 \quad \text{(see Appendix 1)}$$

Chapter 13

1) $s[n] = a^{n+1} s[-1] + \sum_{k=0}^{n} a^k u[n-k]$

Let $\underline{s} = [s[n_1]\ s[n_2]\ \ldots\ s[n_k]]^T$

Assume $n_k > n_{k-1} > \ldots > n_1$

$$\underline{s} = \begin{bmatrix} a^{n_1+1} & a^{n_1} & \ldots & 1 & 0 & 0 & \ldots & 0 \\ a^{n_2+1} & a^{n_2} & \ldots & & 1 & 0 & \ldots & 0 \\ \vdots & & & & & & & \\ a^{n_k+1} & a^{n_k} & \ldots & & & & & 1 \end{bmatrix} \begin{bmatrix} s[-1] \\ u[0] \\ \vdots \\ u[n_k] \end{bmatrix}$$

Since $s[-1], u[0], \ldots, u[n_k]$ are independent, they are jointly Gaussian and thus $\underline{s}$ has a multivariate Gaussian PDF, being a linear transformation.

2) If $\mu_s = 0$, from (13.4), $E(s[n]) = 0$

$$C_s[m, n] = a^{m+n+2}\sigma_s^2 + \sigma_u^2 a^{m-n} \cdot \sum_{k=0}^{n} a^{2k}$$

$$= \frac{a^{m+n+2}\sigma_u^2}{1-a^2} + \frac{\sigma_u^2 a^{m-n}(1-a^{2(n+1)})}{1-a^2}$$

$$= \frac{\sigma_u^2}{1-a^2} a^{m-n}$$

$\Rightarrow$ WSS. By setting the initial conditions as given the process is in steady-state for $n = -1$.

3) $$E(s[n]) = a^{n+1}\mu_s = 5(0.98)^{n+1}$$

$$\mathrm{var}(s[n]) = a^{2n+2}\sigma_s^2 + \sigma_u^2 \sum_{k=0}^{n} a^{2k}$$

$$= (0.98)^{2n+2} + 0.1 \sum_{k=0}^{n} (0.98)^{2k}$$

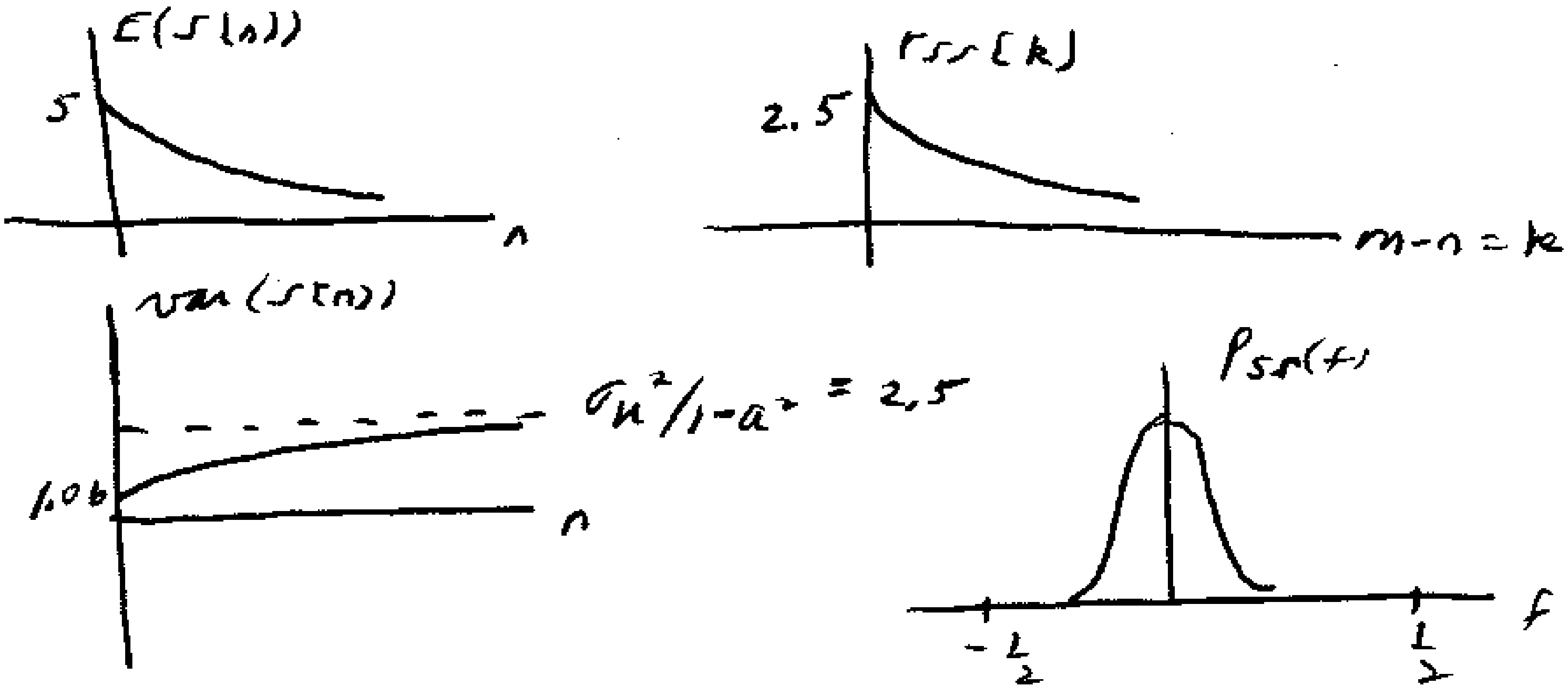
$$c_s[m,n] \to \frac{\sigma_u^2}{1-a^2} a^{m-n}$$

$$= \frac{0.1}{1-0.98^2}\, 0.98^{m-n}$$

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)

$$c_s[m,n] = a^{m-n} \underbrace{\left[a^{2n+2}\sigma_s^2 + \sigma_u^2 \sum_{k=0}^{n} a^{2k} \right]}_{\mathrm{var}(s[n])}$$

$$c_s[m,n] = a^{m-n} c_s[n,n]$$

5) $\underline{s}[n] = \underline{A}\,\underline{s}[n-1] + \underline{B}\,\underline{u}[n]$

$$E(\underline{s}[n]) = \underline{A}\,E(\underline{s}[n-1]) + \underline{B}\,E(\underline{u}[n])$$
$$= \underline{A}\,E(\underline{s}[n-1])$$

$$\underline{C}[n] = E[(\underline{s}[n] - E(\underline{s}[n]))(\underline{s}[n] - E(\underline{s}[n]))^T]$$
$$= E[(\underline{A}\,\underline{s}[n-1] + \underline{B}\,\underline{u}[n] - \underline{A}\,E(\underline{s}[n-1]))$$
$$(\qquad '' \qquad)^T]$$

$$= E[(\underline{A}(\underline{s}[n-1] - E(\underline{s}[n-1])) + \underline{B}\,\underline{u}[n])$$
$$(\qquad '' \qquad)^T)$$

$$= \underline{A}\,\underline{C}[n-1]\underline{A}^T + \underline{B}\,\underline{Q}\,\underline{B}^T \quad \text{since}$$
$$E(\underline{s}[n-1]\,\underline{u}[n]^T) = \underline{0}$$
($\underline{s}[n-1]$ depends on $\underline{u}[k]$ for $k \le n-1$).

6) From (3.14) $E(\underline{s}[n]) = \underline{A}^{n+1}\mu_s$

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 $\underline{A} = \underline{V}\,\underline{\Lambda}\,\underline{V}^T$

$$\Rightarrow \underline{A}^{n+1} = \underline{V}\,\underline{\Lambda}^{n+1}\underline{V}^T$$
$$E(\underline{s}[n]) = \underline{V}\,\underline{\Lambda}^{n+1}\underline{V}^T\underline{\mu}_s$$
$$= \sum_{i=1}^{p} a_i \lambda_i^{n+1} \underline{v}_i$$

where $a_i = (\underline{V}^T \underline{u}_s)_i$

$\underline{V} = (\underline{V}_1\ \underline{V}_2 \dots \underline{V}_p)$

If any $|\lambda_i| > 1$ $\quad E(\underline{s}(n)) \to \infty$

If all $|\lambda_i| < 1$ $\quad E(\underline{s}(n)) \to \underline{0}$

7)
$$\underline{A}^n = (\underline{V}\,\underline{\Lambda}\,\underline{V}^T)^n = \underline{V}\,\underline{\Lambda}^n\,\underline{V}^T$$
$$\underline{A}^n \underline{C}_s \underline{A}^{nT} = \underline{V}\,\underline{\Lambda}^n\,\underline{V}^T \underline{C}_s \underline{V}\,\underline{\Lambda}^n\,\underline{V}^T$$
$$\underline{e}_i^T \underline{A}^n \underline{C}_s \underline{A}^{nT} \underline{e}_j = \underbrace{\underline{e}_i^T \underline{V}\,\underline{\Lambda}^n}_{\underline{b}^T}(\underline{V}^T \underline{C}_s \underline{V})\underbrace{\underline{\Lambda}^n \underline{V}^T \underline{e}_j}_{\underline{a}}$$

But $\underline{b} = \underline{\Lambda}^n \underline{V}\,\underline{e}_i \to \underline{0}$ if $|\lambda_i| < 1$

$\underline{a} = \underline{\Lambda}^n \underline{V}\,\underline{e}_j \to \underline{0}$ if $|\lambda_i| < 1$

$\Rightarrow$ $\quad \underline{b}^T \underline{V}^T \underline{C}_s \underline{V}\,\underline{a} = [\underline{A}^n \underline{C}_s \underline{A}^{nT}]_{ij} \to 0$

for all i, j

8)
$$\underbrace{\begin{bmatrix} r(n-p+1) \\ r(n-p+2) \\ \vdots \\ r(n) \end{bmatrix}}_{\underline{s}(n)} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & & & & \\ -a(p) & -a(p-1) & \dots & & -a(1) \end{bmatrix}}_{\underline{A}} \underbrace{\begin{bmatrix} r(n-p) \\ r(n-p+1) \\ \vdots \\ r(n-1) \end{bmatrix}}_{\underline{s}(n-1)}$$

$$+ \underbrace{\begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ & & & \\ b(q) & b(q-1) & \dots & 1 \end{bmatrix}}_{\underline{B}} \underbrace{\begin{bmatrix} u(n-q) \\ u(n-q+1) \\ \vdots \\ u(n) \end{bmatrix}}_{\underline{u}(n)}$$

Not a Gauss-Markov process since $\underline{u}[n]$ is not vector WGN. This is because $\underline{u}[n]$ is correlated in time. If $q=1$ for example

$$E(\underline{u}[n]\,\underline{u}[n+1]^T) = E\begin{bmatrix} u[n-1] \\ u[n] \end{bmatrix}[u[n]\;u[n+1]]$$

$$= \begin{bmatrix} E[u[n-1]u[n]] & E[u[n-1]u[n+1]] \\ E[u^2[n]] & E[u[n]\,u[n+1]] \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 \\ \sigma_u^2 & 0 \end{bmatrix} \neq \underline{0}$$

9) $$r[n] = s[n] + \sum_{l=1}^{q} b[l]s[n-l]$$

$$\sum_{k=0}^{p} a[k]r[n-k] = \sum_{k=0}^{p} a[k]s[n-k] + \sum_{l=1}^{q} b[l] \sum_{k=0}^{p} a[k]s[n-k+l]$$

$$= u[n] + \sum_{l=1}^{q} b[l]\,u[n-l]$$

Now let the state be given by (13.10)

$$\Rightarrow \quad \underline{s}[n] = \underline{A}\,\underline{s}[n-1] + \underline{B}\,u[n]$$ (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 $q \le p-1$

$$r[n] = \underbrace{(0 \ldots 0\; b[q] \ldots b[1]\; 1)}_{h^T[n] = h^T}\,\underline{s}[n]$$

