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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $\underline{C}$,

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

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

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

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

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

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

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

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

Must minimize $\dfrac{1}{\sum_i \alpha_i^2/\lambda_i}$ subject to constraint $\sum_i \alpha_i^2 = \mathcal{E}_0$. Equivalently we must maximize $\sum_i \alpha_i^2/\lambda_i$. Using Lagrangian multipliers

$$F = \sum_i \alpha_i^2/\lambda_i + \lambda\left(\sum_i \alpha_i^2 - \mathcal{E}_0\right)$$

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

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

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 $\mathcal{E}_0 = \alpha_j^2$, $\operatorname{var}(\hat{A}) = \lambda_j / \mathcal{E}_0$ 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.

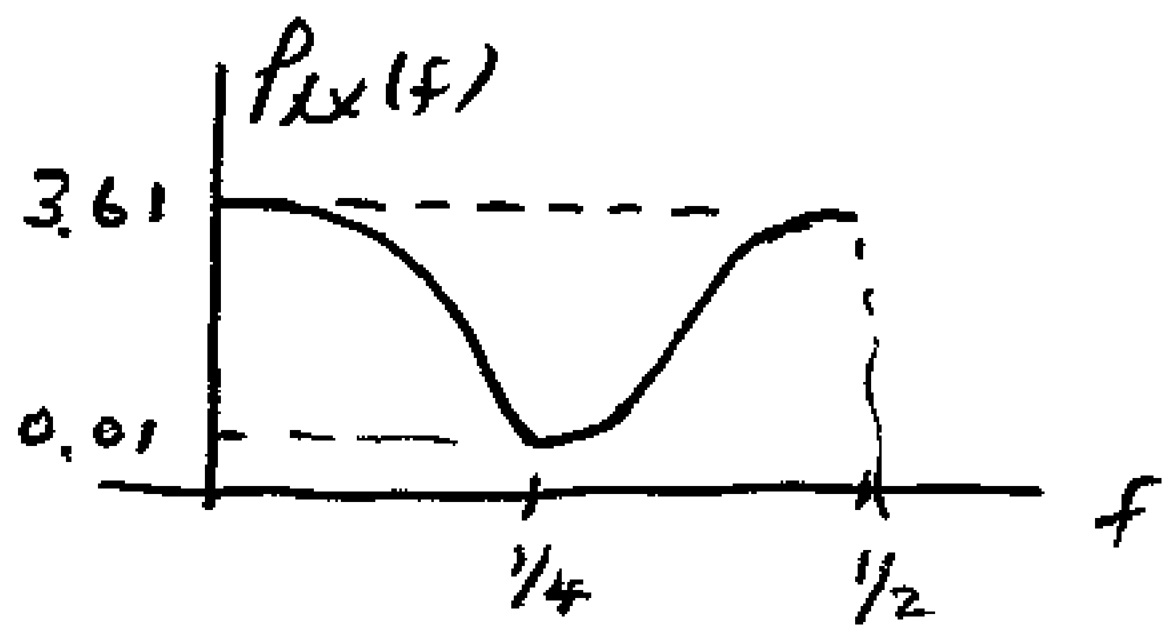
$$\mathrm{var}(\hat{A}) = \frac{1}{\underline{s}^T \underline{C}^{-1} \underline{s}} = \frac{\sigma^2}{\sum_{n=0}^{N-1} s^2[n]}$$

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

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) $$P_{xx}(f) = r_{xx}[0] + r_{xx}[2]e^{-j4\pi f} + r_{xx}[2]e^{j4\pi f}$$
$$= 1.81 + 2(0.9)\cos 4\pi f$$
$$= 1.81 + 1.8\cos 4\pi f$$

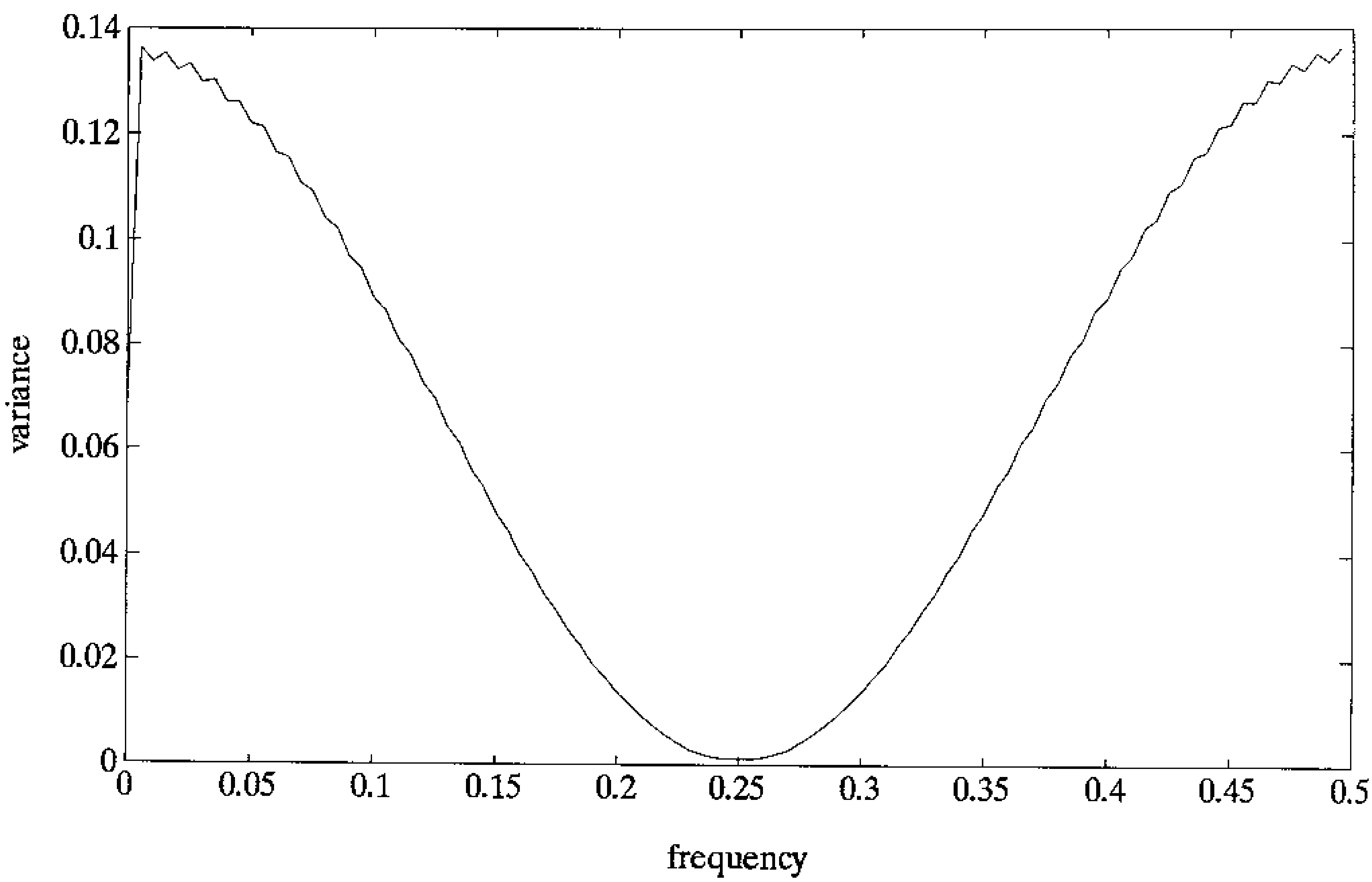


Need to minimize $\frac{1}{\underline{s}^T \underline{C}^{-1} \underline{s}} = \mathrm{var}(\hat{A})$
where $\underline{s} = [1 \ \cos 2\pi f_1 \ \ldots \ \cos 2\pi f_1 (N-1)]^T$
and

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

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

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

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

since $\sum_k h[k] = 1$

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

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

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

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

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

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

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

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

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, $E(w[n]) = 0$.

