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 $\underline{h}'_{k+1} \perp \{\underline{h}_1, \ldots, \underline{h}_k\}$ 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

$$\text{or } \alpha = \frac{\underline{x}^T \underline{h}'_{k+1}}{\|\underline{h}'_{k+1}\|^2} = \frac{\underline{x}^T \underline{P}_k^\perp \underline{h}_{k+1}}{\|\underline{P}_k^\perp \underline{h}_{k+1}\|^2}$$

$$\hat{\underline{s}}_{k+1} = \underline{H}_k \hat{\underline{\theta}}_k + \frac{\underline{x}^T \underline{P}_k^\perp \underline{h}_{k+1}}{\underline{h}_{k+1}^T \underline{P}_k^\perp \underline{h}_{k+1}} \underline{P}_k^\perp \underline{h}_{k+1}$$

$$= \underline{H}_k \hat{\underline{\theta}}_k + \frac{(\underline{I} - \underline{P}_k) \underline{h}_{k+1} \underline{h}_{k+1}^T \underline{P}_k^\perp \underline{x}}{\underline{h}_{k+1}^T \underline{P}_k^\perp \underline{h}_{k+1}}$$

$$= \underline{H}_k \hat{\underline{\theta}}_k + \frac{\underline{h}_{k+1} \underline{h}_{k+1}^T \underline{P}_k^\perp \underline{x}}{\underline{h}_{k+1}^T \underline{P}_k^\perp \underline{h}_{k+1}} - \frac{\underline{H}_k (\underline{H}_k^T \underline{H}_k)^{-1} \underline{H}_k^T \cdot \underline{h}_{k+1} \underline{h}_{k+1} \underline{P}_k^\perp \underline{x}}{\underline{h}_{k+1}^T \underline{P}_k^\perp \underline{h}_{k+1}}$$

$$= [\underline{H}_k \; \underline{h}_{k+1}] \begin{bmatrix} \hat{\underline{\theta}}_k - \dfrac{(\underline{H}_k^T \underline{H}_k)^{-1} \underline{H}_k^T \cdot \underline{h}_{k+1} \underline{h}_{k+1}^T \underline{P}_k^\perp \underline{x}}{\underline{h}_{k+1}^T \underline{P}_k^\perp \underline{h}_{k+1}} \\ \dfrac{\underline{h}_{k+1}^T \underline{P}_k^\perp \underline{x}}{\cdots} \end{bmatrix}$$

$$= \underline{H}_{n+1} \, \hat{\underline{\theta}}_{k+1}$$

15) Clearly, $\hat{A}_1 = \bar{x}$. From (8.28) with

$$\underline{H}_1 = \underline{1} \qquad \underline{h}_2 = [1 \; r \ldots r^{N-1}]^T = \underline{h}$$

$$\hat{A}_2 = \hat{A}_1 - \frac{\frac{1}{N}\,\underline{1}^T \underline{h}\,\underline{h}^T \underline{P}_1^{\perp} \underline{x}}{\underline{h}^T \underline{P}_1^{\perp} \underline{h}}$$

$$\hat{B}_2 = \frac{\underline{h}^T \underline{P}_1^{\perp} \underline{x}}{\underline{h}^T \underline{P}_1^{\perp} \underline{h}}$$

$$\underline{P}_1^{\perp} = \underline{I} - \underline{1}(\underline{1}^T\underline{1})^{-1}\underline{1}^T = \underline{I} - 1/N \, \underline{1}\,\underline{1}^T$$

$$\underline{h}^T \underline{P}_1^{\perp} \underline{x} = \underline{h}^T(\underline{x} - \bar{x}\,\underline{1})$$

$$= \sum_n x[n] r^n - \bar{x} \sum_n r^n$$

$$\underline{h}^T \underline{P}_1^{\perp} \underline{h} = \underline{h}^T(\underline{I} - 1/N \, \underline{1}\,\underline{1}^T)\underline{h}$$

$$= \underline{h}^T \underline{h} - 1/N \, (\underline{1}^T \underline{h})^2$$

$$= \sum_n r^{2n} - 1/N \left(\sum_n r^n\right)^2$$

$$\hat{A}_2 = \bar{x} - \frac{\frac{1}{N} \sum_n r^n \left[\sum_n x[n] r^n - \bar{x} \sum_n r^n \right)}{\sum_n r^{2n} - \frac{1}{N}\left(\sum_n r^n\right)^2}$$

$$\hat{B}_2 = \frac{\sum_n x(n) r^n - \bar{x} \sum_n r^n}{\sum_n r^{2n} - \frac{1}{N}\left(\sum_n r^n\right)^2}$$

16)
$$\underline{H}_1 = [1\,1\,\dots\,1]^T \Rightarrow \hat{A}_1 = (\underline{H}_1^T\underline{H}_1)^{-1}\underline{H}_1^T\underline{x} = \bar{x}$$

$$J_{Min_1} = \sum_{n=0}^{N-1} x^2[n] - N\bar{x}^2$$

$$\underline{H}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ \vdots & 0 \\ \vdots & 1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix} \begin{matrix} \} M\times 1 \\ \\ \} (N-M)\times 1 \end{matrix} \quad \text{for} \quad \underline{\theta}_2 = \begin{bmatrix} A \\ B-A \end{bmatrix}$$

$$\uparrow \qquad \uparrow$$
$$\underline{H}_1 \qquad \underline{h}_2$$

$$\underline{P}_1^{\perp} = \underline{I} - \tfrac{1}{N}\underline{1}\,\underline{1}^T$$

$$\underline{h}_2^T\underline{P}_1^{\perp}\underline{x} = \underline{h}_2^T(\underline{x} - \tfrac{1}{N}\underline{1}\,\underline{1}^T\underline{x})$$
$$= \underline{h}_2^T(\underline{x} - \bar{x}\underline{1})$$
$$= \sum_{n=M}^{N-1}(x[n] - \bar{x}) = \sum_{M}^{N-1} x[n] - (N-M)\bar{x}$$

$$\underline{h}_2^T\underline{P}_1^{\perp}\underline{h}_2 = \underline{h}_2^T\underline{h}_2 - \underline{h}_2^T\underline{1}\,\underline{1}^T\underline{h}_2/N$$
$$= (N-M) - (N-M)^2/N$$
$$= (N-M)M/N$$

$$\Rightarrow \hat{A}_2 = \hat{A}_1 - \frac{\tfrac{1}{N}\underline{1}^T\underline{h}_2\left[\sum_{M}^{N-1} x[n] - (N-M)\bar{x}\right]}{(N-M)M/N}$$

$$= \bar{x} - \frac{\frac{N-M}{N}\left[\sum_{M}^{N-1} x[n] - (N-M)\bar{x}\right]}{(N-M)M/N}$$

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

$$\widehat{B-A} = \frac{\sum_{M}^{N-1} x[n] - (N-M)\bar{x}}{(N-M)M/N}$$

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 $\underline{h}_i^T \underline{P}_k^{\perp} \underline{x} = 0$ for $i = 1, 2, \ldots, k$

$$\begin{bmatrix} \underline{h}_1^T \underline{P}_k^\perp \underline{x} \\ \vdots \\ \underline{h}_k^T \underline{P}_k^\perp \underline{x} \end{bmatrix} = \begin{bmatrix} \underline{h}_1^T \\ \vdots \\ \underline{h}_k^T \end{bmatrix} \underline{P}_k^\perp \underline{x} = \underline{H}_k^T \underline{P}_k^\perp \underline{x}$$

$$= \underline{H}_k^T \left(\underline{I} - \underline{H}_k (\underline{H}_k^T \underline{H}_k)^{-1} \underline{H}_k^T \right) \underline{x}$$
$$= (\underline{H}_k^T - \underline{H}_k^T) \underline{x} = \underline{0}$$

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 $\mathcal{J} = \frac{K^2[N]}{var(\hat{A}[N-1])} + \frac{1}{\sigma_N^2} - \frac{2K[N]}{\sigma_N^2} + \frac{K^2(N)}{\sigma_N^2}$

But from (8.38), $1-K[N] = \frac{\sigma_N^2}{var(\hat{A}[N-1])+\sigma_N^2}$

$$\mathcal{J} = \frac{var(\hat{A}[N-1])}{(var(\hat{A}[N-1])+\sigma_N^2)^2} + \frac{1}{\sigma_N^2}\frac{\sigma_N^4}{(var(\hat{A}[N-1])+\sigma_N^2)^2}$$

$$= \frac{1}{var(\hat{A}[N-1])+\sigma_N^2}$$

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 $\Sigma[n] = \mathrm{var}(\hat{A}[n])$

$$\Rightarrow \quad K[n] = \frac{\mathrm{var}(\hat{A}[n-1])\, r^n}{1 + r^{2n}\,\mathrm{var}(\hat{A}[n-1])} \qquad \text{from } (8.47)$$

$$\mathrm{var}(\hat{A}[n]) = (1 - K[n]\, r^n)\,\mathrm{var}(\hat{A}[n-1]) \quad \text{from } (8.48)$$

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) $$K[N] = \frac{\mathrm{var}(\hat{A}[N-1])}{\mathrm{var}(\hat{A}[N-1]) + \sigma_N^2}$$

$$\text{var}(\hat{A}[N]) = (1 - K[N])\ \text{var}(\hat{A}[N-1])$$

Let $v_N = \text{var}(\hat{A}[N])$

$$v_N = \left(1 - \frac{v_{N-1}}{v_{N-1} + \sigma_N^2}\right) v_{N-1}$$

$$= \frac{\sigma_N^2 v_{N-1}}{v_{N-1} + \sigma_N^2}$$

$$\frac{1}{v_N} = \frac{v_{N-1} + \sigma_N^2}{\sigma_N^2 v_{N-1}} = \frac{1}{\sigma_N^2} + \frac{1}{v_{N-1}}$$

$$= \frac{1}{r^N} + \frac{1}{v_{N-1}}$$

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 $r = 1$, $\text{var}(\hat{A}[N]) \to 0$ as $N \to \infty$

$K[N] \to 0$ as $N \to \infty$

If $0 < r < 1$, $\quad var(\hat{A}[N]) \to 0 \quad$ as $N \to \infty$

$K[N] \to$ Constant as $N \to \infty$

If $r > 1$, $\quad var(\hat{A}[N]) \to \frac{r-1}{r} \quad$ as $N \to \infty$

since $\displaystyle\sum_{n=0}^{\infty} 1/r^n = \frac{1}{1-1/r} = \frac{r}{r-1}$

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) $$\hat{\underline{\theta}}_s(n) = \left(\sum_{k=-p}^{-1} 1/\sigma_k^2\, \underline{h}[k]\, \underline{h}^T[k] + \sum_{k=0}^{n} \frac{1}{\sigma_k^2}\, \underline{h}[k]\, \underline{h}^T[k] \right)^{-1}$$