$\Rightarrow x(n) = \underline{h}^T \underline{s}[n] + \underline{w}[n]$

Now use (13.51) - (13.56) with

$$\underline{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ & & & & \\ -a(p) & -a(p-1) & \cdots & & -a[1] \end{bmatrix} \qquad \underline{B} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

$$\underline{Q} = \sigma_u^2$$

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

$$\hat{A}[n] = \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/n+1} \frac{1}{n+1} \sum_{k=0}^{n} x[k]$$

$$M[n-1] = \frac{\sigma_A^2 \sigma^2}{n\sigma_A^2 + \sigma^2}$$

$$K[n] = \frac{\sigma_A^2}{(n+1)\sigma_A^2 + \sigma^2}$$

11) From (13.39), (13.40), (13.42)

$$M[n|n-1] = 0.81\, M[n-1|n-1] + 1$$

$$K[n] = \frac{M[n|n-1]}{\sigma_n^2 + M[n|n-1]}$$

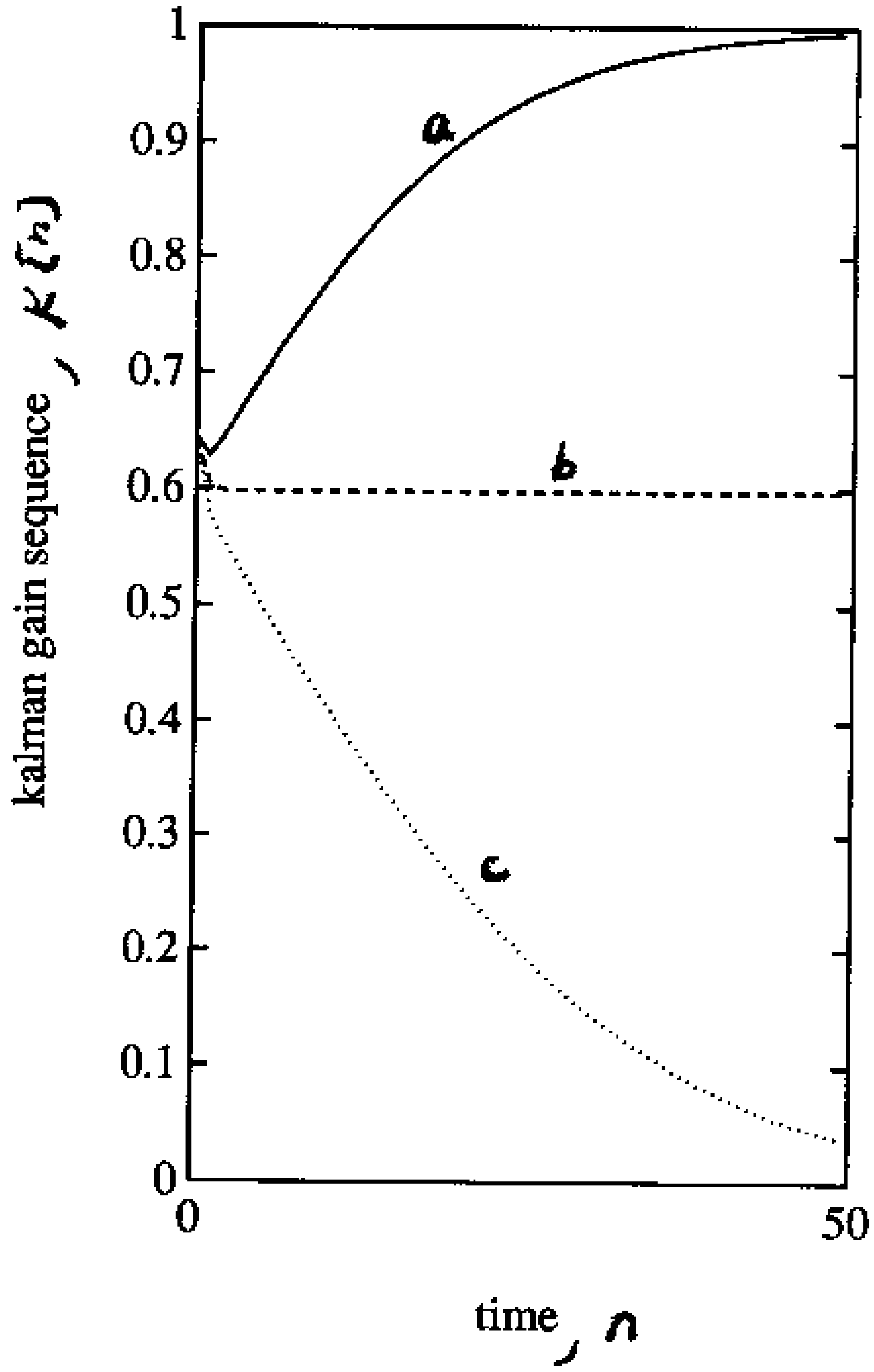
$$M[n|n] = (1 - K[n])\, M[n|n-1]$$

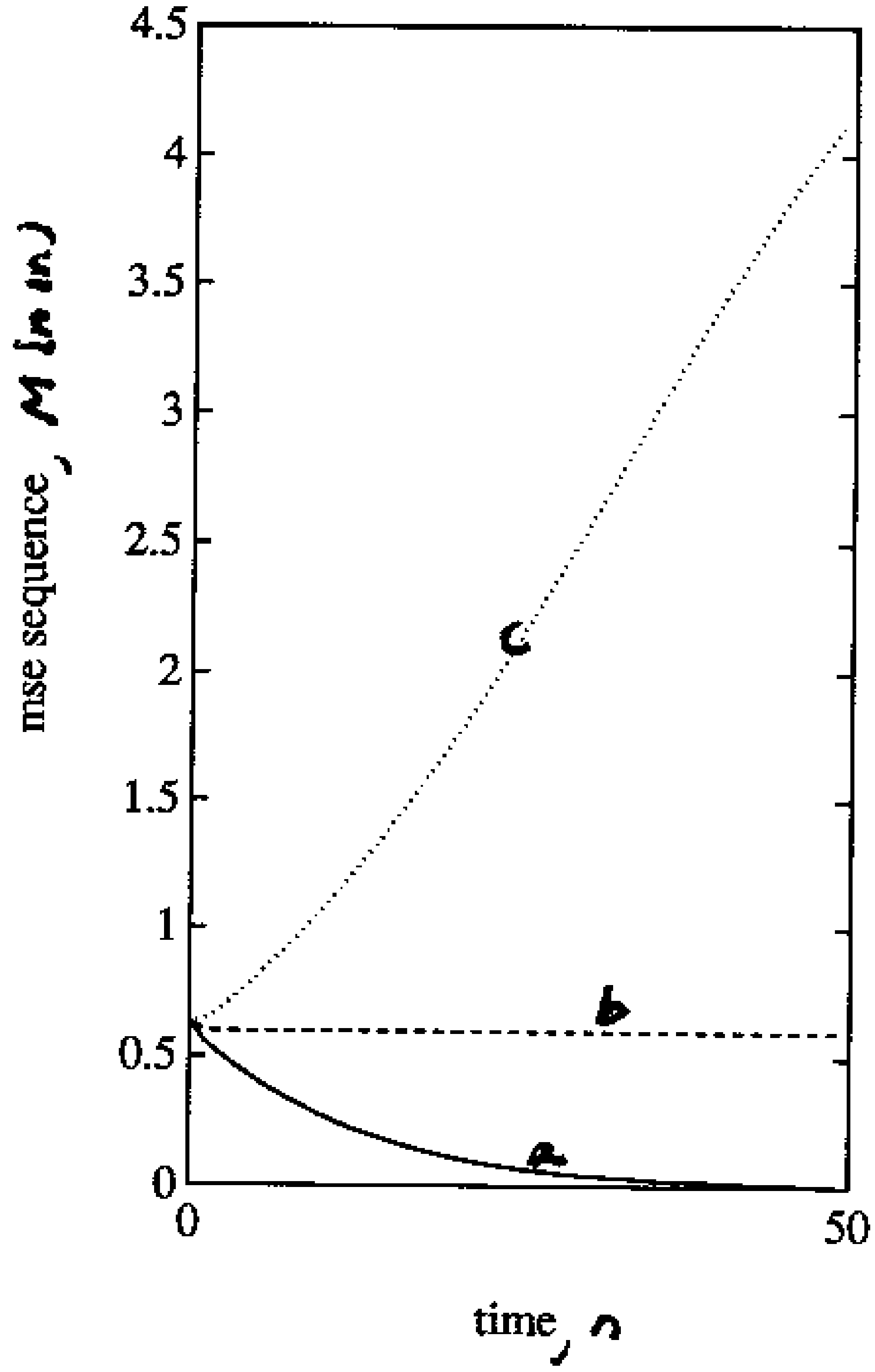
where $M[-1|-1] = \sigma_s^2 = 1$

See next page for plots. For (a) the gain $\rightarrow 1$ since the observations become less noisy and thus $\hat{s}[n|n] \rightarrow x[n]$. 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 $\hat{s}[n|n] \rightarrow \hat{s}[n|n-1]$.

$$= 0.9\, \hat{s}[n-1|n-1]$$

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.

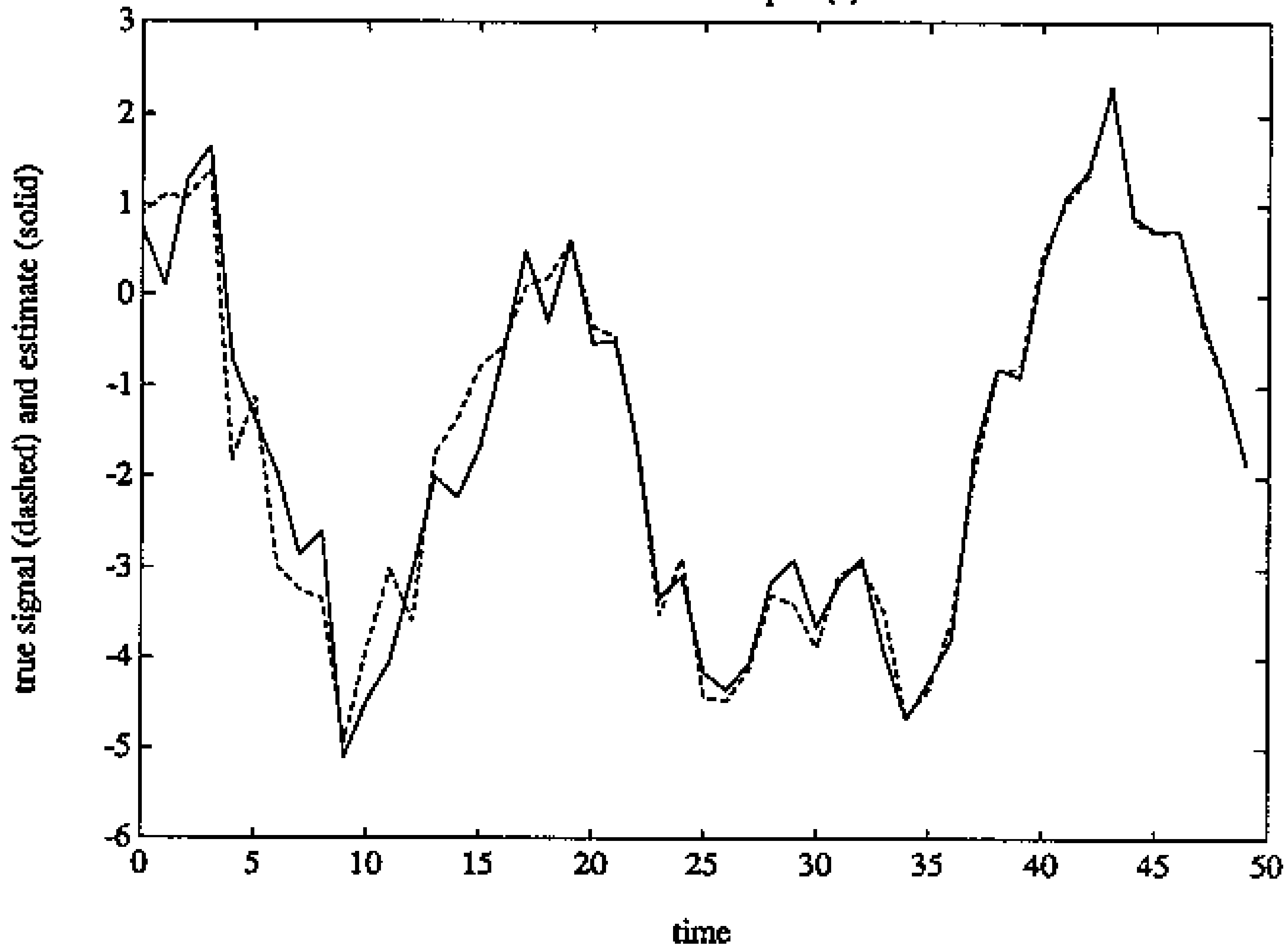
12) If $\sigma_n^2 = 0$, $K[n] = 1$ and $\hat{s}[n|n] = x[n]$