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

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

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

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

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

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

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

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

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

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 $\hat{\underline{C}}$

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

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

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

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

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

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

Clearly, by the minimum variance property of the BLUE

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

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

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

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

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

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

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

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

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

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

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

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.

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

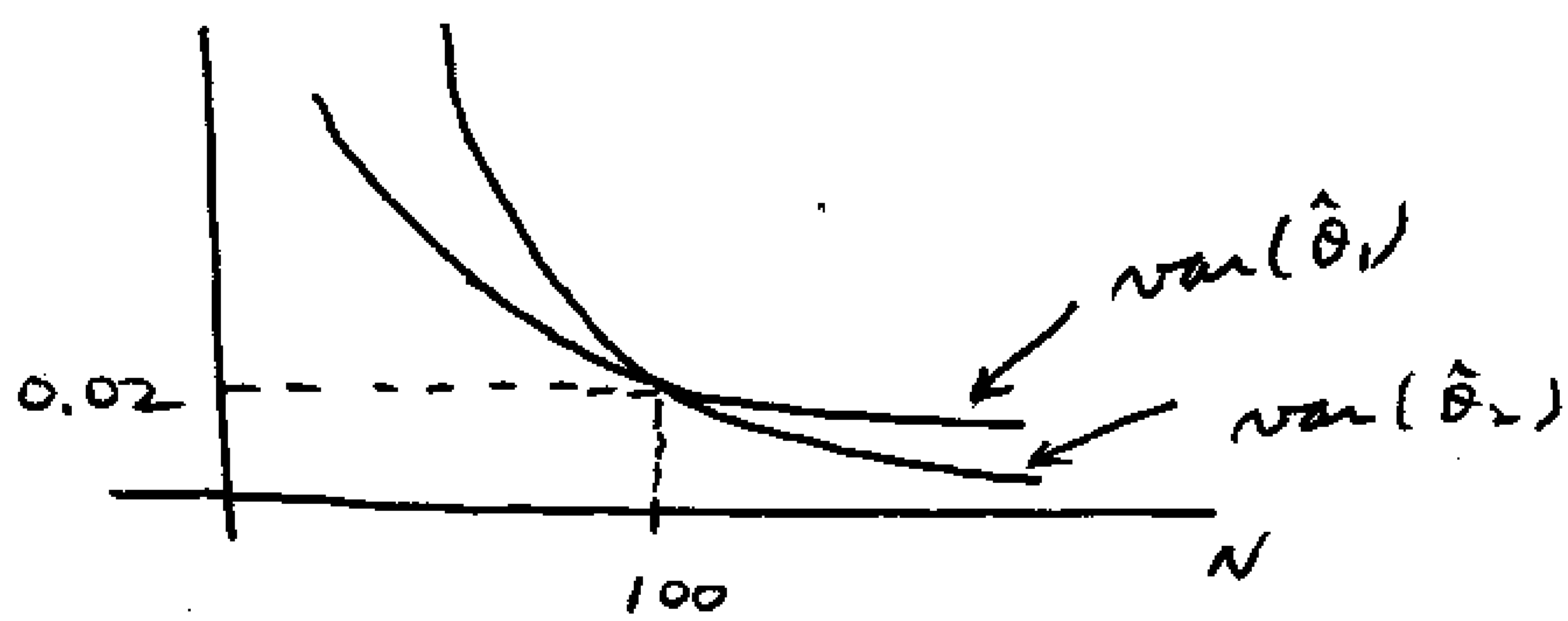
$$\frac{d \ln p}{d\lambda} = N/\lambda - N\bar{x}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Minimize $\sum_n (x[n] - A s[n])^2$ 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, $\hat{A} \sim \mathcal{N}(A, I^{-1}(A))$ 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 $C_{ij} = \sum_n x_i[n]\, x_j(n)$

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

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

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

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

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

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

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 $\hat{\rho}=\frac{1}{N}c_{12}$

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

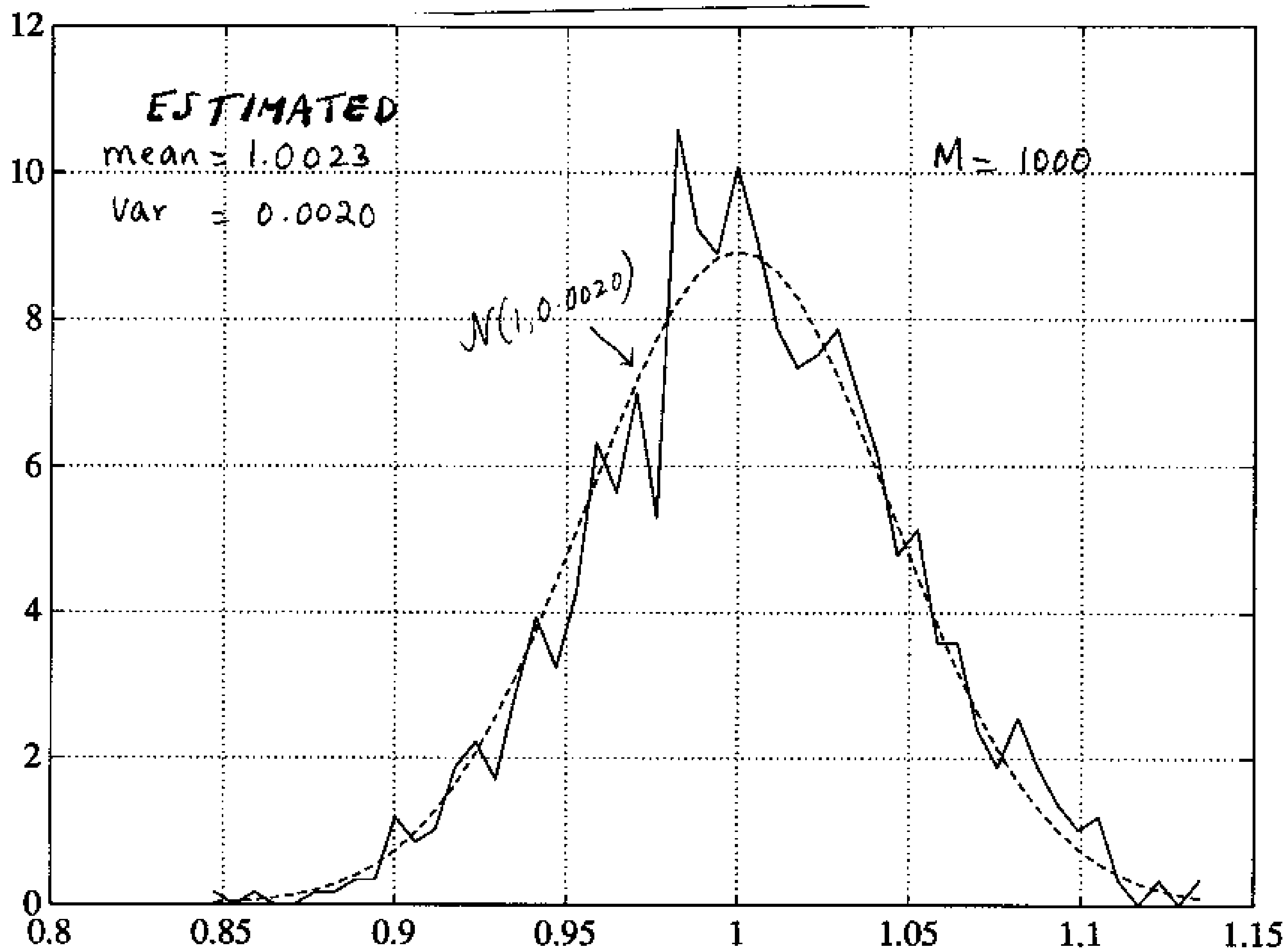
12) To find the MLE we maximize p or equivalently $\ln p$