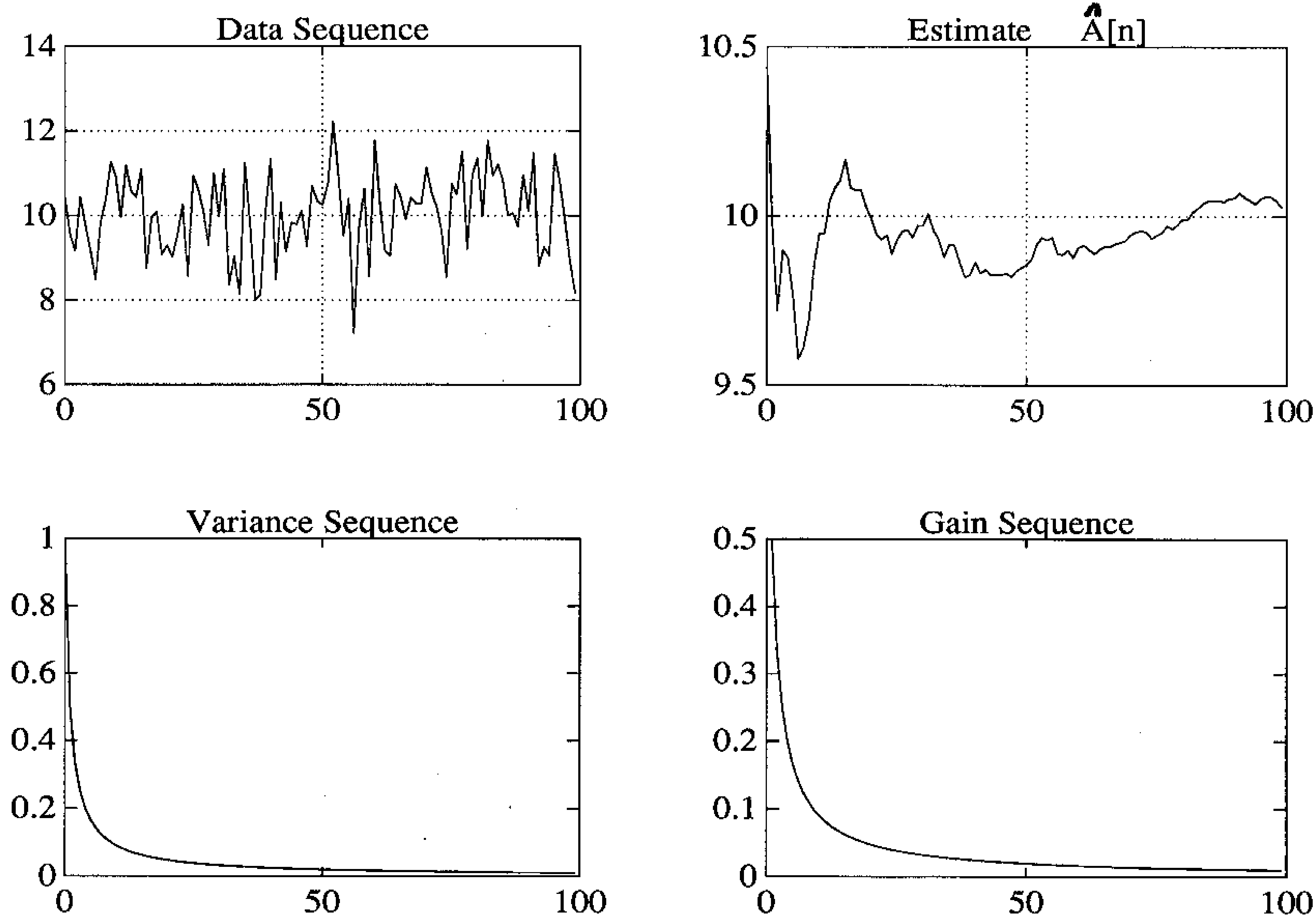
$$\cdot \left(\sum_{k=-p}^{-1} \frac{1}{\sigma_k^2}\, x[k]\, \underline{h}^T[k] + \sum_{k=0}^{n} \frac{1}{\sigma_k^2}\, x[k]\, \underline{h}^T[k] \right)$$

But $$\hat{\underline{\theta}}[-1] = \left(\sum_{k=-p}^{-1} \frac{1}{\sigma_k^2}\, \underline{h}[k]\, \underline{h}^T[k] \right)^{-1}$$

$$\cdot \left(\sum_{k=-p}^{-1} \frac{1}{\sigma_k^2}\, x[k]\, \underline{h}^T[k] \right)$$

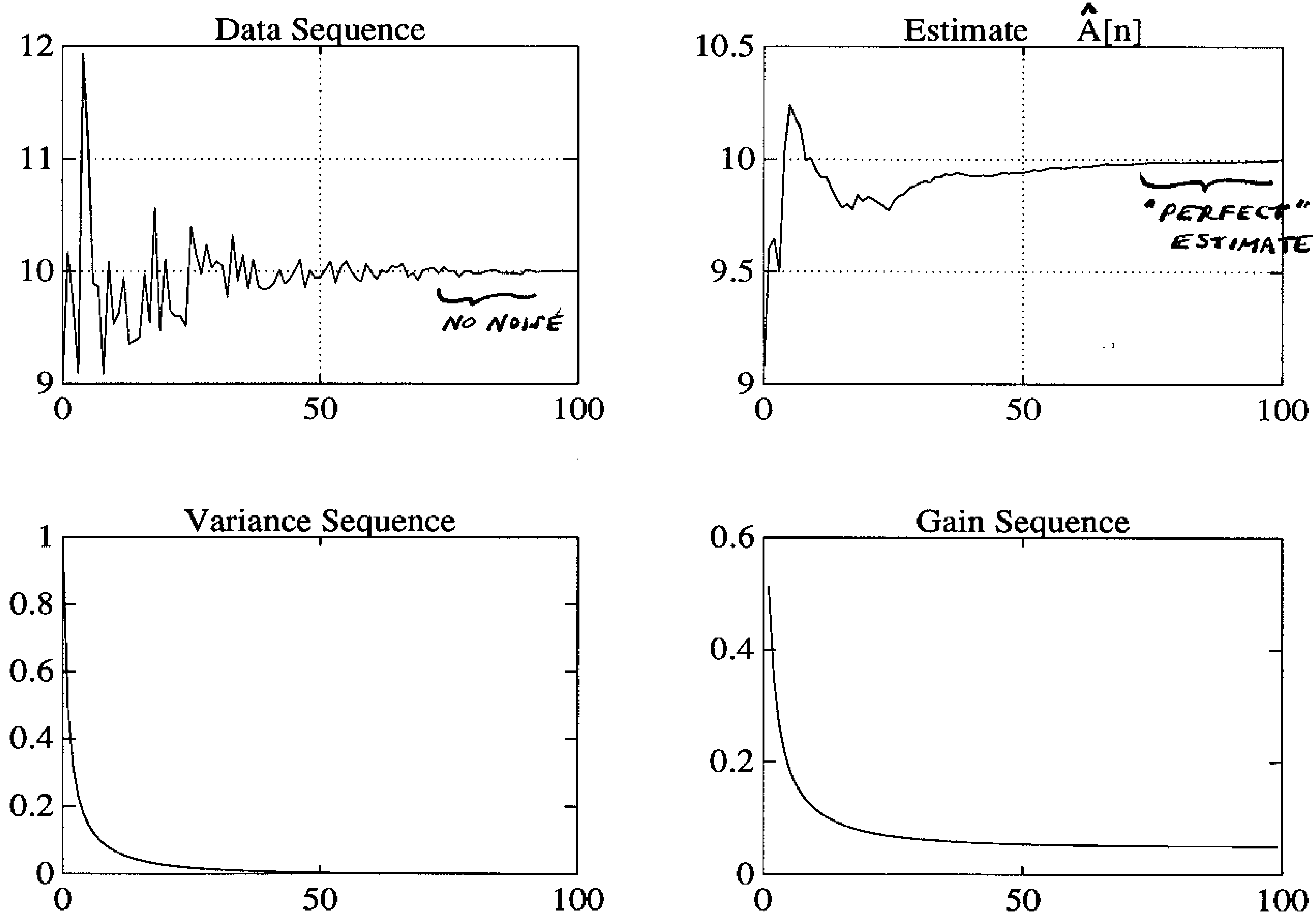
$$\underline{\Sigma}[-1] = \left(\sum_{k=-p}^{-1} \frac{1}{\sigma_k^2}\, \underline{h}[k]\, \underline{h}^T[k] \right)^{-1}$$

ob. 8.22


Data Sequence
Estimate $\hat{A}[n]$
Variance Sequence
Gain Sequence
14
12
10
8
6
10.5
10
9.5
1
0.8
0.6
0.4
0.2
0.5
0.4
0.3
0.2
0.1
0
50
100