Also, $\hat{s}[n|n-1] = a\,\hat{s}[n-1|n-1]$

$$= a\,x[n-1] = a\,s[n-1]$$

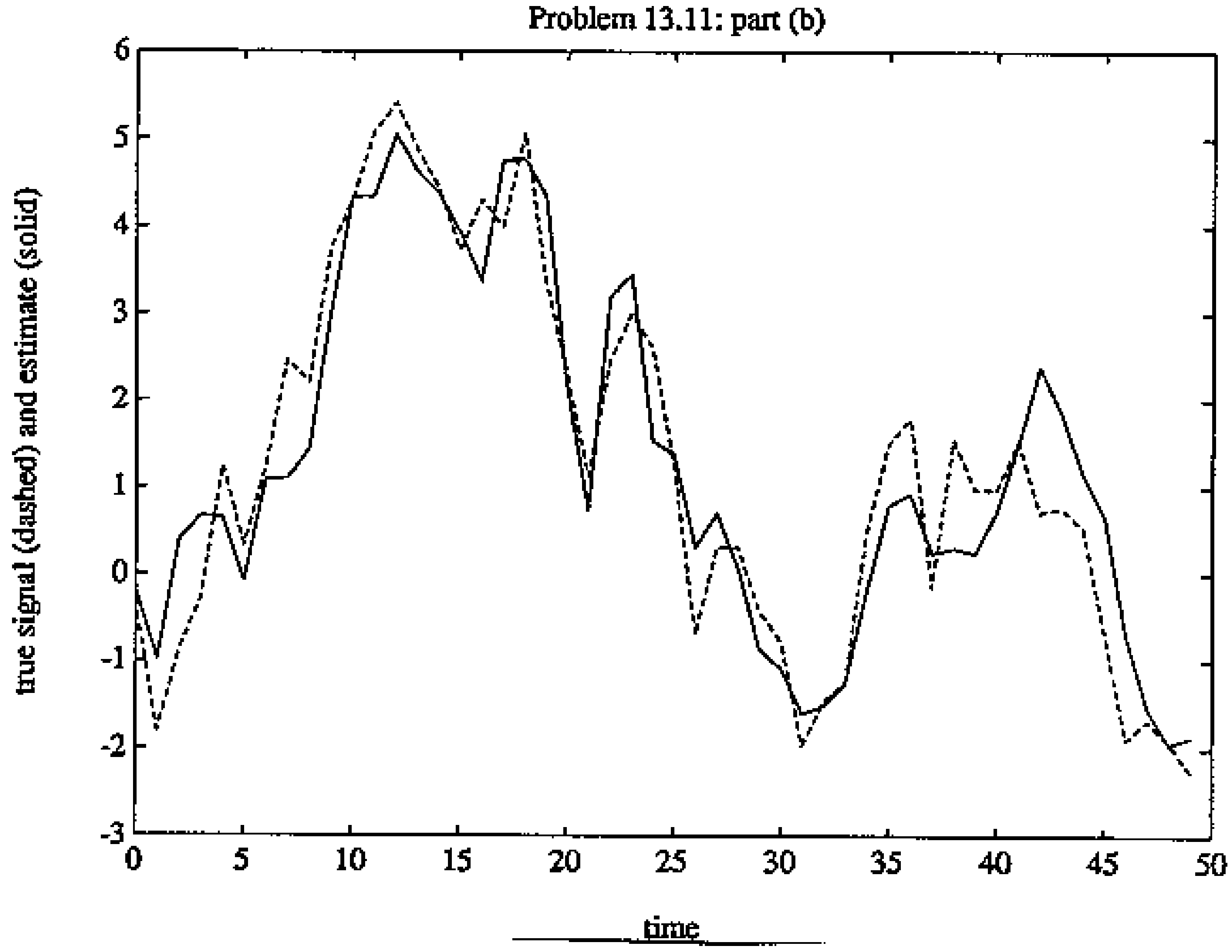
Thus, $\tilde{x}[n] = x[n] - \hat{s}[n|n-1]$

Problem 13.11: part (a)

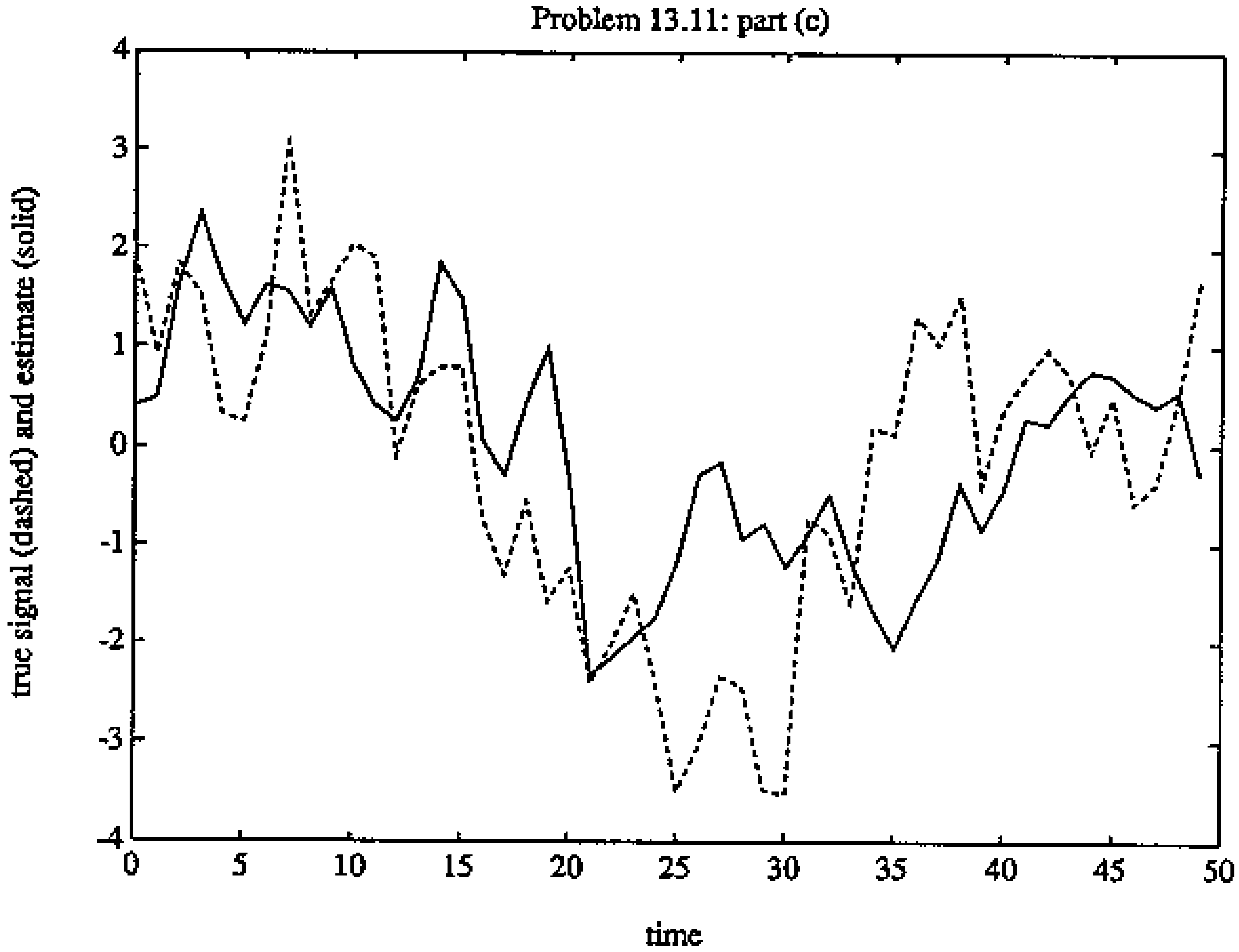


$s[n]$ = DASHED
$\hat{s}[n|n]$ = SOLID

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 MMSE 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: $\hat{s}'[n|n-1] = a\,\hat{s}'[n-1|n-1]$

$$\hat{s}[n|n-1] - E(s[n]) = a\left(\hat{s}[n-1|n-1] - E(s[n-1])\right)$$

or $\hat{s}[n|n-1] = a\,\hat{s}[n-1|n-1]$

since $E(s[n]) = a\,E(s[n-1])$

Correction: $\hat{s}'[n|n] = \hat{s}'[n|n-1]$

$$+ K[n]\left(x'[n] - \hat{s}'[n|n-1]\right)$$

$$\hat{s}[n|n] - E(s[n]) = \hat{s}[n|n-1] - E(s[n])$$

$$+ K[n]\left(x[n] - E(x[n]) - \hat{s}[n|n-1] + E(s[n])\right)$$

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 $\mu_s \neq 0$.

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 $\{x[0], x[1], \ldots, x[n-1]\}$ 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 $\sigma_{n+1}^2 \to \infty$, the future measurements will be useless so that the corrected estimates will be predictions or will be based on only $\{x[0], \ldots, x[n]\}$.

$$\Rightarrow \hat{s}[n+1|n+1] \to \hat{s}[n+1|n] = a\,\hat{s}[n|n]$$
$$\hat{s}[n+2|n+1] = a\,\hat{s}[n+1|n+1] = a^2\,\hat{s}[n|n]$$

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 $x = M[\infty]$

$$0.64x^2 + 2x = 0.64x + 1$$

$$x^2 + 2.125x - 1.5625 = 0$$

$$\Rightarrow \quad x = 0.5781, \ -2.703$$

$$M[\infty] = 0.5781$$

$$M_p[\infty] = a^2 M[\infty] + \sigma_u^2$$

$$= 0.64 M[\infty] + 1 = 1.37$$

$$K[\infty] = \frac{M_p[\infty]}{1 + M_p[\infty]} = 0.5781$$

$$H_\infty(z) = \frac{0.5781}{1 - 0.3375 z^{-1}}$$

$$\hat{s}[n|n] = 0.3375\,\hat{s}[n-1|n-1] + 0.5781\,x[n]$$

17) See plot on next page.