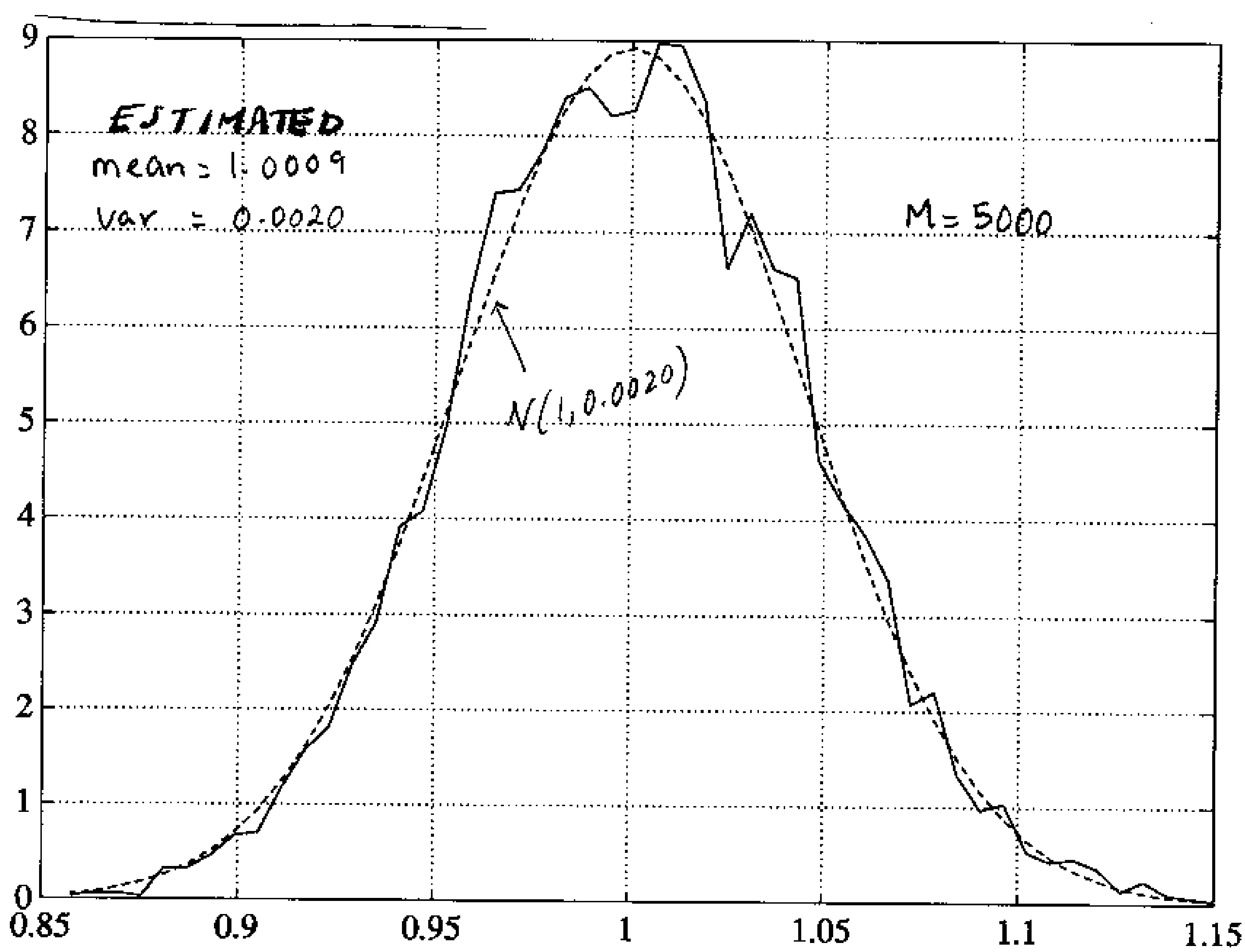
$$\Rightarrow \frac{\partial \ln p}{\partial\theta}=0$$

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:

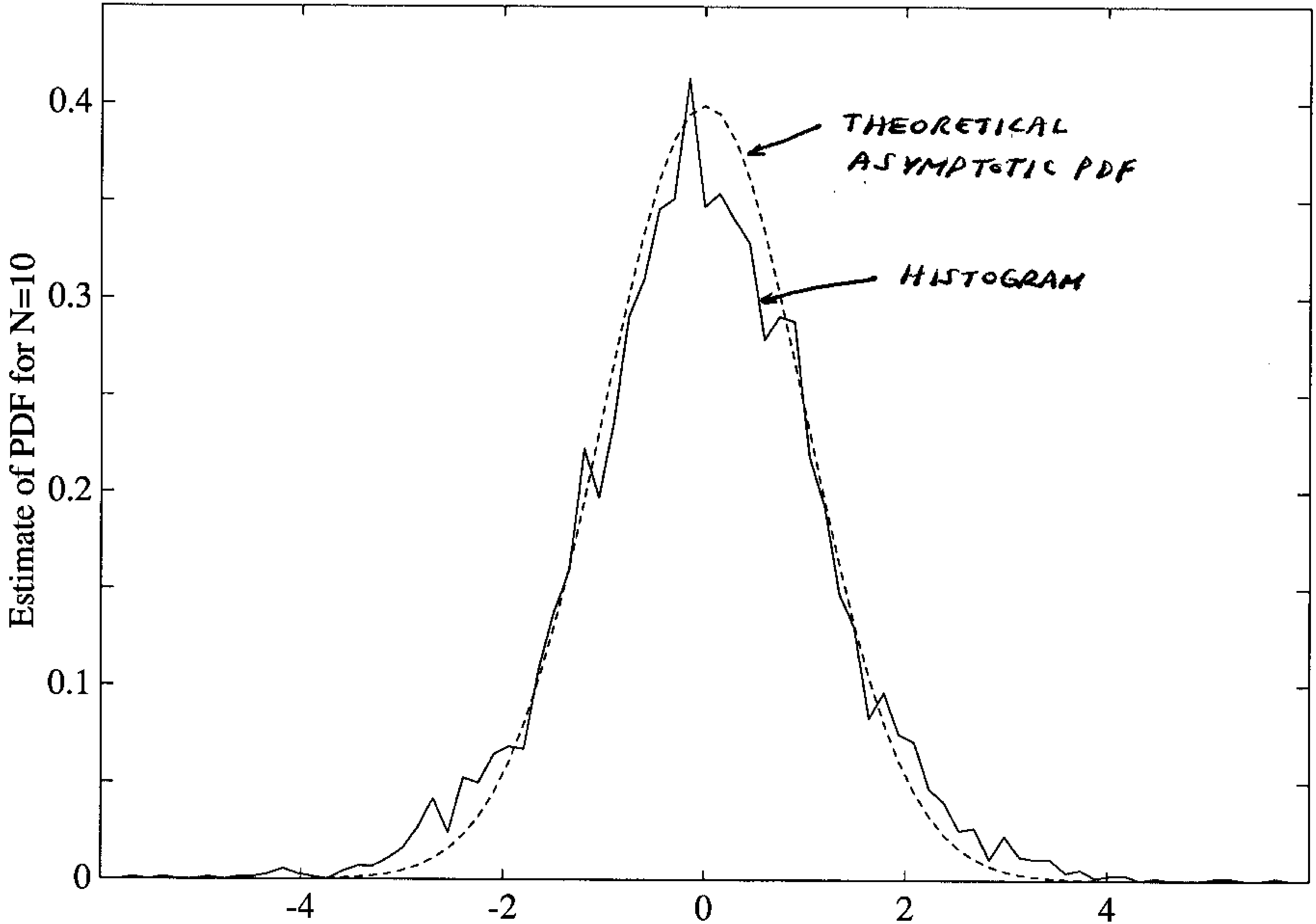
$$\text{var}(\epsilon_s) = E(\epsilon_s^2) = \frac{4}{N^2A^2} \sum_m \sum_n E(w(m)w(n)) \cdot \sin 2\pi f_0 m \sin 2\pi f_0 n$$

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

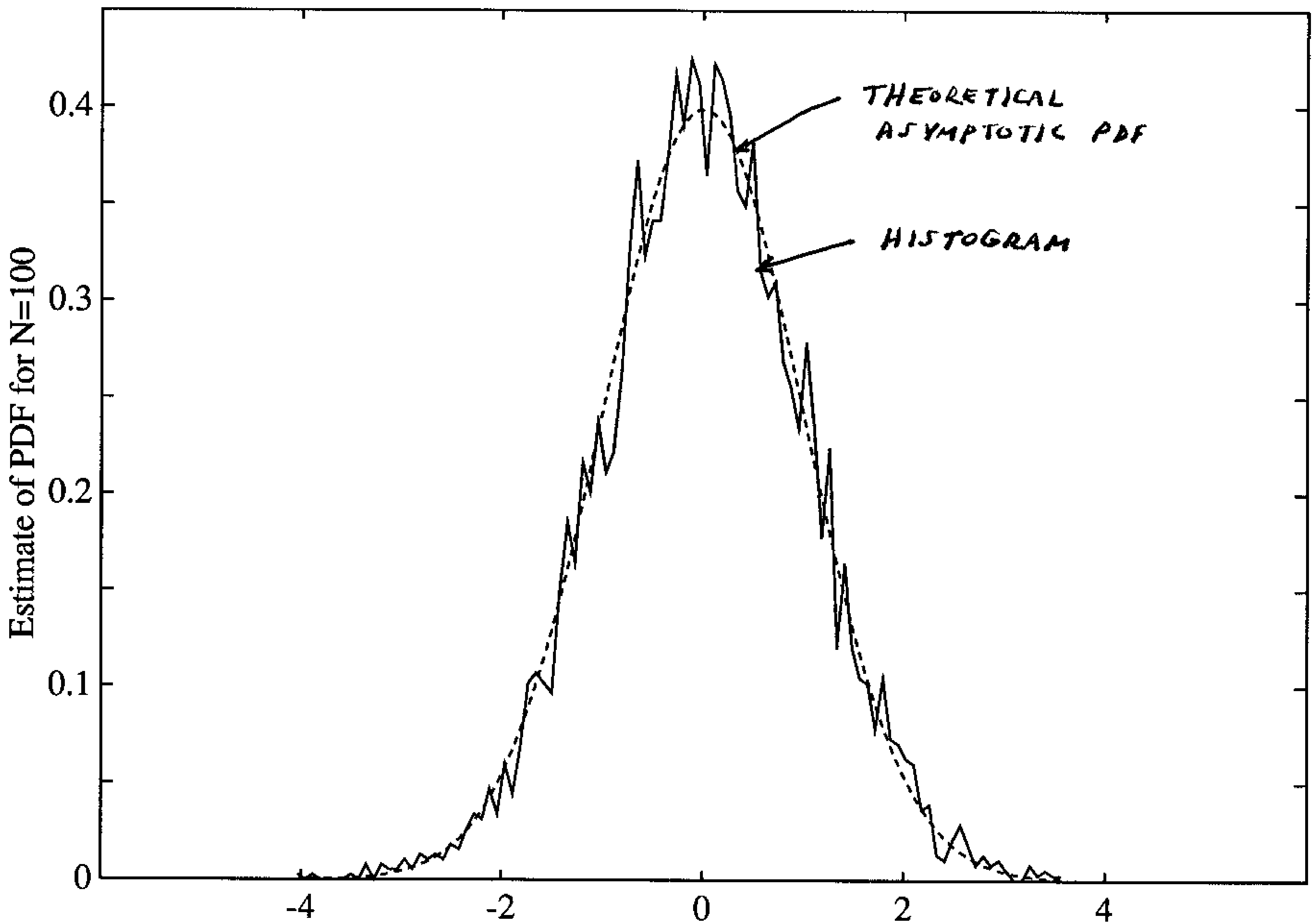
$$= \frac{4\sigma^2}{N^2A^2} \sum_n \left(\tfrac{1}{2} - \tfrac{1}{2}\cos 4\pi f_0 n\right)$$

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

Problem 7.14: Verification of Slutsky"s theorem



Problem 7.14: Verification of Slutsky"s theorem



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

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

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

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

$$= \cos\phi$$

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

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

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

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

The variance is just the CRLB.