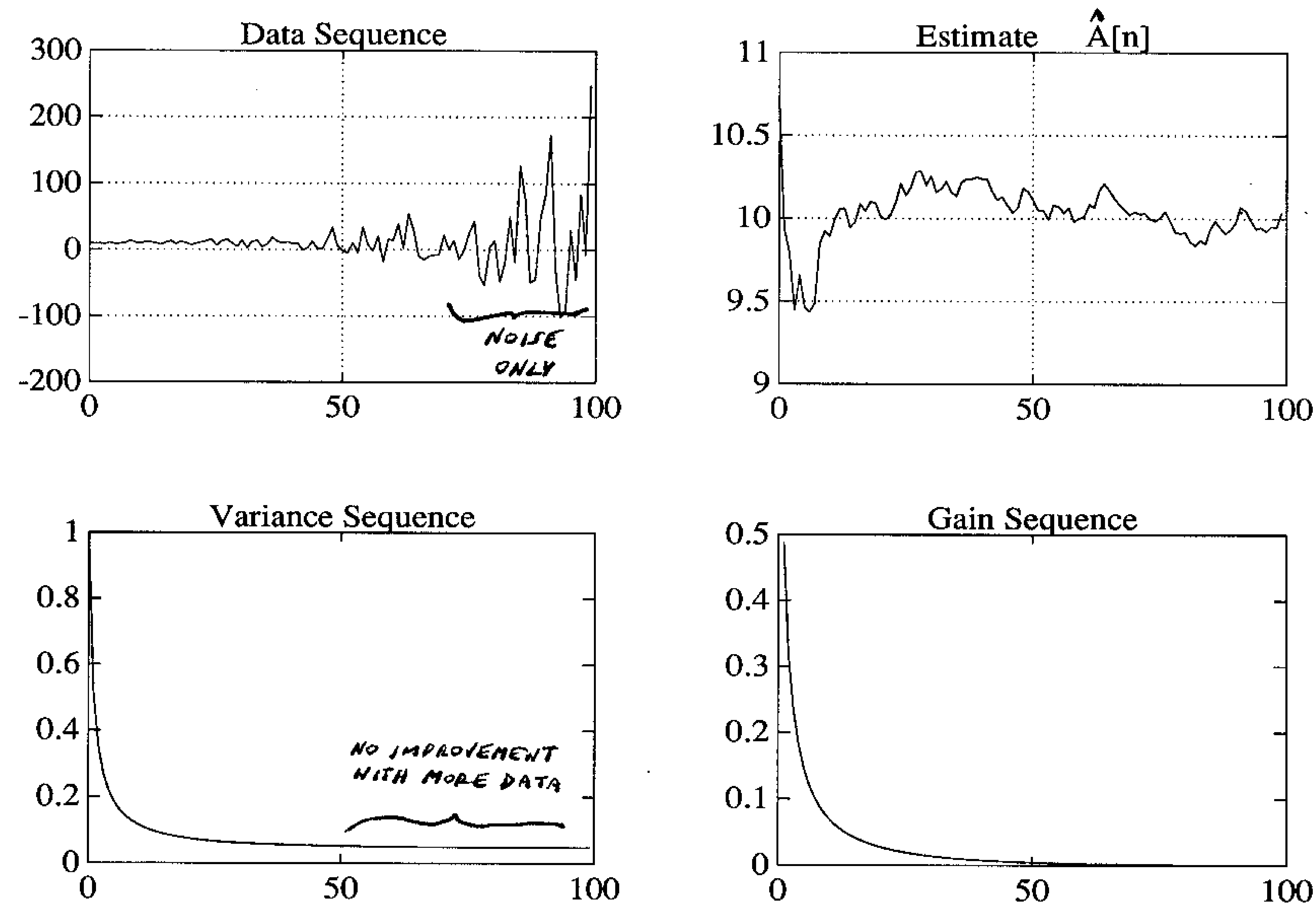
$r = 1.00$

ob. 8.22


Data Sequence
NO NOISE
Estimate $\hat{A}[n]$
"PERFECT" ESTIMATE
Variance Sequence
Gain Sequence

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

$$\hat{\underline{\theta}} = (\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{x} = \begin{bmatrix} N & 0 \\ 0 & N \end{bmatrix}^{-1}\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ \vdots & \\ 1 & 1 \end{bmatrix}^T \underline{x}$$

$$= \begin{bmatrix} \frac{1}{N}\sum_0^{N-1} x[n] \\ \frac{1}{N}\sum_0^{N-1}(-1)^n x[n] \end{bmatrix}$$

Now, if $A = B$ we have $(1\ -1)\,\underline{\theta} = 0$

or $\underline{A} = [1\ -1]$, $b = 0$ and thus from (8.52)

$$\hat{\underline{\theta}}_c = \hat{\underline{\theta}} - (\underline{H}^T\underline{H})^{-1}\underline{A}^T\left(\underline{A}(\underline{H}^T\underline{H})^{-1}\underline{A}^T\right)^{-1}\underline{A}\,\hat{\underline{\theta}}$$

$$= \hat{\underline{\theta}} - \frac{1}{N}\underline{A}^T\left(\frac{1}{N}\underline{A}\underline{A}^T\right)^{-1}\underline{A}\,\hat{\underline{\theta}}$$

$$= \left(\underline{I} - \underline{A}^T(\underline{A}\underline{A}^T)^{-1}\underline{A}\right)\hat{\underline{\theta}}$$

$$\underline{A}^T(\underline{A}\underline{A}^T)^{-1}\underline{A} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}\left([1\ -1]\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right)^{-1}(1\ -1)$$

$$= \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

$$\underline{I} - \underline{A}^T(\underline{A}\underline{A}^T)^{-1}\underline{A} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$\hat{\underline{\theta}}_c = \begin{bmatrix} \frac{1}{2}\left(\bar{x} + \frac{1}{N}\sum_{n=0}^{N-1}(-1)^n x[n]\right) \\ \frac{1}{2}\left(\bar{x} + \frac{1}{N}\sum_{n=0}^{N-1}(-1)^n x[n]\right) \end{bmatrix}$$

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

$h(g^{-1}(\hat{\alpha})) \le h(g^{-1}(\alpha))$ for all α

$\Rightarrow \quad h(\theta_0) \le h(\theta)$ for all θ

where $\theta_0 = g^{-1}(\hat{\alpha})$

But then θ_0 minimizes $h(\theta)$, ie, $\theta_0 = \hat{\theta}$

$\Rightarrow \quad \hat{\theta} = g^{-1}(\hat{\alpha})$

27) From (8.61)

$$\theta_{k+1} = \theta_k + \left(\underline{H}^T(\theta_k)\,\underline{H}(\theta_k) - \sum_{n=0}^{N-1} G_n(\theta_k)\,(x[n] - e^{\theta_k}) \right)^{-1} \cdot \underline{H}^T(\theta_k)\,(\underline{x} - e^{\theta_k}\underline{1})$$

$$[\underline{H}(\theta)]_i = \frac{\partial s[i]}{\partial \theta} = e^{\theta} \qquad i = 0, 1, \ldots, N-1$$

$$[\underline{G}_n(\theta)]_{11} = \frac{\partial^2 s[n]}{\partial \theta^2} = e^{\theta}$$

$$\Rightarrow \quad \underline{H}(\theta) = e^{\theta}\underline{1} \qquad G_n(\theta) = e^{\theta}$$

$$\theta_{k+1} = \theta_k + \left(N e^{2\theta_k} - \sum_{n=0}^{N-1} e^{\theta_k} (x[n] - e^{\theta_k}) \right)^{-1}$$

$$\cdot \; e^{\theta_k} \underline{1}^T (\underline{x} - e^{\theta_k} \underline{1})$$

$$= \theta_k + \frac{e^{\theta_k}(N\bar{x} - N e^{\theta_k})}{N e^{2\theta_k} - N\bar{x} e^{\theta_k} + N e^{2\theta_k}}$$

$$= \theta_k + \frac{\bar{x} - e^{\theta_k}}{2e^{\theta_k} - \bar{x}}$$

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 $G(z)A(z) = 1 \Rightarrow g[n] \star a[n] = \delta[n]$

$$[\underline{A}^T\underline{G}]_{ij} = \sum_{k=1}^{N} (\underline{A}^T)_{ik} (\underline{G})_{kj} \qquad \begin{array}{l} i = 1, 2, \ldots, N-p \\ j = 1, 2, \ldots, q+1 \\ \quad = 1, 2, \ldots, p \end{array}$$

$$= \sum_{k=1}^{N} a[p+i-k]\, g[k-j]$$

$$= \sum_{k=-\infty}^{\infty} a[p+i-k]\, g[k-j]$$

where $a[n] = 0$ for $n < 0$, $n > p$
$g[n] = 0$ for $n < 0$

$$= \sum_{l=-\infty}^{\infty} g[l]\, a[p+i-j-l]$$

$$= g[n] \star a[n] \Big|_{n=p+i-j}$$

$$= \delta[p+i-j]$$

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 $f_{o_{k+1}} = f_{o_k}$, $\phi_{k+1} = \phi_k$, we have

$$\sum_{n=-M}^{M} n\,x[n]\sin(2\pi\hat{f}_o n + \hat{\phi}) = 0$$

$$\sum_{n=-M}^{M} x[n]\sin(2\pi\hat{f}_o n + \hat{\phi}) = 0$$

At high SNR we have

$$\sum_n n\cos(2\pi f_o n + \phi)\sin(2\pi\hat{f}_o n + \hat{\phi}) = 0$$
$$\sum_n \cos(2\pi f_o n + \phi)\sin(2\pi\hat{f}_o n + \hat{\phi}) = 0$$

$$\Rightarrow$$

$$\frac{1}{2}\sum_n n\Big[\sin(2\pi(f_o+\hat{f}_o)n + \phi + \hat{\phi}) + \sin(2\pi(\hat{f}_o - f_o)n + (\hat{\phi}-\phi))\Big] = 0$$

$$\frac{1}{2}\sum_n \sin(2\pi(f_o+\hat{f}_o)n + \phi + \hat{\phi}) + \sin(2\pi(\hat{f}_o - f_o)n + (\hat{\phi}-\phi)) = 0$$

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) $$\rho = \operatorname{cov}(u, v) \quad \text{where} \quad \underline{x} = \begin{bmatrix} u \\ v \end{bmatrix}$$
$$= E(uv)$$

$$\Rightarrow \quad \hat{\rho} = 1/N \sum_{n=0}^{N-1} u_n v_n \quad \text{where} \quad \underline{x}[n] = \begin{bmatrix} u_n \\ v_n \end{bmatrix}$$

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

$$\hat{r}_{xx}[k] = \frac{1}{N}\sum_{n=0}^{N-1-|k|} x[n]\,x[n+|k|]$$

$$\Rightarrow \quad \hat{r}_{xx}[n] = -\sum_{k=1}^{p} a[k]\,\hat{r}_{xx}[n-k] \qquad n > q$$

Choosing $n = q+1, \ldots, q+p$ yields the linear equations

$$\begin{bmatrix} \hat{r}_{xx}[q] & \cdots & \hat{r}_{xx}[q-p+1] \\ \\ \hat{r}_{xx}[q+p-1] & \cdots & \hat{r}_{xx}[q] \end{bmatrix} \begin{bmatrix} a[1] \\ \vdots \\ a[p] \end{bmatrix} = -\begin{bmatrix} \hat{r}_{xx}[q+1] \\ \vdots \\ \hat{r}_{xx}[q+p] \end{bmatrix}$$

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)

$$\mathrm{var}(\hat{\theta}) = \left(\frac{\partial g}{\partial T}\Big|_{T=\mu}\right)^2 \mathrm{var}(T)$$

$$= \left(2T\big|_{T=\mu}\right)^2 \mathrm{var}(T)$$

$$= (2A)^2\sigma^2/N = 4A^2\sigma^2/N$$

7) $$E(x[n]) = \cos\phi$$

$$\Rightarrow \quad \phi = \arccos E(x[n])$$

or

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

$\hat{\phi} = h(\underline{w})$ 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)

$$\mathrm{var}(\hat{\phi}) = \left.\frac{\partial h}{\partial \underline{w}}^T\right|_{\underline{w}=\underline{0}} \sigma^2 \underline{I} \left.\frac{\partial h}{\partial \underline{w}}\right|_{\underline{w}=\underline{0}}$$

$$\frac{\partial h}{\partial w[i]} = -\frac{1}{\sqrt{1-u^2}}\frac{\partial u}{\partial w[i]}$$

where $u = \frac{1}{N}\sum_{n=0}^{N-1}(\cos\phi + w[n])$

$$\left.\frac{\partial h}{\partial w[i]}\right|_{\underline{w}=\underline{0}} = -\frac{1}{\sqrt{1-\cos^2\phi}}\frac{1}{N} = -\frac{1}{N\sin\phi}$$

$$\mathrm{var}(\hat{\phi}) = \sigma^2\sum_{n=0}^{N-1}\left(\left.\frac{\partial h}{\partial w[n]}\right|_{\underline{w}=\underline{0}}\right)^2$$

$$= \sigma^2\sum_{n=0}^{N-1}\frac{1}{N^2\sin^2\phi} = \frac{\sigma^2}{N\sin^2\phi}$$