18) Here $h[n] = r^n$. Using (13.51) - (13.54)

$$x[n] = h[n]A + w[n]$$

$$\underline{Q} = \underline{0}$$

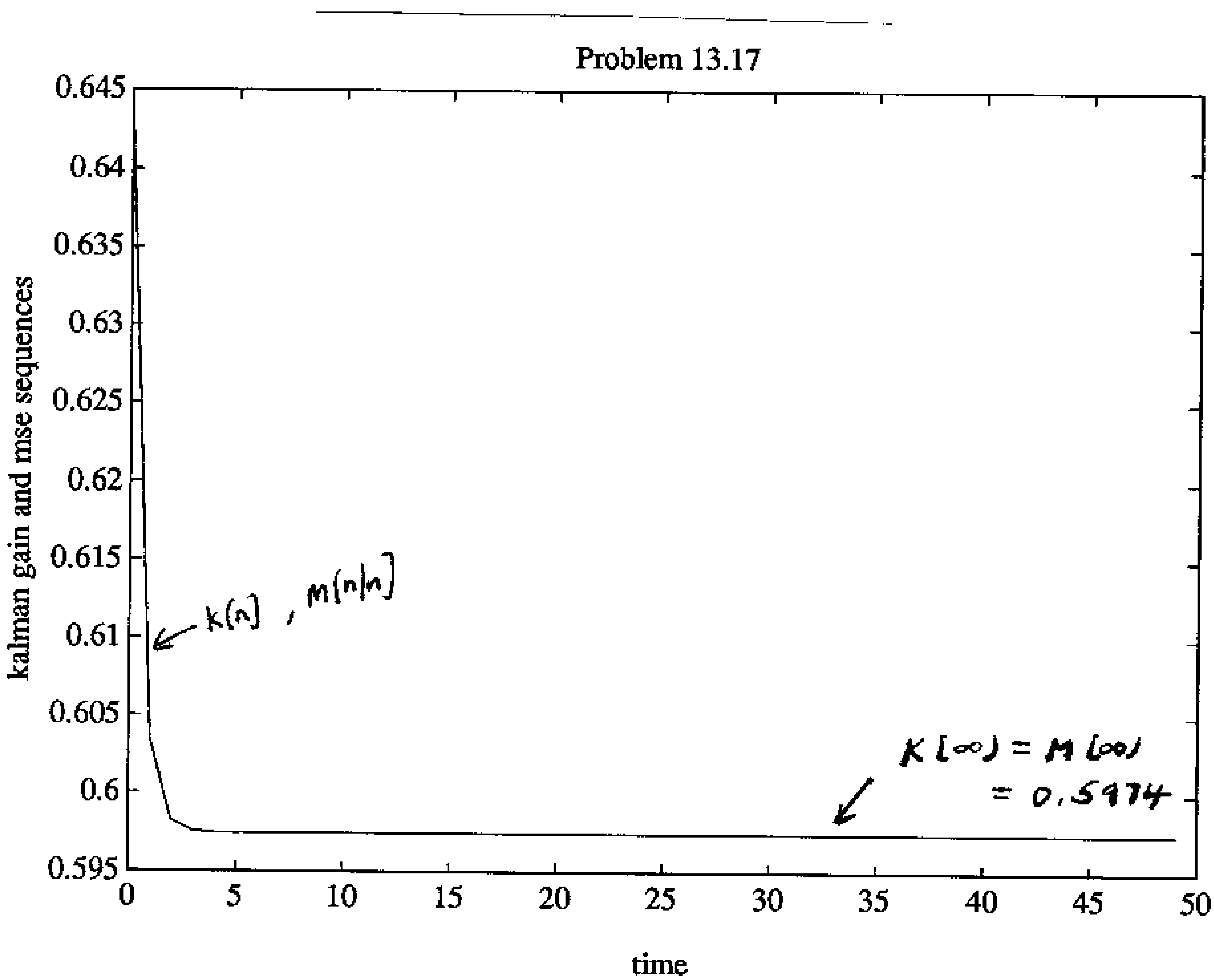
$$\underline{A} = \underline{I}$$

Can omit prediction stage.

$$K[n] = \frac{M[n|n-1]\,r^n}{\sigma^2 + r^{2n} M[n|n-1]}$$

$$M[n|n] = (1 - K[n]r^n)\,M[n|n-1]$$

$$\hat{A}[n|n] = \hat{A}[n|n-1] + K[n](x[n] - r^n \hat{A}[n|n-1])$$



where $\hat{A}[-1|-1] = \mu_A$, $M[-1|-1] = \sigma_A^2$.

19) If $\underline{C}[n] = \underline{0}$, from (13.60)

$$\underline{K}[n] = \underline{M}[n|n-1]\,\underline{H}^T[n]\left(\underline{H}[n]\,\underline{M}[n|n-1]\,\underline{H}^T[n]\right)^{-1}$$

$$= \underline{M}[n|n-1]\,\underline{H}^T[n]\,\underline{H}^{T^{-1}}[n]\,\underline{M}^{-1}[n|n-1]\,\underline{H}^{-1}[n]$$

$$= \underline{H}^{-1}[n]$$

$$\hat{\underline{s}}[n|n] = \hat{\underline{s}}[n|n-1] + \underline{H}^{-1}[n]\left(\underline{x}[n] - H[n]\,\hat{\underline{s}}[n|n-1]\right)$$

$$= \underline{H}^{-1}[n]\,\underline{x}[n]$$

$\Rightarrow$ discard all previous data since

$\underline{x}[n] = \underline{H}[n]\,\underline{s}[n]$ and thus $\underline{s}[n] = \underline{H}^{-1}[n]\,\underline{x}[n]$

$\Rightarrow$ no error in $\hat{\underline{s}}[n|n]$.

If $\underline{C}[n] \to \infty$, $\underline{K}[n] \to \underline{0} \Rightarrow \hat{\underline{s}}[n|n] \to \hat{\underline{s}}[n|n-1]$
or we ignore the data sample $\underline{x}[n]$.

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)

$$\hat{f}_0[n|n-1] = a\hat{f}_0[n-1|n-1]$$

$$M[n|n-1] = a^2 M[n-1|n-1] + \sigma_u^2$$

$$K[n] = \frac{M[n|n-1]H[n]}{\sigma^2 + H^2[n]M[n|n-1]}$$

where $H[n] = \left.\frac{\partial h}{\partial f_0[n]}\right|_{f_0[n]=\hat{f}_0[n|n-1]}$

But $h(f_0[n]) = \cos 2\pi f_0[n]$

$$H[n] = -2\pi \sin 2\pi \hat{f}_0[n|n-1]$$

$$\hat{f}_0[n|n] = \hat{f}_0[n|n-1] + K[n]\left(x[n] - \cos 2\pi \hat{f}_0[n|n-1]\right)$$

$$M[n|n] = (1 - K[n]H[n])M[n|n-1]$$

22) $v_x[n] = v_x[n-1] + u_x[n]$

$$\Rightarrow v_x[n] = \sum_{k=0}^{n} u_x[k] + v_x[-1]$$

$$\operatorname{var}(v_x[n]) = \sum_{k=0}^{n} \operatorname{var}(u_x[k]) + \operatorname{var}(v_x[-1])$$

since $u_x[n]$ is WGN and is independent of $v_x[-1]$

$$\mathrm{var}(v_x[n]) = (n+1)\,\mathrm{var}(u_x[k]) + \sigma_v^2$$
$$= (n+1)\sigma_u^2 + \sigma_v^2$$

a better model would be

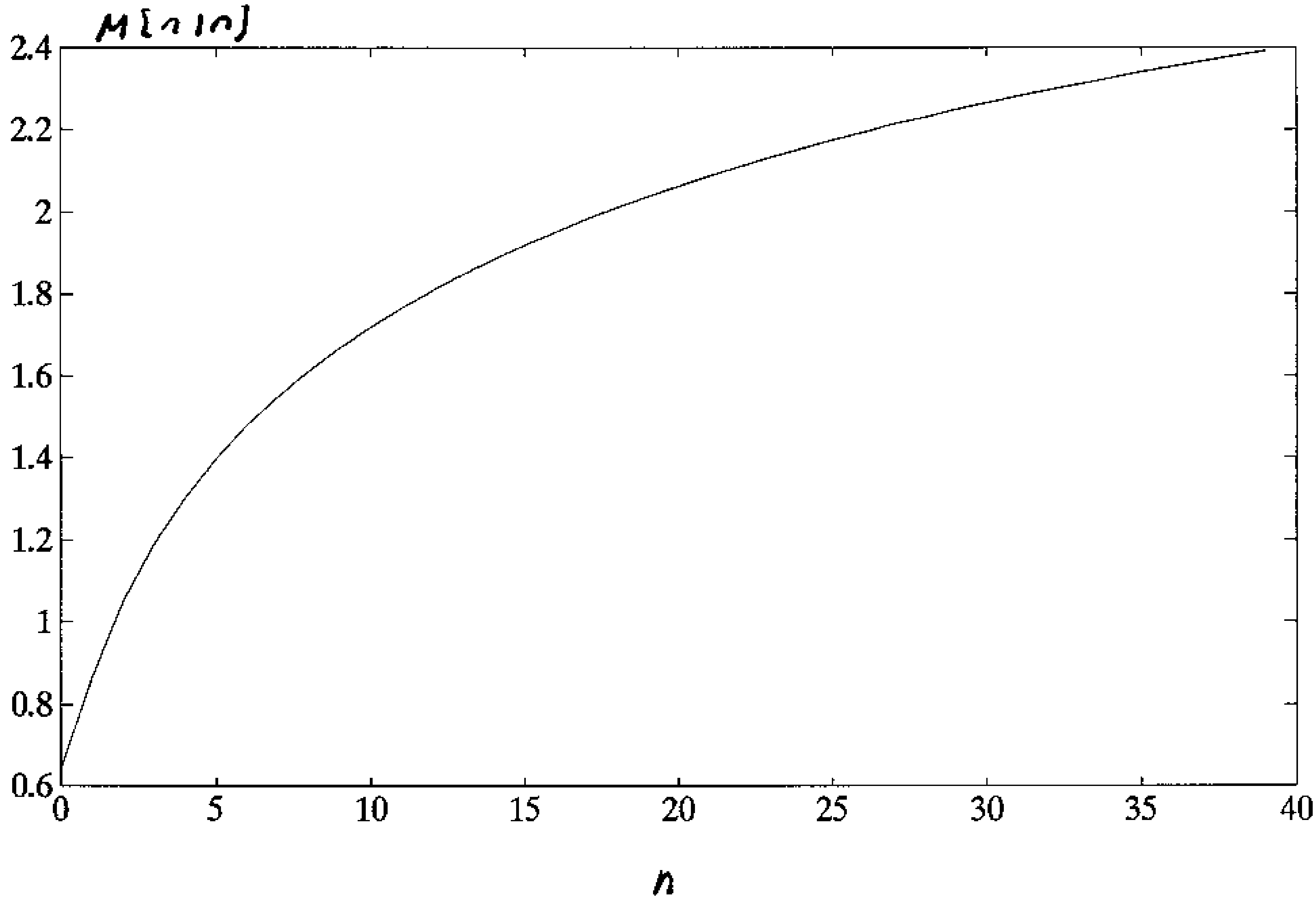
$$v_x[n] = a\, v_x[n-1] + u_x[n]$$

for $0 < a < 1$ and a should be near one. Then in steady-state the variance would be a constant.

23) $$M[n|n] = (1 - K[n])\, M[n|n-1]$$
$$= \left(1 - \frac{M[n|n-1]}{\sigma_n^2 + M[n|n-1]}\right) M[n|n-1]$$
$$= \frac{\sigma_n^2 M[n|n-1]}{\sigma_n^2 + M[n|n-1]}$$
$$= \frac{\sigma_n^2 (a^2 M[n-1|n-1] + \sigma_u^2)}{\sigma_n^2 + a^2 M[n-1|n-1] + \sigma_u^2}$$
$$= \frac{(n+1)(0.81\, M[n-1|n-1] + 1)}{n+1 + 0.81\, M[n-1|n-1] + 1}$$

where $M[-1|-1] = 1$

See plot on next page. Since the observations are progressively noisier, the MMSE increases.