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

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

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

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

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

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

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

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

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

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

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

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

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

Initial freq. (f_{0_0})	$E(\hat{f}_0)$	$\mathrm{var}(\hat{f}_0)$
0.24	0.25	1.2×10^{-6}
0.26	0.25	1.2×10^{-6}
0.28	0.16	10

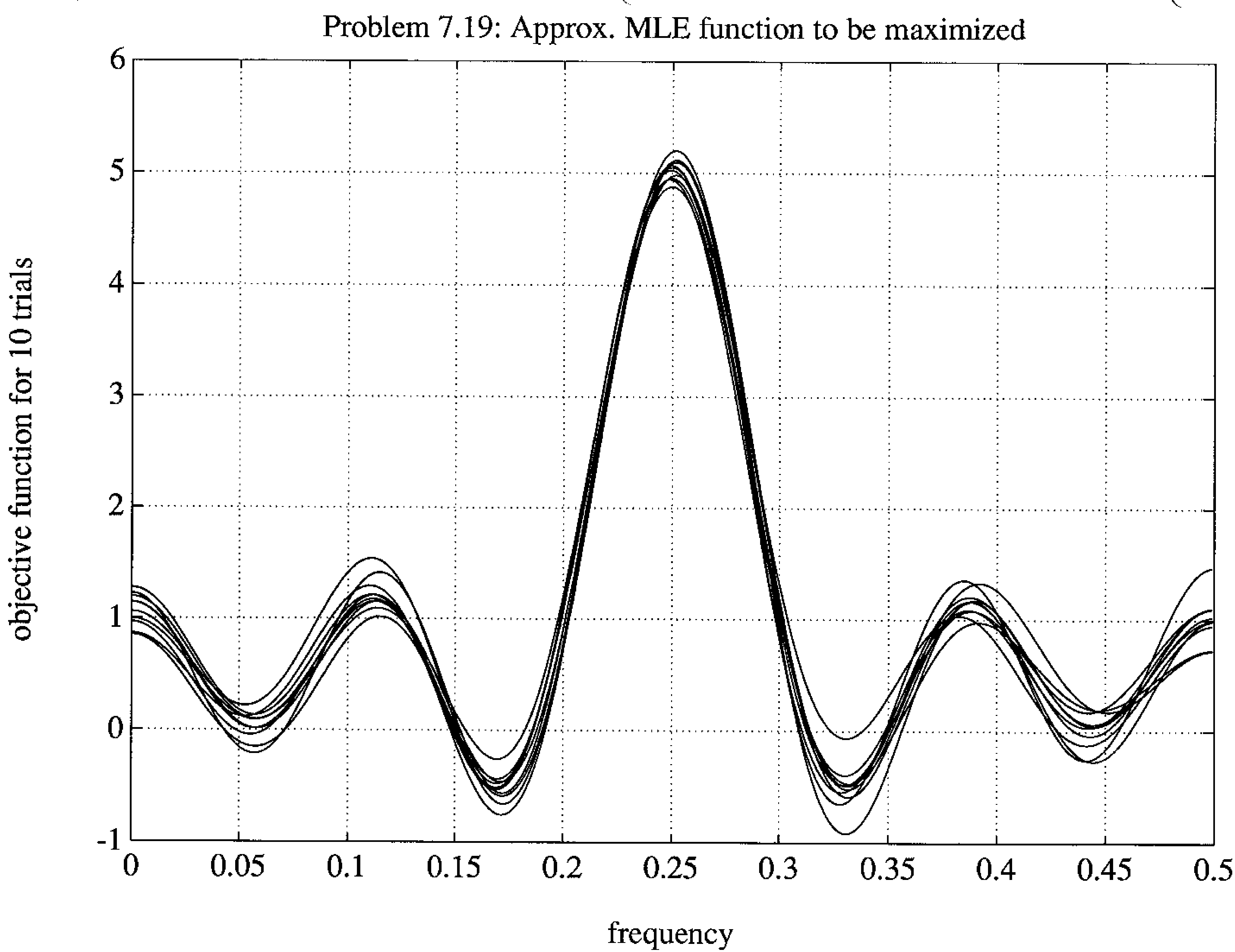
20) $$p(\underline{x};\underline{s}) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-s[n])^2}$$

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

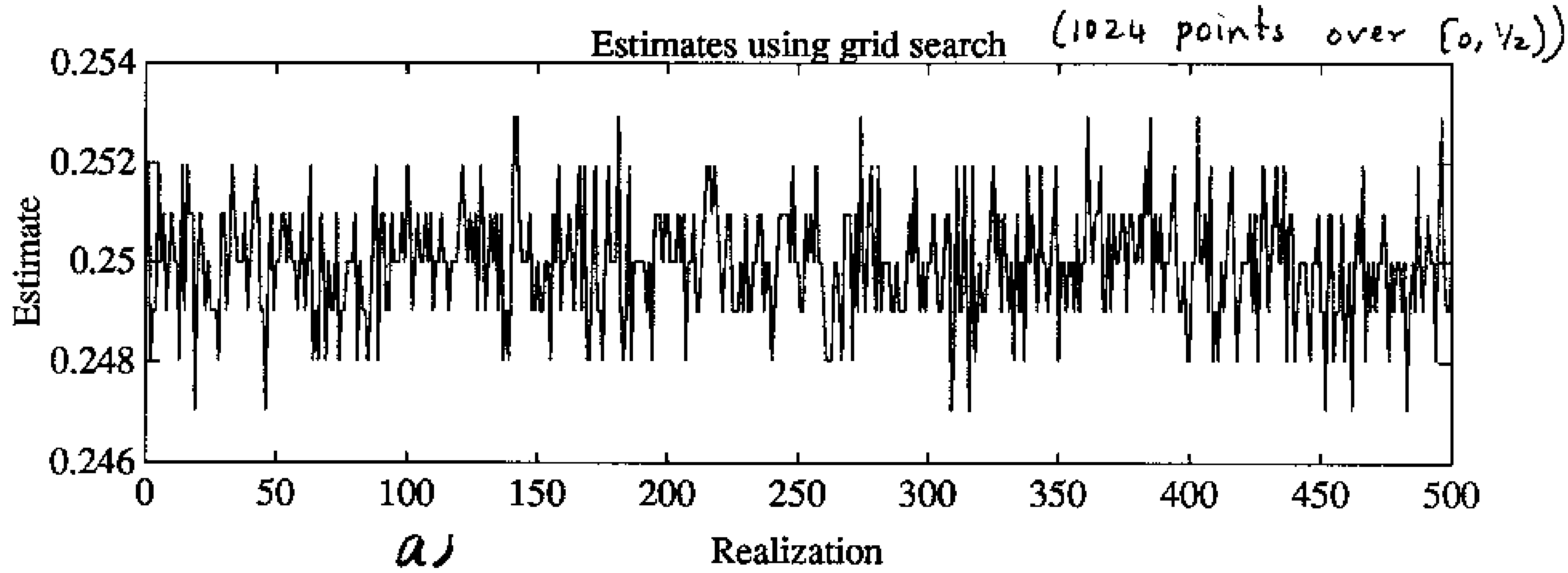
But $\hat{\underline{s}} = \underline{x} \sim N(\underline{s}, \sigma^2 \underline{I})$ and thus $\hat{\underline{s}}$ is unbiased and Gaussian. To determine if it is efficient we find the CRLB.