8) $$\hat{\theta} = g(\underline{T})$$

$$\approx g(\underline{\mu}) + \sum_{k=1}^{r} \frac{\partial g}{\partial T_k}\Big|_{\underline{T}=\underline{\mu}} (T_k - \mu_k)$$

$$+ \frac{1}{2}(\underline{T}-\underline{\mu})^T \frac{\partial^2 g}{\partial \underline{T}\partial \underline{T}^T}\Big|_{\underline{T}=\underline{\mu}} (\underline{T}-\underline{\mu})$$

where $\frac{\partial^2 g}{\partial \underline{T}\partial \underline{T}^T}$ is the Hessian

$$\begin{aligned} E(\hat{\theta}) &= g(\underline{\mu}) + \tfrac{1}{2} E\left[(\underline{T}-\underline{\mu})^T \underline{G}(\mu)(\underline{T}-\underline{\mu})\right] \\ &= g(\underline{\mu}) + \tfrac{1}{2} E\left[\operatorname{tr}\left((\underline{T}-\underline{\mu})^T \underline{G}(\mu)(\underline{T}-\underline{\mu})\right)\right] \\ &= g(\underline{\mu}) + \tfrac{1}{2} \operatorname{tr} E\left[G(\underline{\mu})(\underline{T}-\underline{\mu})(\underline{T}-\underline{\mu})^T\right] \\ &= g(\underline{\mu}) + \tfrac{1}{2} \operatorname{tr}(\underline{G}(\mu)\underline{C}_T) \end{aligned}$$

9) $$\hat{\theta} \approx g(\underline{\mu}) + \frac{\partial g}{\partial \underline{T}}\Big|_{\underline{T}=\underline{\mu}}^T (\underline{T}-\underline{\mu})$$

$$+ \tfrac{1}{2}(\underline{T}-\underline{\mu})\underline{G}(\underline{\mu})(\underline{T}-\underline{\mu})$$

$$\begin{aligned} \operatorname{var}(\hat{\theta}) &= E\left[(\hat{\theta} - E(\hat{\theta}))^2\right] \\ &= E\left[\left(\frac{\partial g}{\partial T}\Big|_{\underline{T}=\underline{\mu}}^T (\underline{T}-\underline{\mu}) + \tfrac{1}{2}(\underline{T}-\underline{\mu})^T \underline{G}(\underline{\mu})(\underline{T}-\underline{\mu})\right.\right. \\ &\qquad \left.\left. - \tfrac{1}{2}\operatorname{tr}(\underline{G}(\underline{\mu})\underline{C}_T)\right)^2\right] \end{aligned}$$

using the results from Prob 9.8

But $\underline{T}-\underline{\mu} \sim N(\underline{0}, C_T)$ so that all odd-order moments are zero. Let $\underline{x} = \underline{T}-\underline{\mu}$, $\underline{b} = \frac{\partial g}{\partial \underline{T}}\Big|_{T=\mu}$ 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 $\mu = E(T_1) = 1/\lambda$

$$C_T = \mathrm{var}(T_1) = \mathrm{var}(x(n))/N = \frac{1}{N\lambda^2}$$

$$G(\mu) = \left.\frac{\partial^2 g}{\partial T^2}\right|_{T=\mu}$$
$$= \left.\frac{2}{T^3}\right|_{T=\mu} = \frac{2}{\mu^3} = 2\lambda^3$$

$$E(\hat{\lambda}) = \lambda + \frac{1}{2}(2\lambda^3)\frac{1}{N\lambda^2} = \lambda + \lambda/N$$

$$= \lambda(1 + 1/N)$$

$$\mathrm{var}(\hat{\lambda}) = \left(\frac{\partial g}{\partial T}\Big|_{T=\mu}\right)^2 \mathrm{var}(T)$$
$$+ \frac{1}{2}\,\mathrm{tr}\left[(G(\mu)C_T)^2\right]$$

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

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) $$\frac{1}{N-1}\sum_{n=0}^{N-2} A\cos(2\pi f_0 n + \phi)\, A\cos(2\pi f_0(n+1) + \phi)$$

$$= \frac{A^2}{2(N-1)}\sum_{n=0}^{N-2}\left[\cos(4\pi f_0 n + 2\pi f_0 + 2\phi)\right.$$
$$\left. + \cos(2\pi f_0)\right]$$

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

$$\hat{f}_0 = \frac{1}{2\pi} \arccos \left[\frac{\hat{r}_{xx}(1)}{\hat{r}_{xx}[0]} \right] \text{ so that}$$

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) $$\mathrm{var}(\hat{f}_0) = \frac{\sigma^2}{(2\pi)^2 (N-1)^2 \sin^2 2\pi f_0}$$

$$\cdot \left[s^2[1] + 4\cos^2 2\pi f_0 \sum_{n=1}^{N-2} s^2[n] + s^2[N-2] \right]$$

$f_0 = 0.25 \Rightarrow \cos 2\pi f_0 = 0$

$$s[1] = \sqrt{2} \cos 2\pi f_0 = 0$$

$$s[N-2] = \sqrt{2} \cos 2\pi \tfrac{1}{4}(N-2)$$

$$= \sqrt{2} \cos \tfrac{\pi}{2}(N-2) = 0$$

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

$$\Rightarrow \qquad = p_{\underline{w}}(x[0]-A, x[1]-A)$$

and $w[0]$ is independent of $w[1]$

$$\Rightarrow \qquad = p_w(x[0]-A) \, p_w(x[1]-A)$$

$$= p_{x[0]}(x[0] \mid A) \, p_{x[1]}(x[1] \mid A)$$

$$\underline{x} = \begin{bmatrix} A + w[0] \\ A + w[1] \end{bmatrix} = A\underline{1} + \underline{w} \sim N(\underline{0}, \underline{1}\underline{1}^T + \underline{I})$$

since $A, w[0], w[1]$ are IID and $\sim N(0,1)$

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) $$p(\underline{x}|\theta) = e^{-\sum_n (x[n]-\theta)} \qquad \text{all } x[n]\text{'s} > \theta$$

$$= e^{-N(\bar{x}-\theta)} \qquad \xi = \min x[n] > \theta$$

$$p(\theta|\underline{x}) = \frac{p(\underline{x}|\theta)\,p(\theta)}{\int p(\underline{x}|\theta)\,p(\theta)\,d\theta}$$

$$= \frac{e^{-N(\bar{x}-\theta)}e^{-\theta}}{\int_0^{\xi} e^{-N(\bar{x}-\theta)}e^{-\theta}\,d\theta}$$

$$= \frac{e^{\theta(N-1)}}{\left.\frac{e^{\theta(N-1)}}{N-1}\right|_0^{\xi}} = \begin{cases} \dfrac{(N-1)e^{\theta(N-1)}}{e^{(N-1)\xi}-1} & 0 < \theta < \xi \\ 0 & \text{otherwise} \end{cases}$$

$$\hat{\theta} = E(\theta|\underline{x}) = \frac{\int_0^{\xi} \theta(N-1)e^{\theta(N-1)}\,d\theta}{e^{(N-1)\xi}-1}$$

$$= \frac{N-1}{e^{(N-1)\xi}-1}\left[\left.\frac{\theta(N-1)-1}{(N-1)^2}e^{\theta(N-1)}\right|_0^{\xi}\right)$$

$$= \frac{1}{(N-1)(e^{(N-1)\mathfrak{z}}-1)}\left[(\mathfrak{z}(N-1)-1)\, e^{(N-1)\mathfrak{z}} + 1 \right]$$

$$= \frac{1}{(N-1)(e^{(N-1)\mathfrak{z}}-1)}\left[1 - e^{(N-1)\mathfrak{z}} + (N-1)\mathfrak{z}\, e^{(N-1)\mathfrak{z}} \right]$$

$$= \frac{\mathfrak{z}\, e^{(N-1)\mathfrak{z}}}{e^{(N-1)\mathfrak{z}}-1} - \frac{1}{N-1}$$

$$\hat{\theta} = \frac{\min x[n]}{1 - e^{-(N-1)\min x[n]}} - \frac{1}{N-1}$$

4) $p(\underline{x} \mid \theta) = \frac{1}{\theta^N}$ 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 $\mathfrak{z} \le \theta \le \beta$

$$p(\theta \mid \underline{x}) = \frac{p(\underline{x} \mid \theta)\, p(\theta)}{\int p(\underline{x} \mid \theta)\, p(\theta)\, d\theta}$$

$$= \frac{\frac{1}{\theta^N}\frac{1}{\beta}}{\int_{\mathfrak{z}}^{\beta} \frac{1}{\theta^N}\frac{1}{\beta}\, d\theta} = \frac{\frac{1}{\theta^N}}{-\frac{\theta^{-(N-1)}}{N-1}\Big|_{\mathfrak{z}}^{\beta}}$$

$$= \frac{-(N-1)\, 1/\theta^N}{\beta^{-(N-1)} - \mathfrak{z}^{-(N-1)}}$$

$$= \frac{1/\theta^N}{\frac{1}{(N-1)}\left(\mathfrak{z}^{-(N-1)} - \beta^{-(N-1)}\right)}$$

$$\hat{\theta} = E(\theta|\underline{x}) = \int \theta p(\theta|\underline{x})\,d\theta$$

$$= c\int_{\xi}^{\beta} \theta \frac{d\theta}{\theta^N} \qquad c = \frac{1}{\frac{1}{N-1}\left(\xi^{-(N-1)} - \beta^{-(N-1)}\right)}$$

$$= c\left.\frac{\theta^{-(N-2)}}{-(N-2)}\right|_{\xi}^{\beta} = -\frac{c}{N-2}\left(\beta^{-(N-2)} - \xi^{-(N-2)}\right)$$

$$= \frac{(\max x[n])^{-(N-2)} - \beta^{-(N-2)}}{(\max x[n])^{-(N-1)} - \beta^{-(N-1)}} \quad \frac{N-1}{N-2}$$

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.