M[1|n]
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0
5
10
15
20
25
30
35
40
n

Chapter 15

1) $\mathrm{Cov}(\tilde{x}_1, \tilde{x}_2) = E[(\tilde{x}_1 - E(\tilde{x}_1))^* (\tilde{x}_2 - E(\tilde{x}_2))]$

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,

$$\mathrm{Cov}(\tilde{x}_1, \tilde{x}_2) = E_{u_1, v_1}(\tilde{x}_1 - E(\tilde{x}_1)) E_{u_2, v_2}(\tilde{x}_2 - E(\tilde{x}_2))$$
$$= 0$$

2) Consider $\tilde{y} = \underline{a}^H \tilde{\underline{x}}$

$E(|\tilde{y}|^2) = E(\underline{a}^H \tilde{\underline{x}} \tilde{\underline{x}}^H \underline{a}) = \underline{a}^H \underline{C}_{\tilde{x}} \underline{a} \geq 0$ for all $\underline{a}$. It will be positive definite if and only if $E(|\tilde{y}|^2) > 0$ or $\tilde{y} \neq \underline{0}$ for any $\underline{a}$. Thus, if any random variable (element of $\tilde{\underline{x}}$) can be expressed as a linear combination of the others, $\underline{C}_{\tilde{x}}$ will only be positive semidefinite.

3) $$\underline{C}_x = \frac{1}{2} \begin{bmatrix} \underline{A} & -\underline{B} \\ \underline{B} & \underline{A} \end{bmatrix} \quad \text{where } \underline{A} = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$$
$$\underline{B} = \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}$$

Note that $\underline{A}^T = \underline{A}$, $\underline{B}^T = -\underline{B}$ as required.

$$\Rightarrow \underline{C}_{\tilde{x}} = \underline{A} + j\underline{B} = \begin{bmatrix} 4 & 2+2j \\ 2-2j & 4 \end{bmatrix}$$

The complex Gaussian vector $\tilde{\underline{x}} = \begin{bmatrix} u_1 + jv_1 \\ u_2 + jv_2 \end{bmatrix}$ will have the covariance matrix $\underline{C}_{\tilde{x}}$.

4) a) $C_x^{-1} = \frac{1}{2}\begin{bmatrix}2 & -1 & 0 & 1\\ -1 & 2 & -1 & 0\\ 0 & -1 & 2 & -1\\ 1 & 0 & -1 & 2\end{bmatrix}$

$$C_{\tilde{x}}^{-1} = \frac{1}{4}\begin{bmatrix}2 & -1-j\\ -1+j & 2\end{bmatrix}$$

Let $\underline{x} = (\underline{u}^T \underline{v}^T)^T$ where $\underline{u} = (u_1 u_2)$, $\underline{v} = (v_1 v_2)^T$

$$2\underline{x}^T C_x^{-1} \underline{x} = (\underline{u}^T \underline{v}^T)\left[\begin{array}{cc|cc}2 & -1 & 0 & 1\\ -1 & 2 & -1 & 0\\ \hline 0 & -1 & 2 & -1\\ 1 & 0 & -1 & 2\end{array}\right]\begin{bmatrix}\underline{u}\\ \underline{v}\end{bmatrix}$$

$$= (\underline{u}^T \underline{v}^T)\begin{bmatrix}C & D^T\\ D & C\end{bmatrix}\begin{bmatrix}\underline{u}\\ \underline{v}\end{bmatrix}$$

$$= \underline{u}^T C \underline{u} + \underline{v}^T C \underline{v} + \underline{u}^T D^T \underline{v} + \underline{v}^T D \underline{u}$$

$$= \underline{u}^T C \underline{u} + \underline{v}^T C \underline{v} + \underline{v}^T D \underline{u} + \underline{v}^T D \underline{u}$$

$$= \underline{u}^T C \underline{u} + \underline{v}^T D \underline{v} + 2\underline{v}^T D \underline{u}$$

$$\tilde{\underline{x}}^H C_{\tilde{x}}^{-1} \tilde{\underline{x}} = (\underline{u} - j\underline{v})^T \frac{1}{4}(C + jD)(\underline{u} + j\underline{v})$$

$$= \frac{1}{4}(\underline{u} - j\underline{v})^T (C\underline{u} + jC\underline{v} + jD\underline{u} - D\underline{v})$$

$$= \frac{1}{4}(\underline{u}^T C \underline{u} + j\underline{u}^T C \underline{v} + j\underline{u}^T D \underline{u} - \underline{u}^T D \underline{v}$$
$$- j\underline{v}^T C \underline{u} + \underline{v}^T C \underline{v} + \underline{v}^T D \underline{u} + j\underline{v}^T D \underline{v})$$

But $\underline{u}^T D \underline{u} = (\underline{u}^T D \underline{u})^T = \underline{u}^T D^T \underline{u} = -\underline{u}^T D \underline{u} = 0$

since $D^T = -D$ and similarly for $\underline{v}^T D \underline{v} = 0$

Also, $\underline{v}^T C \underline{u} = \underline{u}^T C \underline{v}$ since $C^T = C$

$$\tilde{\underline{x}}^H \underline{C}_{\tilde{x}}^{-1} \tilde{\underline{x}} = 1/4 (\underline{u}^T \underline{C} \underline{u} - \underline{u}^T \underline{D} \underline{v} + \underline{v}^T \underline{C} \underline{v} + \underline{v}^T \underline{D} \underline{u})$$

$$= 1/4 (\underline{u}^T \underline{C} \underline{u} - \underline{v}^T \underline{D}^T \underline{u} + \underline{v}^T \underline{C} \underline{v} + \underline{v}^T \underline{D} \underline{u})$$

$$= 1/4 (\underbrace{\underline{u}^T \underline{C} \underline{u} + \underline{v}^T \underline{C} \underline{v} + 2 \underline{v}^T \underline{D} \underline{u}}_{2 \underline{x}^T \underline{C}_x^{-1} \underline{x}})$$

$$= \frac{1}{2} \underline{x}^T \underline{C}_x^{-1} \underline{x}$$

b) $\det(C_{\underline{x}}) = 4$

$$\det(\underline{C}_{\tilde{x}}) = \det \begin{pmatrix} 4 & 2+2j \\ 2-2j & 4 \end{pmatrix}$$

$$= 16 - |2+2j|^2 = 16 - 8 = 8$$

$$\frac{\det^2(C_{\tilde{x}})}{16} = \frac{64}{16} = 4$$

5) $m_1 = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ $\quad m_2 = \begin{bmatrix} c & -d \\ d & c \end{bmatrix}$

$m_1 \rightarrow a + jb \qquad m_2 \rightarrow c + jd$

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 2x2 blocks or let

$$\underline{A} = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} \rightarrow \begin{bmatrix} a_1 + jb_1 \\ a_2 + jb_2 \\ a_3 + jb_3 \end{bmatrix}$$

But $\underline{A}^T\underline{A} = m_1^T m_1 + m_2^T m_2 + m_3^T m_3$

(Note that if $m \rightarrow a+jb$, $m^T \rightarrow a - jb$)

$$\Rightarrow \quad \underline{A}^T\underline{A} \rightarrow (a_1 - jb_1)(a_1 + jb_1) + (a_2 - jb_2)(a_2 + jb_2) + (a_3 - jb_3)(a_3 + jb_3)$$
$$= a_1^2 + b_1^2 + a_2^2 + b_2^2 + a_3^2 + b_3^2 + j\,0$$

Transforming back we have

$$\underline{A}^T\underline{A} = \begin{bmatrix} a_1^2 + b_1^2 + a_2^2 + b_2^2 + a_3^2 + b_3^2 & & 0 \\ & \ddots & \\ 0 & & \end{bmatrix}$$

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 $E(\tilde{x}^3) = 0$, $E(\tilde{x}^*\tilde{x}^2) = 0$.

But $$\frac{\partial^3 \Phi_{\tilde{x}}(\tilde{\omega})}{\partial \tilde{\omega}^{*3}} = (j/2)^3 E(\tilde{x}^3)$$

(see development of (15B.3) in App 15B)

$$\phi_{\tilde{x}}(\tilde{\omega}) = e^{-1/4\,\sigma^2 \tilde{\omega}\tilde{\omega}^*}$$

$$\frac{\partial \phi}{\partial \tilde{\omega}^*} = e^{-1/4\,\sigma^2 |\tilde{\omega}|^2}\left(-1/4\,\sigma^2 \tilde{\omega}\right)$$

$$\frac{\partial^2 \phi}{\partial \tilde{\omega}^{*2}} = e^{-1/4\,\sigma^2 |\tilde{\omega}|^2}\left(-1/4\,\sigma^2 \tilde{\omega}\right)^2$$

$$\frac{\partial^3 \phi}{\partial \tilde{\omega}^{*3}} = e^{-1/4\,\sigma^2 |\tilde{\omega}|^2}\left(-1/4\,\sigma^2 \tilde{\omega}\right)^3$$

$$\Rightarrow \quad \left.\frac{\partial^3 \phi}{\partial \tilde{\omega}^{*3}}\right|_{\tilde{\omega}=0} = 0 \quad \Rightarrow \quad E(\tilde{x}^3) = 0$$

Similarly,

$$E(\tilde{x}^* \tilde{x}^2) = (j/2)^3 \left.\frac{\partial^3 \phi_{\tilde{x}}(\tilde{\omega})}{\partial \tilde{\omega}\,\partial \tilde{\omega}^{*2}}\right|_{\tilde{\omega}=0}$$

$$\frac{\partial^2 \phi}{\partial \tilde{\omega}^{*2}} = e^{-1/4\,\sigma^2 \tilde{\omega}\tilde{\omega}^*}\left(-1/4\,\sigma^2 \tilde{\omega}\right)^2$$

$$\frac{\partial^3 \phi}{\partial \tilde{\omega}\,\partial \tilde{\omega}^{*2}} = e^{-1/4\,\sigma^2 |\tilde{\omega}|^2}\left(\tfrac{1}{16}\,\sigma^4\, 2\tilde{\omega}\right) + \left(-1/4\,\sigma^2 \tilde{\omega}\right)^2 e^{-1/4\,\sigma^2 |\tilde{\omega}|^2}\left(-1/4\,\sigma^2 \tilde{\omega}^*\right)$$

which when evaluated at $\tilde{\omega} = 0$ is zero.

8) $E(\tilde{x}^2) = 0$

$$\Rightarrow E\big((u+jv)(u+jv)\big) = 0$$

$$E(u^2) - E(v^2) + j\,E(uv) + j\,E(uv) = 0$$

$$\Rightarrow E(u^2) = E(v^2), \quad E(uv) = 0$$

or $\mathrm{var}(u) = \mathrm{var}(v)$, $\mathrm{cov}(u,v) = 0$

This is the usual complex Gaussian random variable assumption.