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

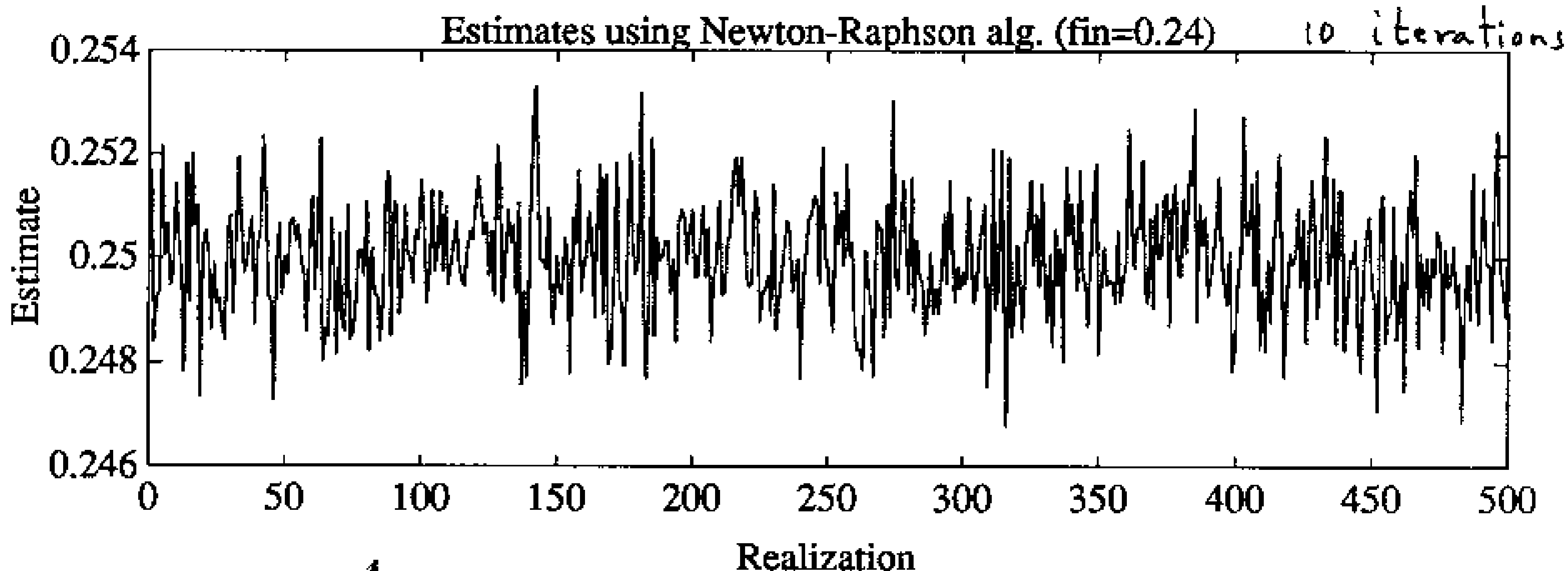
$$\frac{\partial^2 \ln p}{\partial s[k]\partial s[l]} = -\frac{1}{\sigma^2}\delta_{kl}$$


Problem 7.19: Approx. MLE function to be maximized
objective function for 10 trials
frequency
6
5
4
3
2
1
0
-1
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5

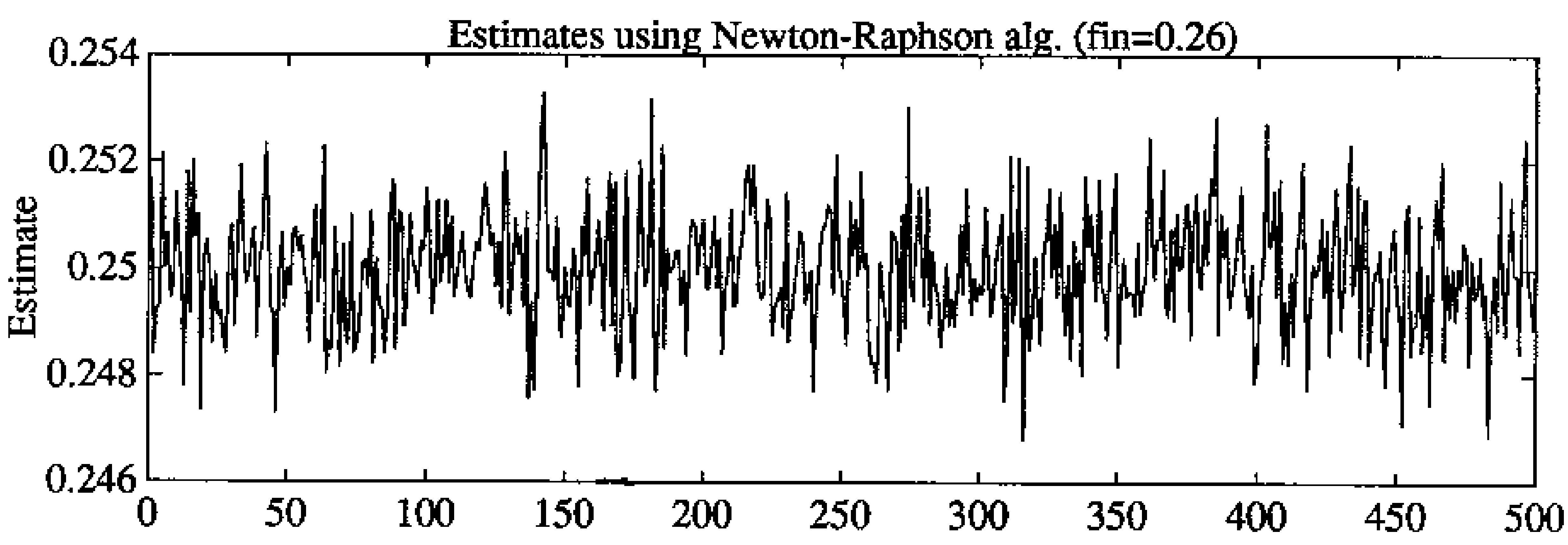
93


Estimates using grid search
(1024 points over [0, 1/2))
Estimate
Realization
0.254
0.252
0.25
0.248
0.246
0 50 100 150 200 250 300 350 400 450 500