5) $$\text{Bmse}(\hat{\theta}) = E_{x,\theta}\left\{\left[(\theta - E(\theta|\underline{x})) + (E(\theta|\underline{x}) - \hat{\theta})\right]^2\right\}$$

$$= E_x\left[E_{\theta|x}\left\{\left[(\theta - E(\theta|\underline{x})) + (E(\theta|\underline{x}) - \hat{\theta})\right]^2\right\}\right]$$

$$= E_x\Big\{E_{\theta|x}\left[(\theta - E(\theta|\underline{x}))^2\right] + 2E_{\theta|x}\left[(\theta - E(\theta|\underline{x}))(E(\theta|\underline{x}) - \hat{\theta})\right] + E_{\theta|x}\left[(E(\theta|\underline{x}) - \hat{\theta})^2\right]\Big\}$$

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.

$$p(w[n]|A) = \begin{cases}\frac{1}{\sqrt{2\pi\sigma_+^2}}e^{-\frac{1}{2\sigma_+^2}w^2[n]} & A\geq 0\\ \frac{1}{\sqrt{2\pi\sigma_-^2}}e^{-\frac{1}{2\sigma_-^2}w^2[n]} & A<0\end{cases}$$

$$\Rightarrow p(x[n]|A) = \begin{cases}\frac{1}{\sqrt{2\pi\sigma_+^2}}e^{-\frac{1}{2\sigma_+^2}(x[n]-A)^2} & A\geq 0\\ \frac{1}{\sqrt{2\pi\sigma_-^2}}e^{-\frac{1}{2\sigma_-^2}(x[n]-A)^2} & A<0\end{cases}$$

or

$$p(\underline{x}|A) = \begin{cases}\frac{1}{(2\pi\sigma_+^2)^{N/2}}e^{-\frac{1}{2\sigma_+^2}\sum_n(x[n]-A)^2} & A\geq 0\\ \frac{1}{(2\pi\sigma_-^2)^{N/2}}e^{-\frac{1}{2\sigma_-^2}\sum_n(x[n]-A)^2} & A<0\end{cases}$$

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

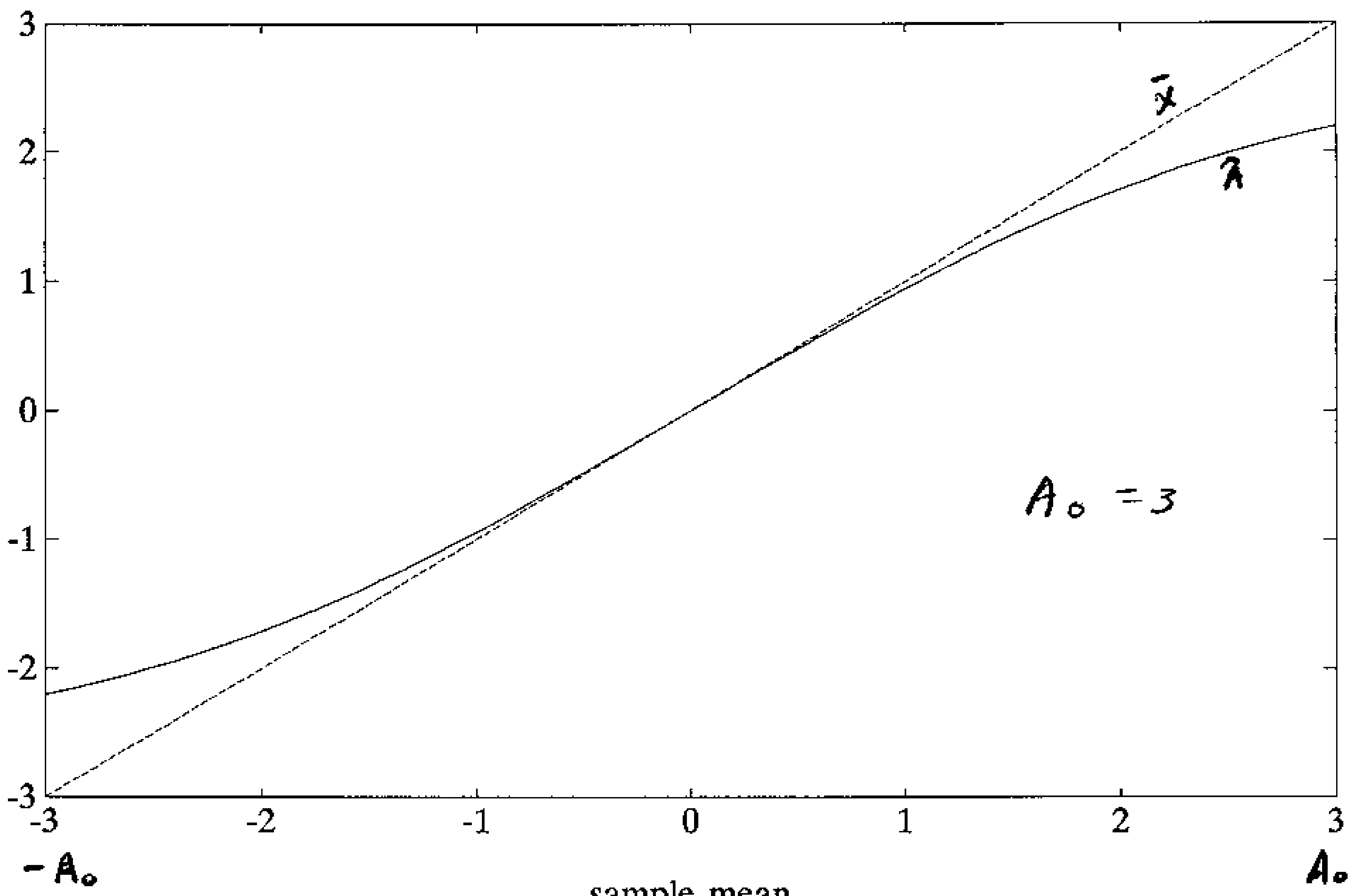
$$-\frac{1}{\sqrt{2\pi\sigma^2/N}} \frac{1}{2\frac{1}{2\sigma^2/N}} \left(e^{-\frac{1}{2\sigma^2/N}(A_0-\bar{x})^2} - e^{-\frac{1}{2\sigma^2/N}(A_0+\bar{x})^2} \right)$$
$$+ \frac{\bar{x}}{\sqrt{2\pi\sigma^2/N}} \int_{-A_0}^{A_0} e^{-\frac{1}{2\sigma^2/N}(A-\bar{x})^2} dA$$

or letting $\Phi(x)$ be the CDF for a $N(0,1)$ random variable

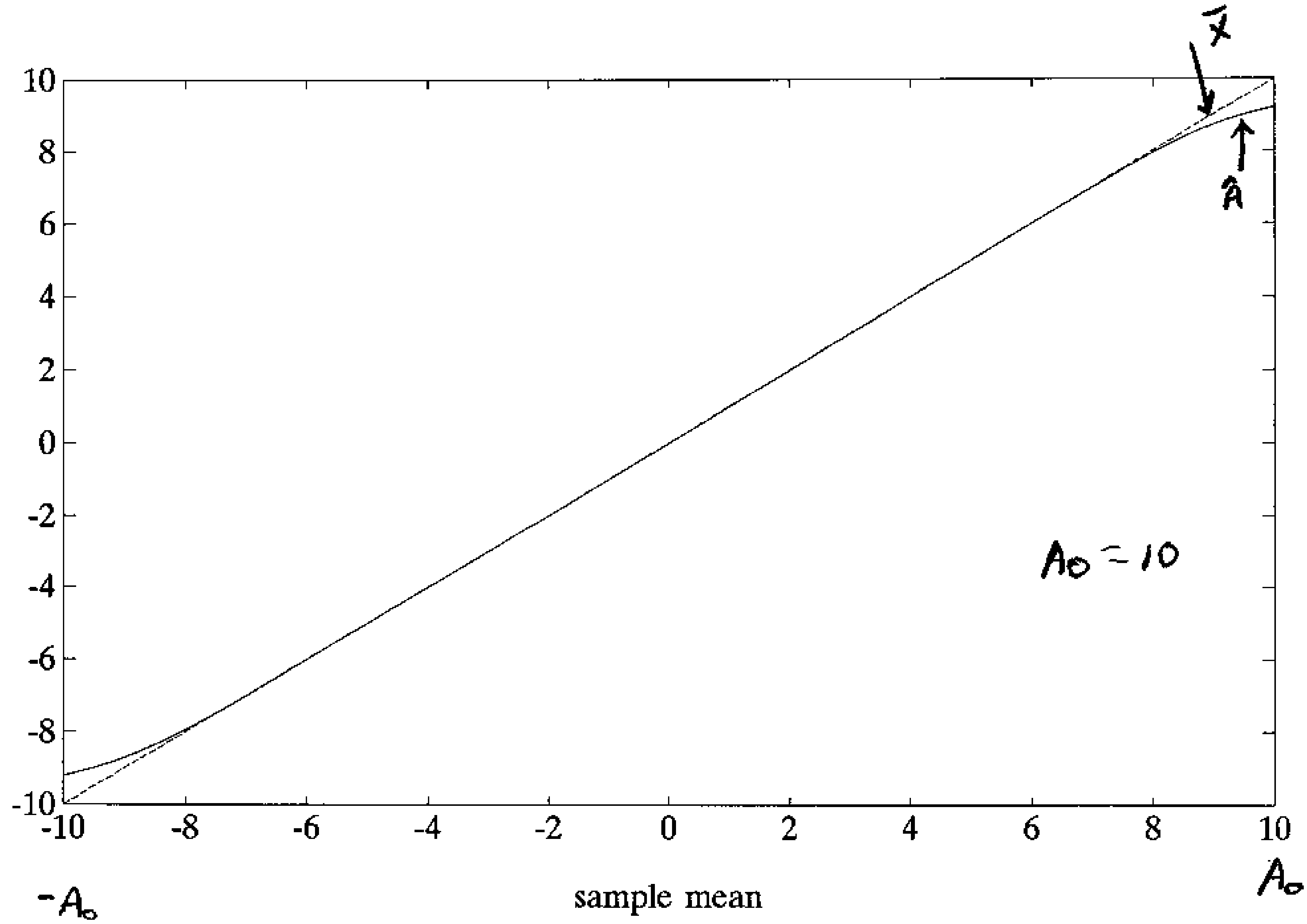
$$\hat{A} = \bar{x} + \frac{\sqrt{\frac{\sigma^2/N}{2\pi}}\left[e^{-\frac{1}{2\sigma^2/N}(A_0+\bar{x})^2} - e^{-\frac{1}{2\sigma^2/N}(A_0-\bar{x})^2}\right]}{\Phi\left(\frac{A_0-\bar{x}}{\sqrt{\sigma^2/N}}\right) - \Phi\left(\frac{-A_0-\bar{x}}{\sqrt{\sigma^2/N}}\right)}$$

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 $|\bar{x}| \leq A_0$. 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,

$$E[(\theta - E(\theta))(E(\theta) - \hat{\theta})] = E[(\theta - E(\theta))](E(\theta) - \hat{\theta})$$

$$= (E(\theta) - E(\theta))(E(\theta) - \hat{\theta}) = 0$$

$$J = E\left[(\theta - E(\theta))^2\right] + (E(\theta) - \hat{\theta})^2$$
$$\geq E\left[(\theta - E(\theta))^2\right]$$

$\Rightarrow$ $\hat{\theta} = E(\theta)$. The minimum MSE is just $E\left[(\theta - E(\theta))^2\right] = \text{var}(\theta)$.