9) $E[(\tilde{\underline{x}}-\tilde{\underline{\mu}})(\tilde{\underline{x}}-\tilde{\underline{\mu}})^T] =$

$$E[((\underline{u}-\underline{\mu}_u)+j(\underline{v}-\underline{\mu}_v))(\quad)^T]$$

$$= E[(\underline{u}-\underline{\mu}_u)(\underline{u}-\underline{\mu}_u)^T] - E[(\underline{v}-\underline{\mu}_v)(\underline{v}-\underline{\mu}_v)^T]$$
$$+ j(E[(\underline{u}-\underline{\mu}_u)(\underline{v}-\underline{\mu}_v)^T] + E[(\underline{v}-\underline{\mu}_v)(\underline{u}-\underline{\mu}_u)^T])$$
$$= \underline{0}$$

$$\Rightarrow \quad \underline{C}_{uu} = \underline{C}_{vv}, \quad \underline{C}_{uv} = -\underline{C}_{vu}$$
$$= \underline{A}/2 \qquad\qquad = -\underline{B}/2$$

or $\underline{C}_x = E\left[\left(\begin{bmatrix}\underline{u}\\ \underline{v}\end{bmatrix} - E\begin{bmatrix}\underline{u}\\ \underline{v}\end{bmatrix}\right)\left(\begin{bmatrix}\underline{u}\\ \underline{v}\end{bmatrix} - E\begin{bmatrix}\underline{u}\\ \underline{v}\end{bmatrix}\right)^T\right]$

$$= \begin{bmatrix} \underline{A}/2 & -\underline{B}/2 \\ \underline{B}/2 & \underline{A}/2 \end{bmatrix}$$

10) $E(\hat{\sigma}^2) = E(\tilde{\underline{x}}^H \underline{A} \tilde{\underline{x}}) = \operatorname{tr}(\underline{A}\,\sigma^2 \underline{B})$ (see (15.29))

$$= \sigma^2 \operatorname{tr}(\underline{A}\,\underline{B})$$
$$\Rightarrow \operatorname{tr}(\underline{A}\,\underline{B}) = 1$$

$\operatorname{var}(\hat{\sigma}^2) = \operatorname{tr}(\underline{A}\,\sigma^2\underline{B}\,\underline{A}\,\sigma^2\underline{B})$ (see (15.30))

$$= \sigma^4 \operatorname{tr}((\underline{A}\,\underline{B})^2)$$

But $\operatorname{tr}(\underline{A}\,\underline{B}) = \sum_{i=1}^{N} \lambda_i = 1$

$$\operatorname{tr}((\underline{A}\,\underline{B})^2) = \sum_{i=1}^{N} \lambda_i^2$$

Using the constraint we have to minimize

$$tr((\underline{A}\underline{B})^2) = \sum_{i=2}^{N} \lambda_i^2 + \left(1 - \sum_{i=2}^{N} \lambda_i\right)^2$$

$$\frac{\partial\, tr((\underline{A}\underline{B})^2}{\partial \lambda_k} = 2\lambda_k + 2\left(1 - \sum_2^N \lambda_i\right)(-1) = 0$$

$$k = 2, 3, \ldots, N$$

$$\Rightarrow \quad \lambda_k = 1 - \sum_2^N \lambda_i$$

Since $\sum_2^N \lambda_i$ is a constant, λ_k must be the same for all k. Thus,

$$\lambda_k = 1/N \text{ for } tr(\underline{A}\underline{B}) = 1.$$

Since

$$\underline{V}^T \underline{A}\, \underline{B}\, \underline{V} = \underline{\Lambda} \qquad \underline{V} = \text{modal matrix}$$

$$= 1/N\, \underline{I} \qquad \underline{\Lambda} = \text{eigenvalue matrix}$$

$$\Rightarrow \quad \underline{A}\underline{B} = 1/N\, \underline{V}^{T^{-1}} \underline{V}^{-1} = 1/N\, \underline{I}$$

$$\Rightarrow \quad \underline{A} = 1/N\, \underline{B}^{-1} \text{ or } \hat{\sigma}^2 = 1/N\, \tilde{\underline{x}}^H \underline{B}^{-1} \tilde{\underline{x}}$$

If $\underline{B} = \underline{I}$, $\hat{\sigma}^2 = 1/N\, \tilde{\underline{x}}^H \tilde{\underline{x}}$.

11) $$\sum_{n=0}^{N-1} |\tilde{x}[n]|^2 = \tilde{\underline{x}}^H \tilde{\underline{x}}$$

$$E(\tilde{\underline{x}}^H \tilde{\underline{x}}) = N\sigma^2$$

$$\text{var}(\tilde{\underline{x}}^H \tilde{\underline{x}}) = tr(\underline{C}_{\tilde{x}} \underline{C}_{\tilde{x}}) \qquad (\text{see } (15.30))$$

$$= tr(\sigma^4 \underline{I})$$

$$= N\sigma^4$$

$$\Rightarrow \hat{\sigma}^2 = 1/N \, \tilde{\underline{x}}^H \tilde{\underline{x}}$$

where $\mathrm{var}(\hat{\sigma}^2) = \sigma^4/N$

For real WGN, $\hat{\sigma}^2 = 1/N \, \underline{x}^T \underline{x}$ and $\mathrm{var}(\hat{\sigma}^2) = 2\sigma^4/N$. Now, since $\tilde{x}(n) \sim CN(0, \sigma^2)$, $\tilde{x}[n] = u[n] + j v[n]$, where $u[n] \sim N(0, \sigma^2/2)$, $v[n] \sim N(0, \sigma^2/2)$. Hence, for the complex case $\hat{\sigma}^2 = 1/N \sum_{n=0}^{N-1} (u^2[n] + v^2[n])$ and we are averaging $2N$ independent samples as opposed to only N for the real case. Thus, $\mathrm{var}(\hat{\sigma}^2)\big|_{\text{COMPLEX}} = 1/2 \, \mathrm{var}(\hat{\sigma}^2)\big|_{\text{REAL}}$.

(2) $$v[n] = \sum_{k=-\infty}^{\infty} h[k] u[n-k]$$

$$E(v[n]) = \sum_k h[k] E(u[n-k]) = 0$$

$$E(v[n] v[n+m]) = E\Big[\sum_k h[k] u[n-k] \cdot \sum_l h[l] u[n+m-l] \Big]$$

$$= \sum_k \sum_l h[k] h[l] \underbrace{E(u[n-k] u[n+m-l])}_{r_{uu}[m+k-l]}$$

doesn't depend on $n \Rightarrow$ WSS

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

$$P_{vv}(f) = |H(f)|^2 P_{uu}(f)$$
$$= 1 \cdot P_{uu}(f)$$
$$\Rightarrow \quad r_{uu}[k] = r_{vv}[k]$$

also, $P_{uv}(f) = H(f) P_{uu}(f)$

$$= \begin{cases} -j P_{uu}(f) & f \geq 0 \\ j P_{uu}(f) & f < 0 \end{cases}$$

But $P_{vu}(f) = P_{uv}^*(f)$,

$$\Rightarrow \quad P_{vu}(f) = \begin{cases} j P_{uu}(f) & f \geq 0 \\ -j P_{uu}(f) & f < 0 \end{cases}$$

$$= -P_{uv}(f)$$

and thus, $r_{uv}[k] = -r_{vu}[k]$. The PSD of $\tilde{x}[n]$ is from (15.3

$$P_{\tilde{x}\tilde{x}}(f) = 2(P_{uu}(f) + j P_{uv}(f))$$
$$= \begin{cases} 2(P_{uu}(f) + j(-j P_{uu}(f))) & f \geq 0 \\ 2(P_{uu}(f) + j(j P_{uu}(f))) & f < 0 \end{cases}$$

$$= \begin{cases} 4 P_{uu}(f) & f \geq 0 \\ 0 & f < 0 \end{cases}$$

B) $$\frac{\partial \theta^*}{\partial \theta} = \frac{1}{2}\left(\frac{\partial}{\partial \alpha} - j\frac{\partial}{\partial \beta}\right)(\alpha - j\beta)$$

$$= \frac{1}{2}\left(\frac{\partial \alpha}{\partial \alpha} - j\frac{\partial \beta}{\partial \alpha} - j\frac{\partial \alpha}{\partial \beta} - \frac{\partial \beta}{\partial \beta}\right)$$
$$= \frac{1}{2}\left(1 - 0 - 0 - 1\right) = 0$$

$$\frac{\partial}{\partial \theta} = \frac{\partial}{\partial \alpha} + j\frac{\partial}{\partial \beta}$$

$$\frac{\partial \theta}{\partial \theta} = \left(\frac{\partial}{\partial \alpha} + j\frac{\partial}{\partial \beta}\right)(\alpha + j\beta)$$
$$= 1 - 1 = 0$$

$$\frac{\partial \theta^*}{\partial \theta} = \left(\frac{\partial}{\partial \alpha} + j\frac{\partial}{\partial \beta}\right)(\alpha - j\beta)$$
$$= 1 + 1 = 2$$

14) $$\frac{\partial \underline{\theta}^H \underline{b}}{\partial \theta_k} = \frac{\partial}{\partial \theta_k} \sum_{l} b_l \theta_l^*$$
$$= \sum_{l} b_l \frac{\partial \theta_l^*}{\partial \theta_k} = 0 \quad \text{all } k$$

$$\Rightarrow \frac{\partial \underline{\theta}^H \underline{b}}{\partial \underline{\theta}} = \underline{0}$$

15) Same proof as in deriving (15.46)
$$\frac{\partial \underline{\theta}^H \underline{A}\,\underline{\theta}}{\partial \underline{\theta}} = \underline{A}^T \underline{\theta}^*$$
Since $\underline{A}^H \neq \underline{A}$, $$\frac{\partial \underline{\theta}^H \underline{A}\,\underline{\theta}}{\partial \underline{\theta}} = (\underline{A}^H \underline{\theta})^*$$

16) To find LSE
$$J = \sum_{n=0}^{N-1} |\tilde{x}[n] - \tilde{A} r^n|^2$$

From Example 15.2

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

$$= \frac{\sum_{n=0}^{N-1} \tilde{x}(n)\gamma^{*n}}{\sum_{n=0}^{N-1} |\gamma|^{2n}}$$

MLE will be the same.

$$E(\hat{\tilde{A}}) = \frac{\sum_n E(\tilde{x}[n])\gamma^{*n}}{\sum_n |\gamma|^{2n}}$$

$$= \frac{\tilde{A}\sum_n \gamma^n \gamma^{*n}}{\sum_n |\gamma|^{2n}} = \tilde{A}$$

$$\mathrm{var}(\hat{\tilde{A}}) = \frac{\sum_n \mathrm{var}(\tilde{x}(n))|\gamma|^{2n}}{(\sum_n |\gamma|^{2n})^2}$$

$$= \frac{\sigma^2}{\sum_n |\gamma|^{2n}}$$

If $|\gamma| < 1$, $\mathrm{var}(\hat{\tilde{A}}) \to \dfrac{\sigma^2}{(1-|\gamma|^2)^{-1}}$

$$= \sigma^2(1-|\gamma|^2)$$

If $|\gamma| \geq 1$, $\mathrm{var}(\hat{\tilde{A}}) \to 0$

17) Equivalently, we have

$$\tilde{x}[n] = A + \tilde{w}[n] \quad \text{or}$$

$$u[n] = A + w_R[n]$$

$$v[n] = w_I[n]$$

Now let $\underline{x} = \begin{bmatrix} \underline{u} \\ \underline{v} \end{bmatrix}$ and $\underline{w} = \begin{bmatrix} \underline{w}_R \\ \underline{w}_I \end{bmatrix}$ so that

$$\underline{x} = A \begin{bmatrix} \underline{1} \\ \underline{0} \end{bmatrix} + \underline{w}$$

where $E(\underline{w}) = \underline{0}$, $\underline{C}_w = \sigma^2/2\, \underline{I} = \underline{C}_x$

The real BLUE can now be used.

$$\hat{A} = \frac{[\underline{1}^T \ \underline{0}^T]\, \underline{C}_w^{-1}\, \underline{x}}{[\underline{1}^T \ \underline{0}^T]\, \underline{C}_w^{-1} \begin{bmatrix} \underline{1} \\ \underline{0} \end{bmatrix}}$$

$$= \frac{1}{N} \sum_{n=0}^{N-1} u[n] = \frac{1}{N} \sum_{n=0}^{N-1} \mathrm{Re}(\tilde{x}[n])$$