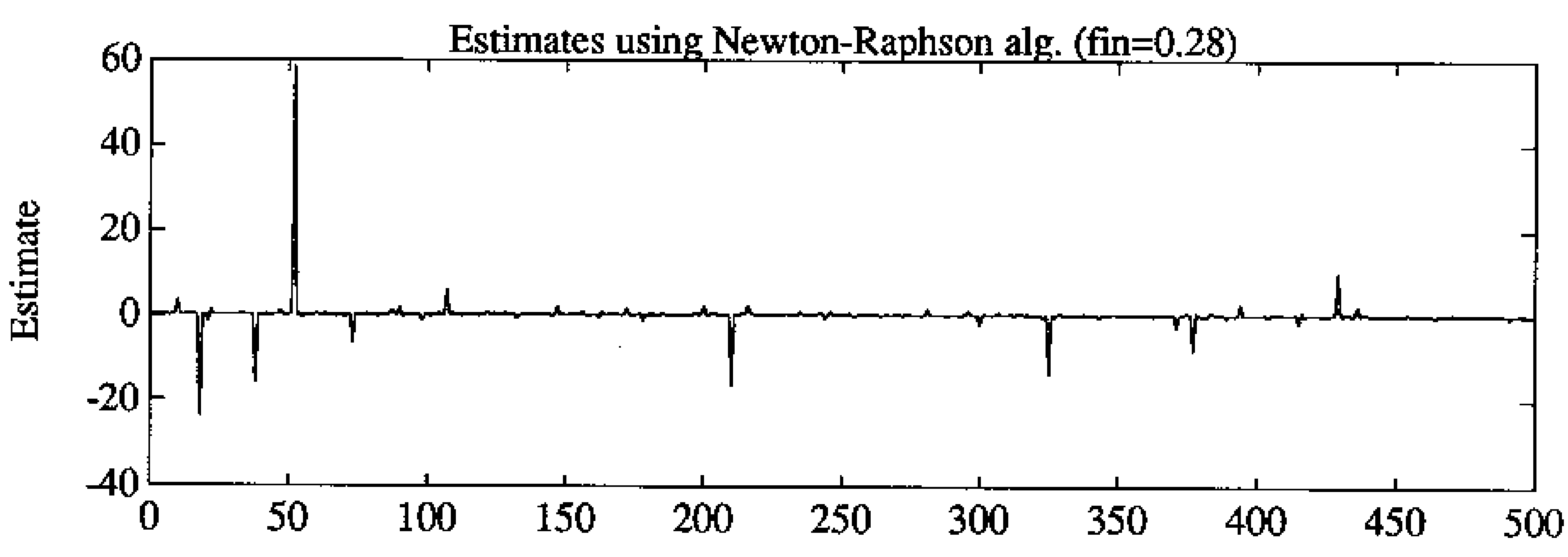
a)


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

b)


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

c)


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

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

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

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

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

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

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

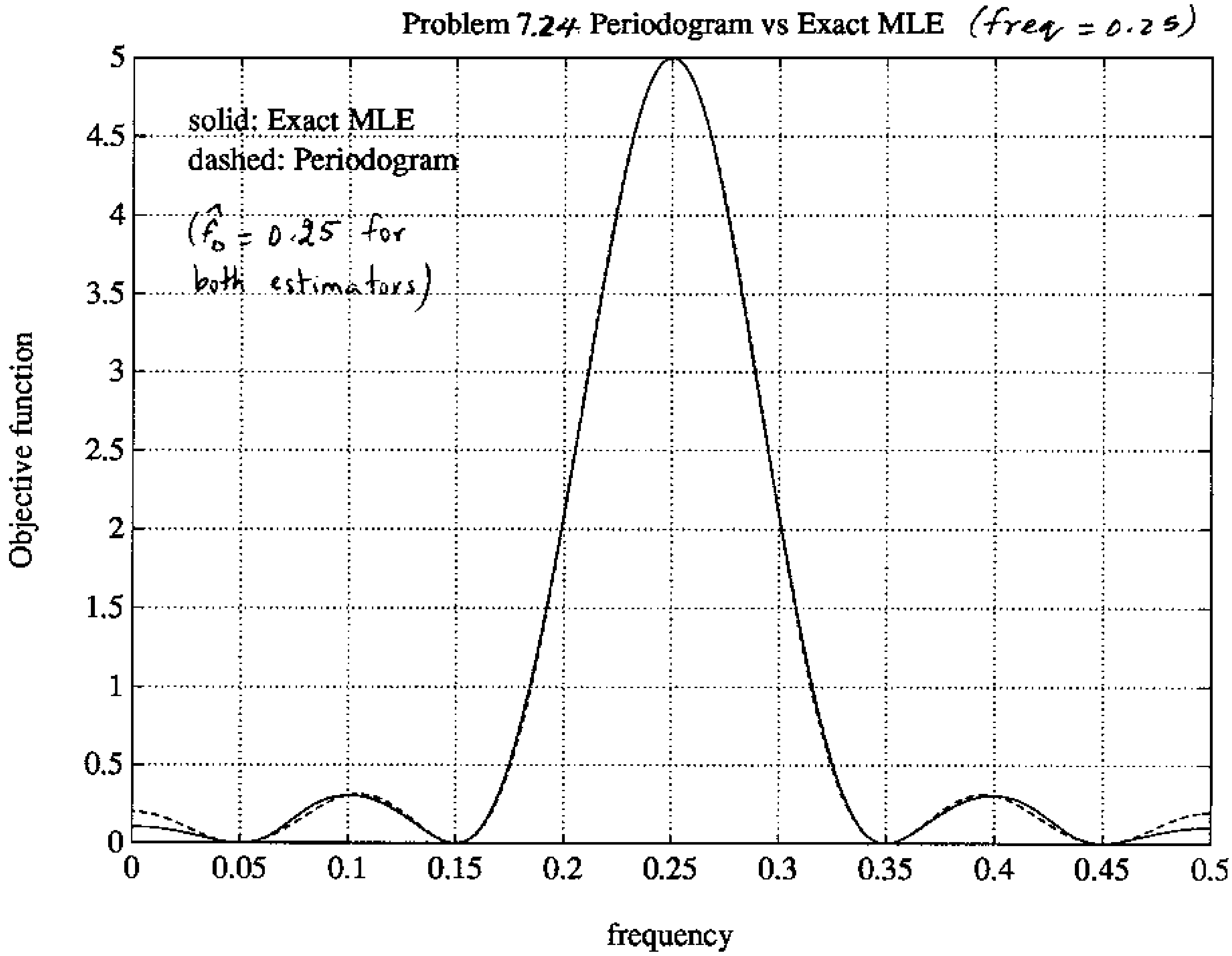
$$\Rightarrow \mathcal{F}^{-1}\{ I(f) \} = \frac{1}{N} x'[n] \star x'[-n]$$
$$= \frac{1}{N} \sum_{n=-\infty}^{\infty} x'[-n]\, x'[k-n]$$
$$= \frac{1}{N} \sum_{n=-\infty}^{\infty} x'[n]\, x'[n+k]$$