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: $N(\underset{\mu_R}{100}, \underset{\sigma_R^2}{0.011})$

Data model: $x[n] = R + w[n]$ $n = 0, 1, \ldots, N-1$
where $w[n] \sim N(0, \underset{\sigma^2}{1})$ and $w[n]$'s are independent

This is just Example 10.1.

$$\Rightarrow \text{Bmse}(\hat{R}) = \frac{\sigma_R^2 \sigma^2/N}{\sigma_R^2 + \sigma^2/N}$$

For the error to be 0.1 on the average
$\Rightarrow$ $\text{Bmse}(\hat{R}) = 0.01$

$$0.01 = \frac{0.011/N}{0.011 + 1/N} \Rightarrow N = 9.09$$

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 $\int_0^\infty p(\theta)d\theta = 1 \Rightarrow \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\lambda\theta} d\theta = 1$

$$\int_0^\infty \theta^{\overbrace{N+\alpha}^{\alpha'}-1} e^{-\underbrace{(N\bar{x}+\lambda)}_{\lambda'}\theta} d\theta = \frac{\Gamma(\alpha')}{\lambda'^{\alpha'}}$$

$$= \frac{\Gamma(N+\alpha)}{(N\bar{x}+\lambda)^{N+\alpha}}$$

$$p(\theta|\underline{x}) = \begin{cases} \frac{(N\bar{x}+\lambda)^{N+\alpha}}{\Gamma(N+\alpha)} \theta^{N+\alpha-1} e^{-(N\bar{x}+\lambda)\theta} & \theta > 0 \\ 0 & \theta < 0 \end{cases}$$

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

$$g(y) = p(x_0, y) = \frac{1}{2\pi \det^{1/2}(\underline{C})} e^{-\frac{1}{2}h(y)}$$

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 $E(x) = E(y) = 0$, $Cov(x,y) = \rho$, $var(x) = 1$

$$\Rightarrow E(y|x) = \rho x$$

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

$\underline{H} = [1 \ r \ \ldots \ r^{N-1}]^T$

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.

$$
\begin{aligned}
I &= H(\theta) - H(\theta|\underline{x}) \\
&= -\int p(\theta)\ln p(\theta)\,d\theta + \iint p(\underline{x},\theta)\ln p(\theta|\underline{x})\,d\underline{x}\,d\theta \\
&= -\iint p(\underline{x},\theta)\,d\underline{x}\,\ln p(\theta)\,d\theta + \quad '' \\
&= -\iint p(\underline{x},\theta)\ln p(\theta)\,d\underline{x}\,d\theta \\
&\quad + \iint p(\underline{x},\theta)\ln p(\theta|\underline{x})\,d\underline{x}\,d\theta \\
&= \iint p(\underline{x},\theta)\ln\frac{p(\theta|\underline{x})}{p(\theta)}\,d\underline{x}\,d\theta \\
&= \int \underbrace{\int p(\theta|\underline{x})\ln\frac{p(\theta|\underline{x})}{p(\theta)}\,d\theta}_{\geq 0 \text{ by given inequality}}\; p(\underline{x})\,d\underline{x}
\end{aligned}
$$

≥ 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.

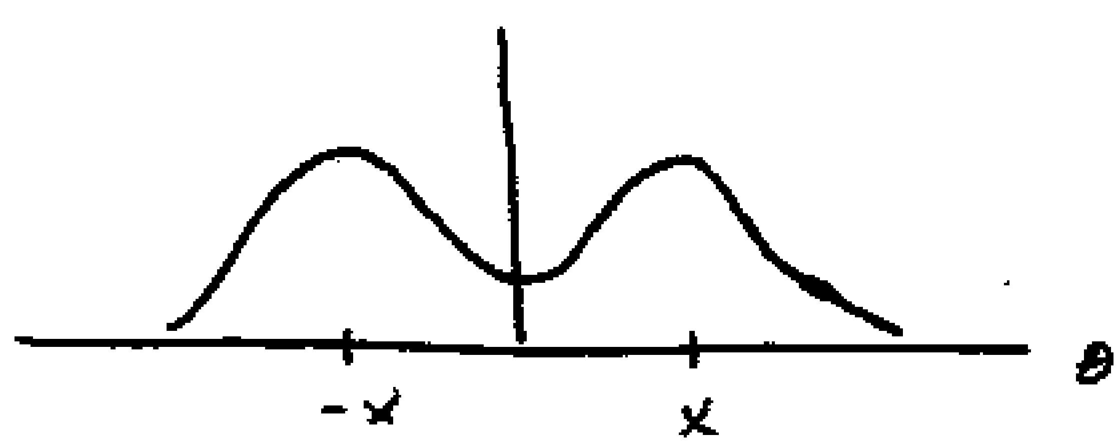
Chapter 11

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

$$\hat{\theta} = E(\theta|x) = \int_{-\infty}^{\infty} \theta \frac{\epsilon}{\sqrt{2\pi}} e^{-\frac{1}{2}(\theta-x)^2} d\theta$$

$$+ \int_{-\infty}^{\infty} \theta \frac{1-\epsilon}{\sqrt{2\pi}} e^{-\frac{1}{2}(\theta+x)^2} d\theta$$

$$= \epsilon x + (1-\epsilon)(-x) = x(2\epsilon - 1) = x/2$$

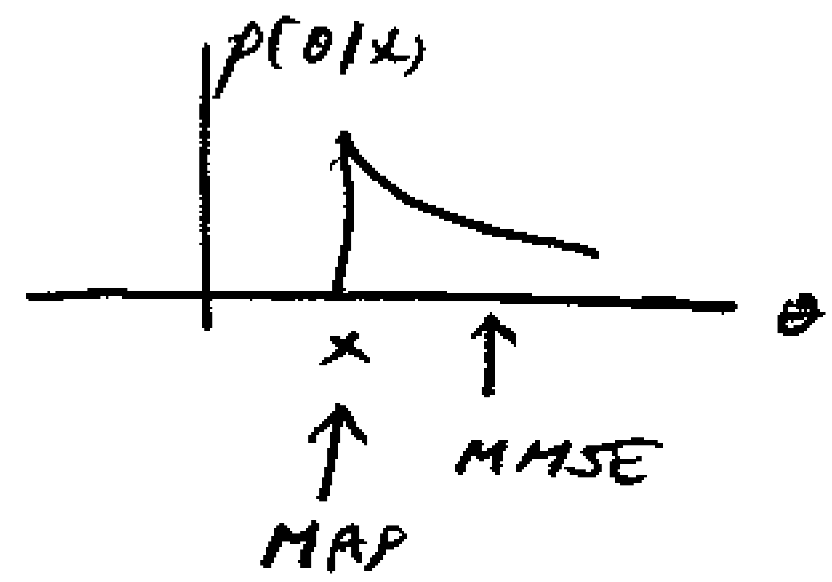
The MAP estimator is $\hat{\theta} = x$. Note that MAP $\neq$ MMSE.

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 $\hat{\theta} = x$



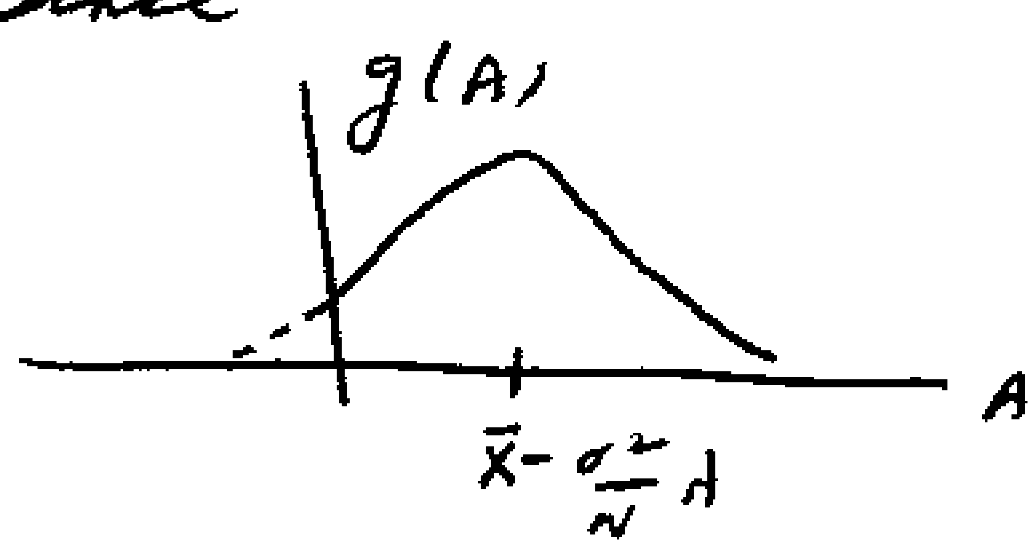
4) $g(A) = p(\underline{x} \mid A) p(A)$

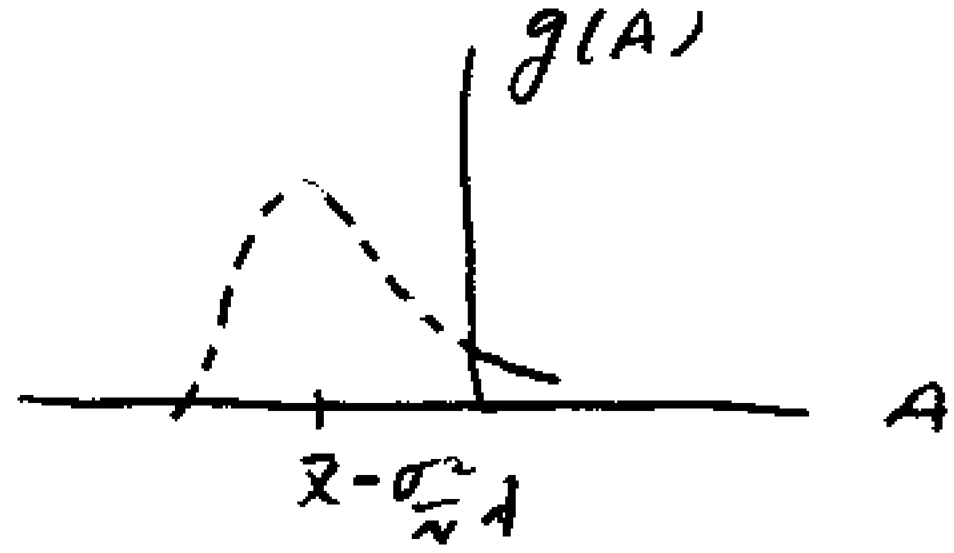
$$= \begin{cases} \dfrac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} \sum_n (x(n) - A)^2} \lambda e^{-\lambda A} & A > 0 \\ 0 & A < 0 \end{cases}$$

$$\frac{\partial \ln g}{\partial A} = \frac{1}{\sigma^2} \sum_n (x(n) - A) - \lambda = 0$$

$$\Rightarrow \hat{A} = \bar{x} - \frac{\sigma^2}{N} \lambda$$

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 $\hat{\underline{\theta}}(n) = \underline{A}^n \hat{\underline{\theta}}(0)$