In Example 15.7 $\tilde{A}$ is complex and if $\underline{C} = \sigma^2 \underline{I}$, which conforms to the assumptions for this problem,

$$\hat{\tilde{A}} = \frac{\underline{1}^T \tilde{\underline{x}}}{\underline{1}^T \underline{1}} = \frac{1}{N} \sum_{n=0}^{N-1} \tilde{x}[n]$$

18) From (15.52)

$$I(\theta) = 2\,\mathrm{Re}\left[\frac{\partial \tilde{\underline{\mu}}^H(\theta)}{\partial \theta} \underline{C}_{\tilde{x}}^{-1} \frac{\partial \tilde{\underline{\mu}}(\theta)}{\partial \theta} \right]$$

$$= \frac{2}{\sigma^2} \mathrm{Re}\left[\sum_{n=0}^{N-1} \left| \frac{\partial [\tilde{\underline{\mu}}(\theta)]_n}{\partial \theta} \right|^2 \right]$$

$$= \frac{2}{\sigma^2} \sum_{n=0}^{N-1} \left| \frac{\partial \tilde{s}[n;\theta]}{\partial \theta} \right|^2$$

$$\operatorname{var}(\hat{\theta}) \geq \frac{\sigma^2/2}{\sum_{n=0}^{N-1} \left(\frac{\partial s_R[n;\theta]}{\partial \theta}\right)^2 + \left(\frac{\partial s_I[n;\theta]}{\partial \theta}\right)^2}$$

For real data (see (3.14))

$$\operatorname{var}(\hat{\theta}) \geq \frac{\sigma^2}{\sum_{n=0}^{N-1} \left(\frac{\partial s[n;\theta]}{\partial \theta}\right)^2}$$

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 $C_{\tilde{x}} = \sigma^2 I$

$$I(\sigma^2) = \operatorname{tr}\left[C_{\tilde{x}}^{-1} \frac{\partial C_{\tilde{x}}}{\partial \sigma^2} C_{\tilde{x}}^{-1} \frac{\partial C_{\tilde{x}}}{\partial \sigma^2} \right]$$

$$= \operatorname{tr}\left[\left(\frac{1}{\sigma^2} I\right)(I)\left(\frac{1}{\sigma^2} I\right)(I) \right]$$

$$= N/\sigma^4$$

$$\operatorname{var}(\hat{\sigma^2}) \geq \sigma^4/N$$

$$p(\tilde{x};\sigma^2) = \frac{1}{\pi^N \det(\sigma^2 I)} e^{-\frac{1}{\sigma^2}\tilde{x}^H \tilde{x}}$$

$$\frac{\partial \ln p}{\partial \sigma^2} = -\frac{\partial \ln \det(\sigma^2 \underline{I})}{\partial \sigma^2} - \frac{\partial}{\partial \sigma^2}\left(\frac{1}{\sigma^2} \underline{\tilde{x}}^H \underline{\tilde{x}}\right)$$

$$= -\frac{\partial \ln \sigma^{2N}}{\partial \sigma^2} - \frac{\partial}{\partial \sigma^2}\left(\frac{1}{\sigma^2} \underline{\tilde{x}}^H \underline{\tilde{x}}\right)$$

$$= -N/\sigma^2 + 1/\sigma^4 \, \underline{\tilde{x}}^H \underline{\tilde{x}}$$

$$= N/\sigma^4 \Big(\underbrace{\frac{1}{N} \sum_{n=0}^{N-1} |\tilde{x}(n)|^2}_{\text{efficient estimator}} - \sigma^2 \Big)$$

20) $\hat{\theta}_i = \underline{a}_i^H \underline{\tilde{x}} \Rightarrow \underline{\hat{\theta}} = \underline{A}\,\underline{\tilde{x}} \qquad \underline{A} = \begin{bmatrix} \underline{a}_1^H \\ \vdots \\ \underline{a}_p^H \end{bmatrix}$

$E(\underline{\hat{\theta}}) = \underline{\theta} \Rightarrow E(\underline{A}(\underline{H}\,\underline{\theta} + \underline{\tilde{w}}))$

$= \underline{A}\,\underline{H}\,E(\underline{\theta}) = \underline{\theta}$

or $\underline{A}\,\underline{H} = \underline{I}$

$\mathrm{var}(\hat{\theta}_i) = E[\,|\theta_i - \hat{\theta}_i|^2]$

$= E[\,|\underline{a}_i^H \underline{\tilde{x}} - \underline{a}_i^H E(\underline{\tilde{x}})|^2]$

$= E[\,\underline{a}_i^H (\underline{x} - E(\underline{x}))(\underline{x} - E(\underline{x}))^H \underline{a}_i\,]$

$= \underline{a}_i^H \underline{C}\, \underline{a}_i$

$\underline{A}\,\underline{H} = \underline{I} \Rightarrow \begin{bmatrix} \underline{a}_1^H \\ \vdots \\ \underline{a}_p^H \end{bmatrix} \underline{H} = \underline{I}$

or $\underline{H}^H [\underline{a}_1 \ldots \underline{a}_p] = \underline{I}$

$\underline{H}^H \underline{a}_i = \underline{e}_i$ where $\underline{e}_i = [0\,0 \ldots 0\,1\,0 \ldots 0]^T$
↑ ith place

Using (15.51) we have

$\underline{a}_{i_{opt}} = \underline{C}^{-1} \underline{H} (\underline{H}^H \underline{C}^{-1} \underline{H})^{-1} \underline{e}_i$

$$\hat{\theta}_i = \underline{a}_{i_{opt}} \tilde{\underline{x}} = \underline{e}_i^H (\underline{H} \underline{C}^{-1} \underline{H})^{-1} \underline{H}^H \underline{C}^{-1} \tilde{\underline{x}}$$

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

21) $\tilde{X} = \tilde{A}_1 \underline{e}_1 + \tilde{A}_2 \underline{e}_2 + \tilde{\underline{w}} = \underline{E} \tilde{\underline{A}} + \tilde{\underline{w}}$

where $\underline{E} = (\underline{e}_1\ \underline{e}_2)$, $\tilde{\underline{A}} = (\tilde{A}_1\ \tilde{A}_2)^T$

The MLE will minimize

$$J(\tilde{\underline{A}}, f) = (\tilde{\underline{x}} - \underline{E}\tilde{\underline{A}})^H (\tilde{\underline{x}} - \underline{E}\tilde{\underline{A}})$$

$$= \tilde{\underline{x}}^H \tilde{\underline{x}} - \tilde{\underline{x}}^H \underline{E} \tilde{\underline{A}} - \tilde{\underline{A}}^H \underline{E}^H \tilde{\underline{x}} + \tilde{\underline{A}}^H \underline{E}^H \underline{E} \tilde{\underline{A}}$$

Taking the complex gradient we have

$$\frac{\partial J}{\partial \tilde{\underline{A}}} = \underline{0} - \frac{\partial}{\partial \tilde{A}} \tilde{\underline{x}}^H \underline{E} \tilde{\underline{A}} - \underline{0} + \frac{\partial}{\partial \tilde{A}} \tilde{\underline{A}}^H \underline{E}^H \underline{E} \tilde{\underline{A}}$$

Using (15.44) and (15.46)

$$\frac{\partial J}{\partial \tilde{\underline{A}}} = -(\underline{E}^H \tilde{\underline{x}})^* + (\underline{E}^H \underline{E} \tilde{\underline{A}})^* = \underline{0}$$

$$\Rightarrow \hat{\tilde{\underline{A}}} = (\underline{E}^H \underline{E})^{-1} \underline{E}^H \tilde{\underline{x}}$$

$$J(\hat{\tilde{\underline{A}}}, f) = (\tilde{\underline{x}} - \underline{E}(\underline{E}^H \underline{E})^{-1} \underline{E}^H \tilde{\underline{x}})^H \cdot (\tilde{\underline{x}} - \underline{E}(\underline{E}^H \underline{E})^{-1} \underline{E}^H \tilde{\underline{x}})$$

$$= \tilde{\underline{x}}^H (\underline{I} - \underline{E}(\underline{E}^H \underline{E})^{-1} \underline{E}^H) \tilde{\underline{x}}$$

since $\underline{I} - \underline{E}(\underline{E}^H \underline{E})^{-1} \underline{E}^H$ is idempotent

To minimize over $\underline{f}$ we need to maximize

$J(\underline{f}) = \tilde{\underline{x}}^H \underline{E} (\underline{E}^H \underline{E})^{-1} \underline{E}^H \tilde{\underline{x}}$

This will require a 2-D search. An approximate MLE for $|f_1 - f_2| >> 1/N$ is found as follows:

$$\underline{E}^H \underline{E} = \begin{bmatrix} \underline{e}_1^H \\ \underline{e}_2^H \end{bmatrix} [\underline{e}_1 \ \underline{e}_2] = \begin{bmatrix} N & \underline{e}_1^H \underline{e}_2 \\ \underline{e}_2^H \underline{e}_1 & N \end{bmatrix}$$

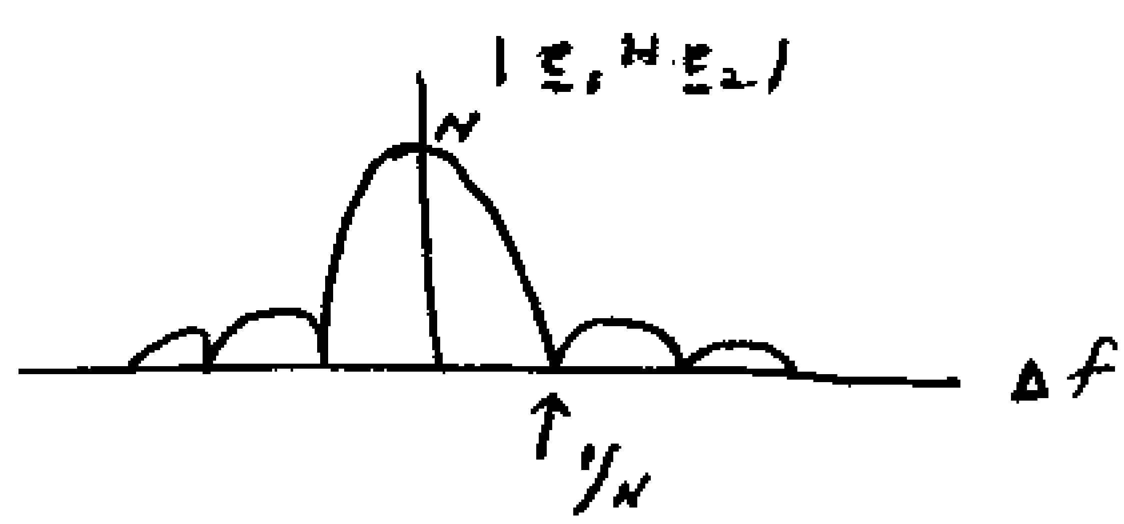
$$\underline{e}_1^H \underline{e}_2 = \sum_{n=0}^{N-1} e^{-j2\pi f_1 n} e^{j2\pi f_2 n}$$

$$= \sum_{n=0}^{N-1} e^{j2\pi \Delta f n} \qquad \Delta f = f_2 - f_1$$

$$= \frac{1 - e^{j2\pi \Delta f N}}{1 - e^{j2\pi \Delta f}}$$

$$= \frac{e^{j\pi \Delta f N}}{e^{j\pi \Delta f}} \frac{e^{-j\pi \Delta f N} - e^{j\pi \Delta f N}}{e^{-j\pi \Delta f} - e^{j\pi \Delta f}}$$