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) $$\mathcal{R} = E[c(\underline{\epsilon})]$$
$$= \iint c(\underline{\epsilon})\, p(\underline{x},\underline{\theta})\, d\underline{x}\, d\underline{\theta}$$
$$= \int \underbrace{\int c(\underline{\epsilon})\, p(\underline{\theta}|\underline{x})\, d\underline{\theta}}_{\text{minimize over } \hat{\underline{\theta}}}\; p(\underline{x})\, d\underline{x}$$

$$\int c(\underline{\epsilon})\, p(\underline{\theta}|\underline{x})\, d\underline{\theta} = \int_{\{\underline{\theta}:\, \|\underline{\epsilon}\| > \delta\}} p(\underline{\theta}|\underline{x})\, d\underline{\theta}$$

$$= \int_{\{\underline{\theta}:\, \|\underline{\theta}-\hat{\underline{\theta}}\| > \delta\}} p(\underline{\theta}|\underline{x})\, d\underline{\theta} = \int_{\{\underline{\theta}:\, \|\underline{\theta}-\hat{\underline{\theta}}\| \le \delta\}^c} p(\underline{\theta}|\underline{x})\, d\underline{\theta}$$

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 $\underline{\alpha} = \underline{A}\,\underline{\theta}$, $\dfrac{\partial \underline{\alpha}}{\partial \underline{\theta}} = \underline{A}$

$$p(\underline{x},\underline{\alpha}) = \frac{p(\underline{x},\underline{\theta})}{\left|\det \frac{\partial \underline{\alpha}}{\partial \underline{\theta}}\right|} = \frac{p(\underline{x},\underline{\theta})}{|\det \underline{A}|}$$

But $\underline{A}$ does not depend on $\underline{\alpha}$ and $\underline{\theta} = \underline{A}^{-1}\underline{\alpha}$ so that
$$p(\underline{x},\underline{\alpha}) = \frac{p_{x,\theta}(\underline{x}, \underline{A}^{-1}\underline{\alpha})}{|\det \underline{A}|}$$

The MAP estimator of $\underline{\alpha}$ maximizes $p_{x,\theta}(\underline{x}, \underline{A}^{-1}\underline{\alpha})$ or because $\underline{\theta} = \underline{A}^{-1}\underline{\alpha}$ 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

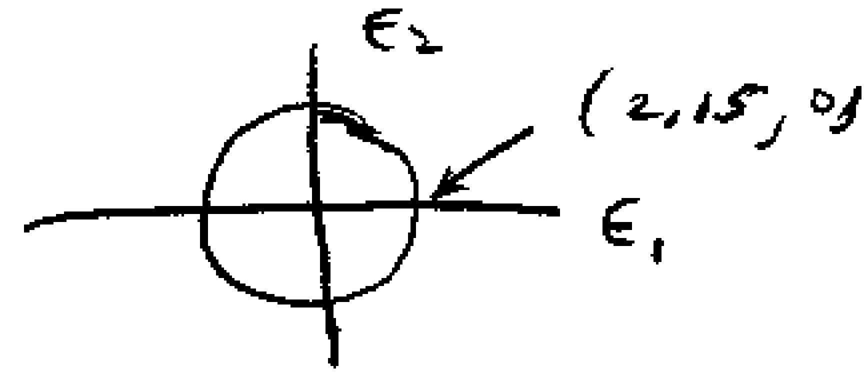
$$\underline{y}^T\underline{y} = y_1^2 + y_2^2 \sim \chi_2^2$$

since $\left.\begin{array}{l} y_1 \sim N(0,1) \\ y_2 \sim N(0,1) \end{array}\right\}$ independent

14) $$\underline{\epsilon}^T \underline{M}_{\hat{\theta}}^{-1} \underline{\epsilon} = 2 \ln \frac{1}{1-P}$$

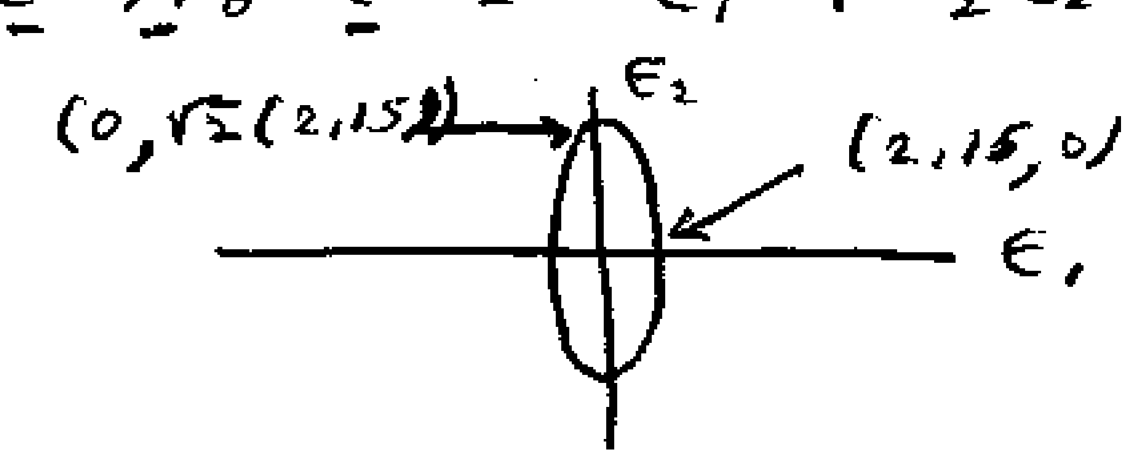
$$= (2,15)^2$$

a) $\underline{\epsilon}^T \underline{\epsilon} = (2,15)^2$



b) $\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}$

$$\underline{\epsilon}^T \underline{M}_{\hat{\theta}}^{-1} \underline{\epsilon} = \epsilon_1^2 + \tfrac{1}{2}\epsilon_2^2 = (2,15)^2$$



c) $\underline{M}_{\hat\theta}^{-1} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$

$$[\epsilon_1\ \epsilon_2]\ \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} = (2.15)^2$$

$$2\epsilon_1^2 + 2\epsilon_2^2 - 2\epsilon_1\epsilon_2 = 3(2.15)^2$$

$$\epsilon_1^2 - \epsilon_1\epsilon_2 + \epsilon_2^2 = \frac{3}{2}(2.15)^2$$

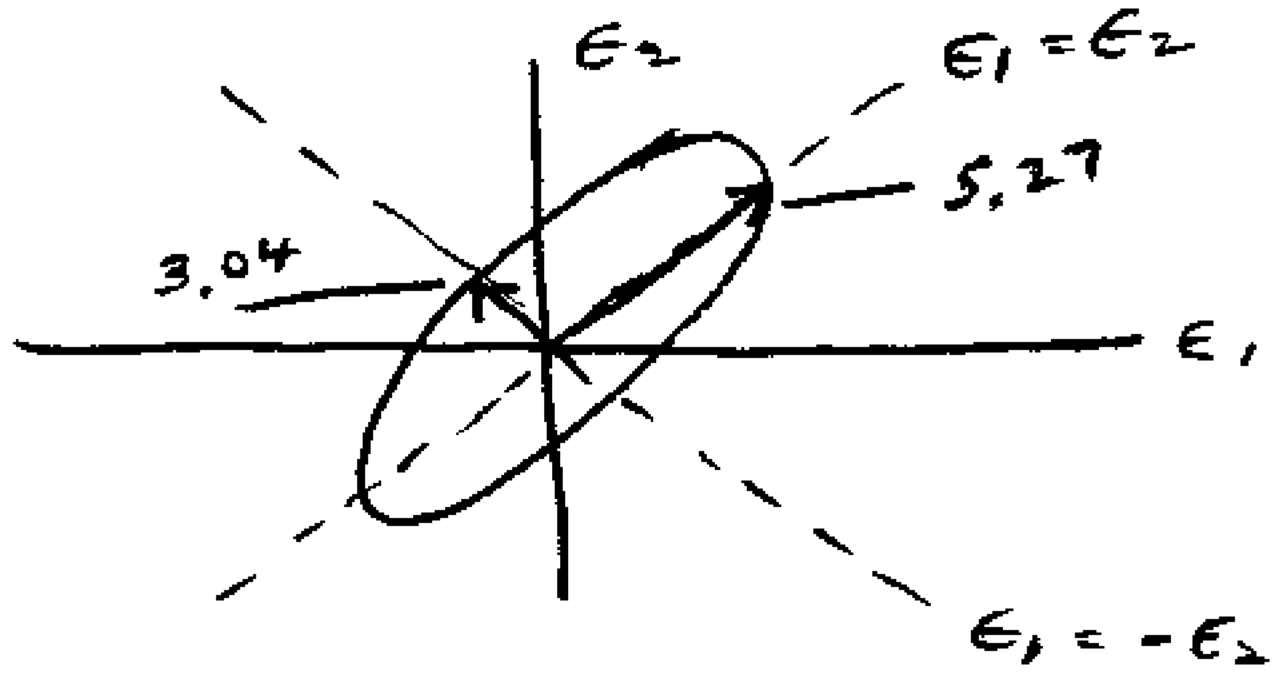
This is a rotated ellipse. It can be rewritten in standard form as

$$\frac{(\epsilon_1+\epsilon_2)^2}{4} + \frac{(\epsilon_1-\epsilon_2)^2}{4/3} = \frac{3}{2}(2.15)^2$$

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) $$\underline{M}_{\hat\theta} = (\underline{C}_\theta^{-1} + \underline{H}^T\underline{C}_W^{-1}\underline{H})^{-1} = (\underline{C}_\theta^{-1} + \underline{C}_W^{-1})^{-1}$$

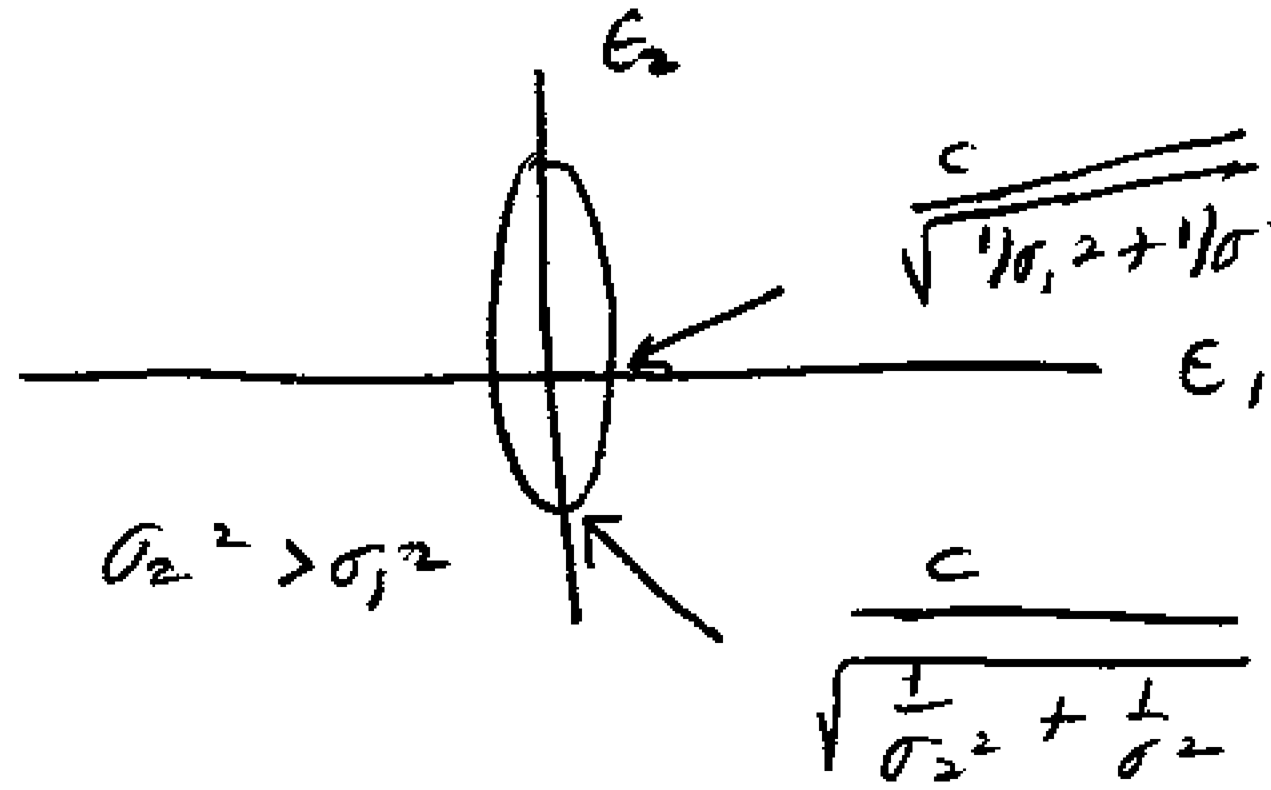
$$\underline{M}_{\hat\theta}^{-1} = \underline{C}_\theta^{-1} + \underline{C}_W^{-1} = \begin{bmatrix} 1/\sigma_1^2 + 1/\sigma^2 & 0 \\ 0 & 1/\sigma_2^2 + 1/\sigma^2 \end{bmatrix}$$

$$\underline{\epsilon}^T\underline{M}_{\hat\theta}^{-1}\underline{\epsilon} = 2\ln\frac{1}{1-P}$$

$$\epsilon_1^2\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma^2}\right) + \epsilon_2^2\left(\frac{1}{\sigma_2^2} + \frac{1}{\sigma^2}\right) = 2\ln\frac{1}{1-P}$$

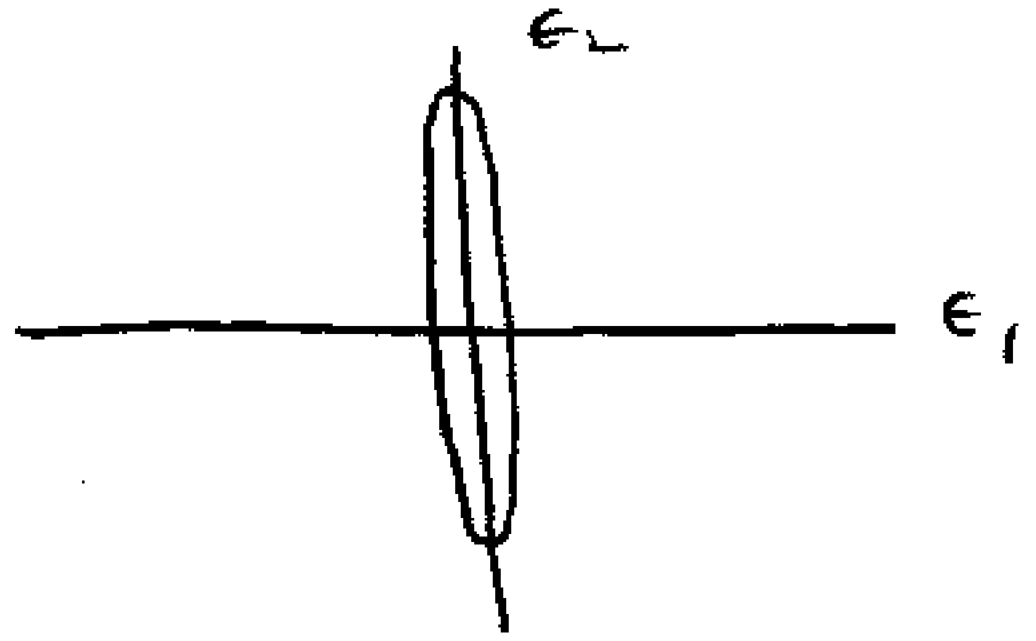
If $\sigma_1^2 = \sigma_2^2$, we have a circle.

If $\sigma_2^2 > \sigma_1^2$, we have an ellipse



$$c = \sqrt{2 \ln \frac{1}{1-P}}$$

For $\sigma_2^2 \gg \sigma_1^2$ 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 $N = 2M+1$

$$(\underline{C}_\theta^{-1} + \underline{H}^T \underline{C}_w^{-1} \underline{H})^{-1} = \begin{bmatrix} \frac{1}{\sigma_A^2} + \frac{N}{\sigma^2} & 0 \\ 0 & \frac{1}{\sigma_B^2} + \frac{\sum n^2}{\sigma^2} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \dfrac{1}{\frac{1}{\sigma_A^2}+\frac{N}{\sigma^2}} & 0 \\ 0 & \dfrac{1}{\frac{1}{\sigma_B^2}+\frac{\sum n^2}{\sigma^2}} \end{bmatrix}$$

$$\underline{H}^T \underline{C}_w^{-1}(\underline{x}-\underline{H}\underline{\mu}_0) = \frac{1}{\sigma^2}(\underline{H}^T\underline{x} - \underline{H}^T\underline{H}\underline{\mu}_0)$$

$$\underline{H}^T\underline{H} = \begin{bmatrix} N & 0 \\ 0 & \sum n^2 \end{bmatrix}$$

$$\underline{H}^T \underline{C}_w^{-1}(\underline{x}-\underline{H}\underline{\mu}_0) = \frac{1}{\sigma^2}\begin{bmatrix} \sum x(n) - NA_0 \\ \sum n x(n) - \sum n^2 B_0 \end{bmatrix}$$

$$\hat{A} = A_0 + \frac{\frac{1}{\sigma^2}\left[\sum x(n) - NA_0\right]}{\frac{1}{\sigma_A^2}+\frac{N}{\sigma^2}}$$

$$= A_0 + \frac{N/\sigma^2}{\frac{1}{\sigma_A^2}+\frac{N}{\sigma^2}}(\bar{x} - A_0)$$

$$\hat{B} = B_0 + \frac{\frac{1}{\sigma^2}\left(\sum n x(n) - \sum n^2 B_0\right)}{\frac{1}{\sigma_B^2}+\frac{\sum n^2}{\sigma^2}}$$

$$= B_0 + \frac{\sum_{n=-M}^{M} n^2/\sigma^2}{\frac{1}{\sigma_B^2}+\sum_{n=-M}^{M} n^2/\sigma^2}\left[\frac{\sum_{n=-M}^{M} n x(n)}{\sum_{n=-M}^{M} n^2} - B_0\right]$$

$$Bmse(\hat{A}) = \left[(\underline{C}_\theta^{-1} + \underline{H}^T \underline{C}_w^{-1}\underline{H})^{-1}\right]_{11}$$

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

$$Bmse(\hat{B}) = \left[\left(\underline{C}_\theta^{-1} + \underline{H}^T \underline{C}_W^{-1} \underline{H}\right)^{-1}\right]_{22}$$

$$= \left(\frac{1}{\sigma_B^2} + \sum_{n=-M}^{M} n^2 / \sigma^2\right)^{-1}$$

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

$$E(\underline{\epsilon}\underline{\epsilon}^T) = \underline{H}\, E\left[(\underline{\theta} - \hat{\underline{\theta}})(\underline{\theta} - \hat{\underline{\theta}})^T\right] \underline{H}^T$$

$$= \underline{H}\, \underline{M}_{\hat{\theta}}\, \underline{H}^T$$

$$= \begin{bmatrix} 1 & -M \\ \vdots & \vdots \\ 1 & M \end{bmatrix} \begin{bmatrix} \frac{1}{1/\sigma_A^2 + N/\sigma^2} & 0 \\ 0 & \frac{1}{\frac{1}{\sigma_B^2} + \frac{\sum n^2}{\sigma^2}} \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 \\ -M & \cdots & M \end{bmatrix}$$

$$= \frac{1}{\frac{1}{\sigma_A^2} + \frac{N}{\sigma^2}} \underline{1}\underline{1}^T + \frac{1}{\frac{1}{\sigma_B^2} + \frac{\sum n^2}{\sigma^2}} \underline{m}\underline{m}^T$$

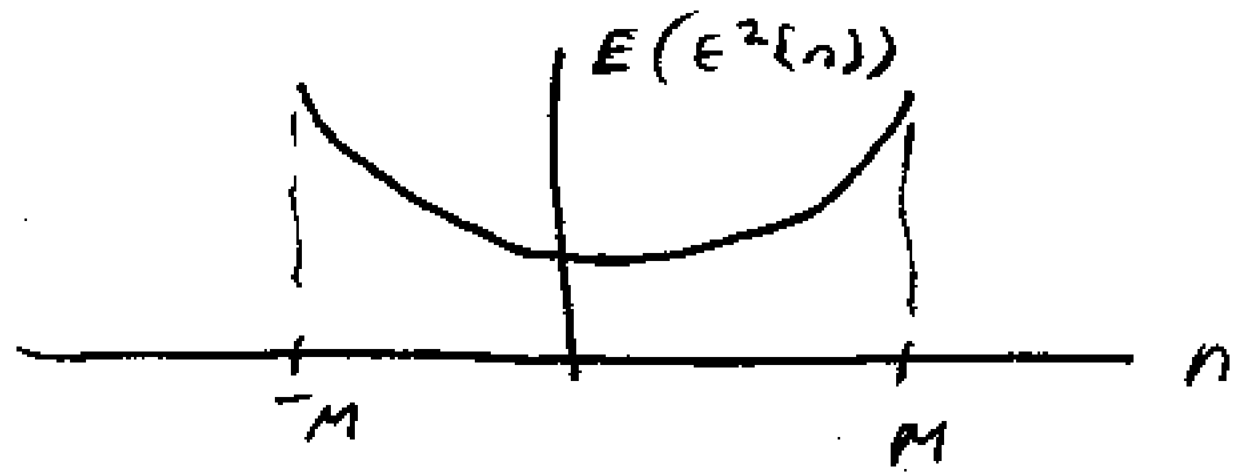
where $\underline{m} = [-M \ldots M]^T$

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)

$$\hat{\underline{s}} = \underline{C}_s(\underline{C}_s + \sigma^2 \underline{I})^{-1}\underline{x}$$

$$= [(\underline{C}_s + \sigma^2\underline{I})\underline{C}_s^{-1}]^{-1}\underline{x}$$

$$= (\underline{I} + \sigma^2\underline{C}_s^{-1})^{-1}\underline{x}$$

$$(\underline{I} + \sigma^2\underline{C}_s^{-1})\hat{\underline{s}} = \underline{x}$$

$$(\underline{C}_s + \sigma^2\underline{I})\hat{\underline{s}} = \underline{C}_s\underline{x}$$