$$= e^{j\pi \Delta f (N-1)} \frac{\sin \pi \Delta f N}{\sin \pi \Delta f}$$



For $|\Delta f| >> 1/N$, $\underline{e}_1^H \underline{e}_2 << N$ and thus

$$\underline{E}^H \underline{E} \simeq N \underline{I}$$

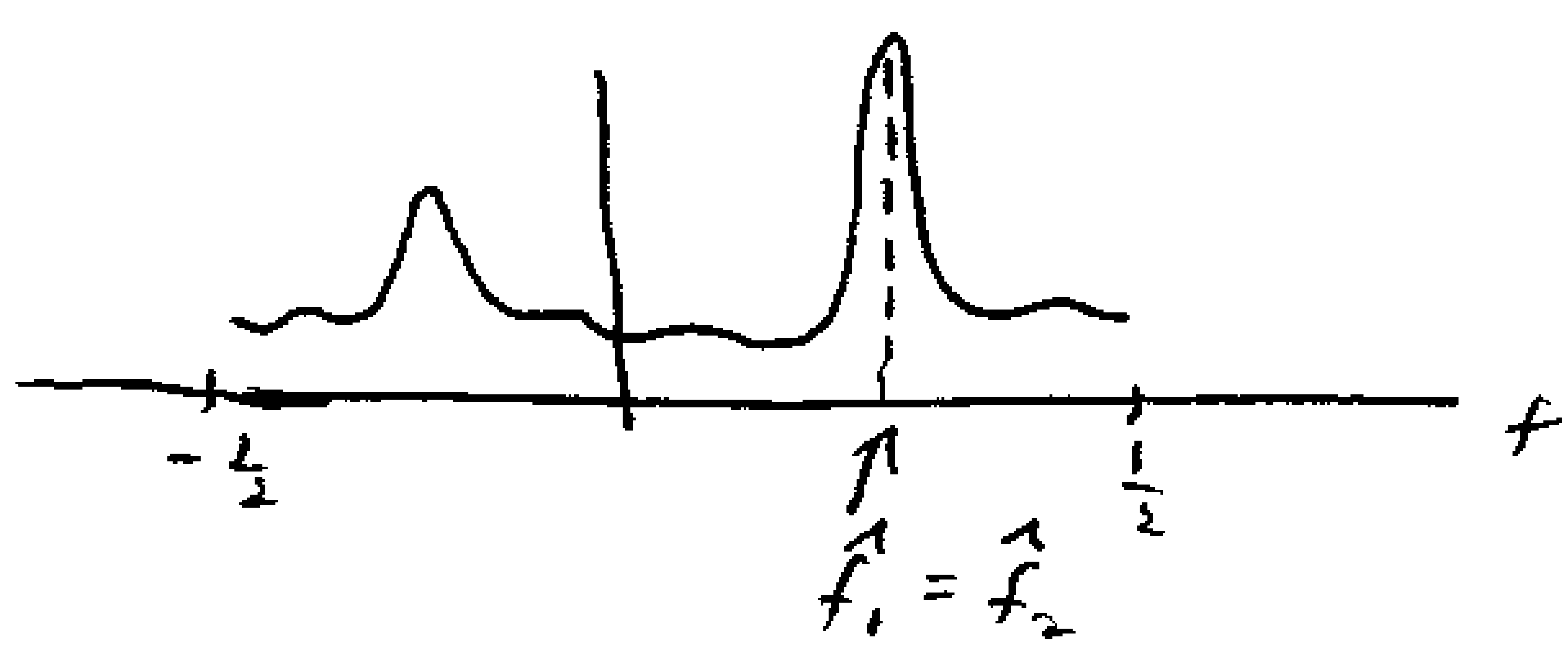
$$J(\underline{f}) \simeq 1/N \, \tilde{\underline{x}}^H \underline{E} \, \underline{E}^H \tilde{\underline{x}} = \frac{1}{N} \|\underline{E}^H \tilde{\underline{x}}\|^2$$

$$= \frac{1}{N}\left(|\underline{e}_1^H \tilde{\underline{x}}|^2 + |\underline{e}_2^H \tilde{\underline{x}}|^2\right)$$

$$= \frac{1}{N}\left|\sum_{n=0}^{N-1} \tilde{x}[n] e^{-j2\pi f_1 n}\right|^2$$

$$+ \frac{1}{N}\left|\sum_{n=0}^{N-1} \tilde{x}[n] e^{-j2\pi f_2 n}\right|^2$$

Approximate MLE finds peak locations of the two largest peaks of a single periodogram, with the constraint $|f_1 - f_2| >> 1/N$.
Without constraint we could have



22) An MLE will minimize
$$\sum_{n=0}^{N-1} |\tilde{x}[n] - \tilde{A}\tilde{s}[n]|^2$$

where $\tilde{s}[n] = e^{j2\pi(f_0 n + 1/2 \alpha n^2)}$

This is a linear LS problem with respect to $\tilde{A}$. Thus, from Example 15.2

$$\hat{\tilde{A}} = \frac{\sum_n \tilde{x}[n]\tilde{s}^*[n]}{\sum_n |\tilde{s}[n]|^2}$$

Substituting for $\tilde{A}$

$$\sum_{n=0}^{N-1} \left(\tilde{x}[n] - \hat{\tilde{A}}\,\tilde{s}[n]\right)^* \left(\tilde{x}[n] - \hat{\tilde{A}}\,\tilde{s}[n]\right)$$

$$= \sum_{n=0}^{N-1} \tilde{x}^*[n] \left(\tilde{x}[n] - \hat{\tilde{A}}\,\tilde{s}[n]\right)$$

$$- \underbrace{\hat{\tilde{A}} \sum_{n=0}^{N-1} \tilde{s}^*[n] \left(\tilde{x}[n] - \hat{\tilde{A}}\,\tilde{s}[n]\right)}_{=0}$$

$$= \sum_n |\tilde{x}[n]|^2 - \hat{\tilde{A}} \sum_n \tilde{x}^*[n]\,\tilde{s}[n]$$

$$= \sum_n |\tilde{x}[n]|^2 - \frac{\left|\sum_n \tilde{x}[n]\, s^*[n]\right|^2}{\sum_n |s[n]|^2}$$

$$= \sum_n |\tilde{x}[n]|^2 - \frac{\left|\sum_n \tilde{x}[n]\, e^{-j2\pi(f_0 n + \frac{1}{2}\alpha n^2)}\right|^2}{N}$$

Hence, we need to maximize over f_0, α

$$\left|\sum_{n=0}^{N-1} \tilde{x}[n]\, e^{-j2\pi(f_0 n + 1/2\,\alpha n^2)}\right|^2.$$ To

compute this efficiently let $\tilde{y}[n] = \tilde{x}[n]\, e^{-j\pi\alpha n^2}$

$\Rightarrow \left|\sum_{n=0}^{N-1} \tilde{y}[n]\, e^{-j2\pi f_0 n}\right|^2$, can use

FFT on $\tilde{y}[n]$ for each α.

23) $\hat{\alpha} = E(\alpha \mid \underline{u}, \underline{v}) = \int \alpha\, p(\alpha \mid \underline{u}, \underline{v})\, d\alpha$

$\hat{\beta} = E(\beta \mid \underline{u}, \underline{v}) = \int \beta\, p(\beta \mid \underline{u}, \underline{v})\, d\beta$

or $\hat{\alpha} = \iint \alpha\, p(\alpha, \beta \mid \underline{u}, \underline{v})\, d\alpha\, d\beta$

$\hat{\beta} = \iint \beta\, p(\alpha, \beta \mid \underline{u}, \underline{v})\, d\alpha\, d\beta$

$$\hat{\theta} = \hat{\alpha} + j\hat{\beta} = \iint (\alpha + j\beta)\, p(\alpha, \beta \mid \underline{u}, \underline{v})\, d\alpha\, d\beta$$
$$= \iint \theta\, p(\alpha, \beta \mid \underline{u}, \underline{v})\, d\alpha\, d\beta$$
$$= E(\theta \mid \underline{u}, \underline{v}) = E(\underline{\theta} \mid \tilde{\underline{x}})$$

Note that $p(\theta \mid \tilde{\underline{x}})$ is just $p(\alpha, \beta \mid \underline{u}, \underline{v})$ in disguise.

24) From (15.52) with $\underline{C}_{\tilde{x}} = P_0 \underline{Q}$ where $\underline{Q}$ is the covariance matrix corresponding to a process with PSD $Q(f)$,

$$I(P_0) = \operatorname{tr}\left(\underline{C}_{\tilde{x}}^{-1} \frac{\partial \underline{C}_{\tilde{x}}}{\partial P_0} \underline{C}_{\tilde{x}}^{-1} \frac{\partial \underline{C}_{\tilde{x}}}{\partial P_0} \right)$$
$$= \operatorname{tr}\left(\frac{1}{P_0} \underline{Q}^{-1} \underline{Q} \frac{1}{P_0} \underline{Q}^{-1} \underline{Q} \right)$$
$$= \frac{1}{P_0^2} \operatorname{tr}(\underline{I}) = N / P_0^2$$

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

From (15.68)

$$I(P_0) = N \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\frac{\partial \ln P_{\tilde{x}\tilde{x}}(f)}{\partial P_0} \right)^2 df$$

$$= N \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\frac{\partial \ln P_0 Q(f)}{\partial P_0} \right)^2 df$$

$$= N \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{P_0^2}\, df = N / P_0^2$$

same result (<u>not</u> true in general)

25) $$X(f_k) = \sum_n \tilde{x}(n) e^{-j2\pi f_k n} \qquad (f_k = k/N)$$

$$= \sum_n \left(\tilde{A} e^{j2\pi f_c n} + \tilde{w}(n) \right) e^{-j2\pi f_k n}$$

$$= \sum_n \tilde{w}(n) e^{-j2\pi f_k n} \qquad f_k \neq f_c$$

$$\sum_n \left(\tilde{A} + \tilde{w}(n) e^{-j2\pi f_k n} \right) \qquad f_k = f_c$$

$$= \begin{cases} W(f_k) & k \neq l \\ N\tilde{A} + W(f_k) & k = l \end{cases}$$

where $W(f_k)$ is the DFT of $\tilde{w}(n)$.
But $\underline{W} \sim CN(\underline{0}, N\sigma^2 \underline{I})$ from Sect 15.9

$$\Rightarrow \underline{X} \sim CN\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \\ N\tilde{A} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, N\sigma^2 \underline{I} \right)$$

where $\underline{X} = [\, X(f_0)\ X(f_1) \ldots X(f_{N-1}) \,]^T$

Now, $$SNR(\text{input}) = |E(\tilde{x}[n])|^2 / \sigma^2$$

$$= |\tilde{A}|^2 / \sigma^2$$

$$SNR(\text{output}) = \frac{|E(X(f_c))|^2}{\text{var}(X(f_c))}$$

$$= \frac{|N\tilde{A}|^2}{N\sigma^2} = \frac{N|\tilde{A}|^2}{\sigma^2}$$

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,

$$E\left[\frac{1}{N}|X(f_k)|^2\right] = \frac{1}{N}\operatorname{var}(X(f_k))$$
$$= \sigma^2 \quad \text{for all } k$$

and for a signal present

$$E\left[\frac{1}{N}|X(f_k)|^2\right] = \frac{1}{N}\left[\operatorname{var}(X(f_k)) + |E(X(f_k))|^2\right]$$

$$= \frac{1}{N}\left(N\sigma^2 + N^2|\tilde{A}|^2\right)$$
$$= \sigma^2 + N|\tilde{A}|^2 \qquad \text{for } k = l$$
$$= \sigma^2 \qquad \text{for } k \neq l$$

This statistic is just a sampled in frequency periodogram!