For $k \geq 0$

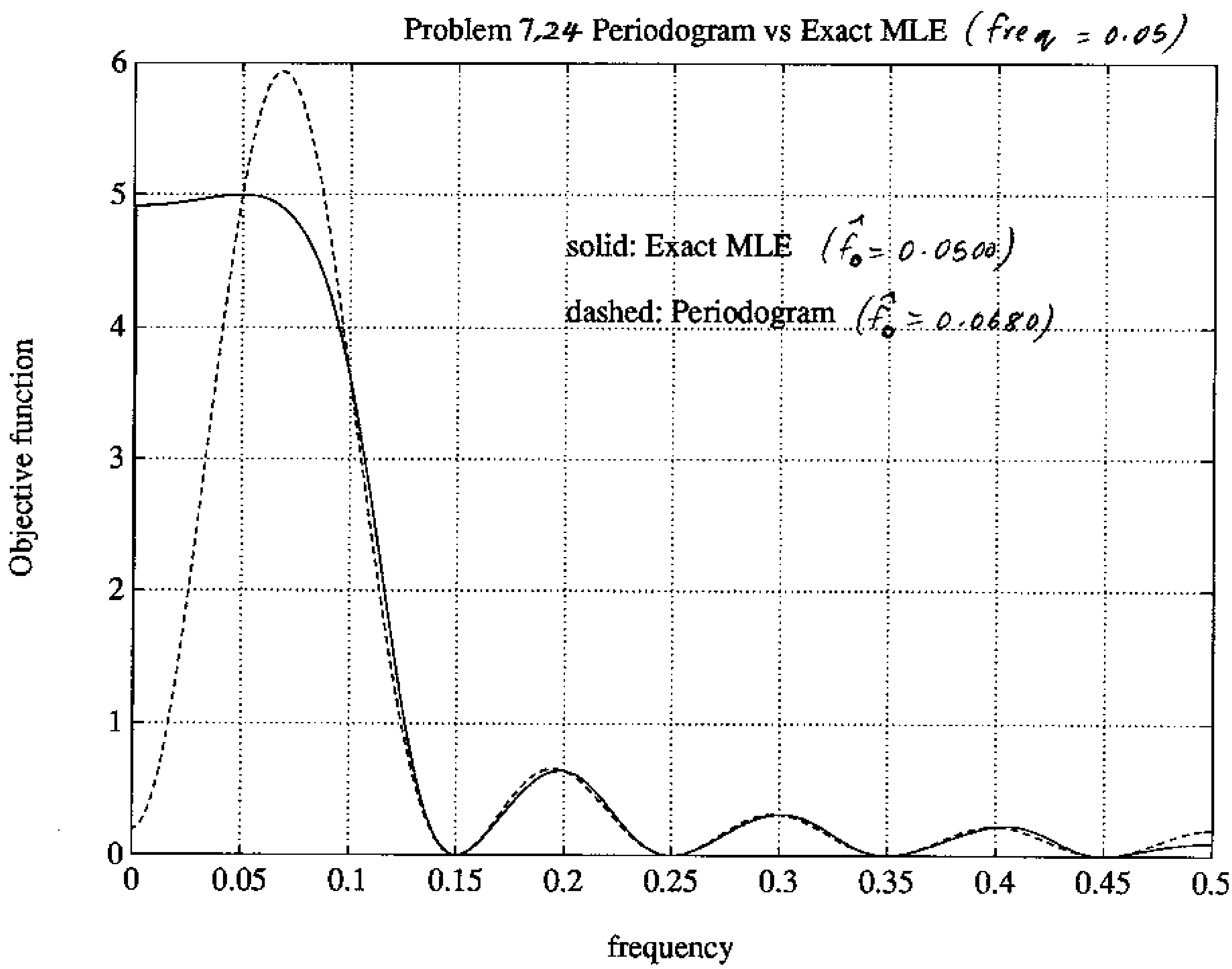
$$\mathcal{F}^{-1}\{ I(f) \} = \frac{1}{N} \sum_{n=0}^{N-1-k} x[n]\, x[n+k]$$

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



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.

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

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

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

$$C_{\hat{\alpha}} - \frac{\partial g}{\partial \underline{\theta}} I^{-1}(\underline{\theta}) \frac{\partial g}{\partial \underline{\theta}}^T \geq 0$$

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

and $g(a[1], \sigma_u^2) = \dfrac{\sigma_u^2}{|1+a[1]e^{-j2\pi f_0}|^2}$

so that

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

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

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

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

$$= \frac{\sigma_u^2}{N r_{xx}[0]} \; \frac{\sigma_u^4}{|A(f)|^8} \; 4\mathrm{Re}^2\left(A(f)\, e^{j2\pi f_0}\right)$$

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

$$= \frac{4 P_{xx}^4(f_0)}{N r_{xx}[0]\, \sigma_u^2} \; \mathrm{Re}^2\left(A(f)\, e^{j2\pi f_0}\right)$$

$$+ \frac{2}{N} P_{xx}^2(f_0)$$

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

$$E(\hat{A}) = 1/N\left[MA - (N-M)A \right] = A$$

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

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

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

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

$$\underline{H}^T\underline{H} = \frac{N}{2}\underline{I}$$

$$\Rightarrow \quad \hat{\underline{\theta}} = \frac{2}{N}\underline{H}^T\underline{x} \Rightarrow \hat{A}_i = \frac{2}{N}\sum_{n=0}^{N-1} x(n)\cos 2\pi f_i n$$

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

$$= \underline{x}^T(\underline{I} - \tfrac{2}{N}\underline{H}\underline{H}^T)\underline{x}$$

$$= \underline{x}^T\underline{x} - 2/N\,\|\underline{H}^T\underline{x}\|^2$$

$$= \underline{x}^T\underline{x} - \frac{2}{N}\left(\frac{N}{2}\right)^2\|\hat{\underline{\theta}}\|^2$$

$$= \sum_{n=0}^{N-1} x^2(n) - \frac{N}{2}\sum_{i=1}^{p}\hat{A}_i^2$$

For $w(n)$ WGN the PDF is

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

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

$$C_{\hat{\underline{\theta}}} = E[(\hat{\underline{\theta}}-\underline{\theta})(\hat{\underline{\theta}}-\underline{\theta})^T]$$

$$= E[(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underbrace{(\underline{x}-\underline{H}\underline{\theta})}_{\underline{w}}\underbrace{(\underline{x}-\underline{H}\underline{\theta})^T}_{\underline{w}^T}\underline{H}(\underline{H}^T\underline{H})^{-1}]$$

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

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

$$\hat{\underline{\theta}} \sim \mathcal{N}(\underline{\theta},\ 2\sigma^2/N\ \underline{I})$$

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

From Prob 8.5 we have $\hat{\underline{\theta}} \sim \mathcal{N}(\underline{\theta}, \sigma^2(\underline{H}^T\underline{H})^{-1})$

Yes, it is unbiased.

7)
$$E(\hat{\sigma}^2) = \frac{1}{N} E\left[\underline{x}^T(\underline{I} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T)\underline{x}\right)$$
$$= \frac{1}{N} \operatorname{tr}\left[(\underline{I} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T) E(\underline{x}\,\underline{x}^T)\right]$$

since $E(\underline{x}^T\underline{y}) = \operatorname{tr}(E(\underline{y}\,\underline{x}^T))$

$$E(\hat{\sigma}^2) = \frac{1}{N} \operatorname{tr}\Big[(\underline{I} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T) E((H\theta + w)(H\theta + w)^T)\Big]$$

$$= \frac{1}{N} \operatorname{tr}\left[(\underline{I} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T)(\underline{H}\underline{\theta}\,\underline{\theta}^T\underline{H}^T + \sigma^2\underline{I})\right]$$

$$= \frac{1}{N} \operatorname{tr}\Big[\underline{H}\underline{\theta}\,\underline{\theta}^T\underline{H}^T + \sigma^2\underline{I} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{H}\underline{\theta}\,\underline{\theta}^T\underline{H}^T - \sigma^2\underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T\Big]$$

$$= \frac{\sigma^2}{N} \, tr \left[\underline{I} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T \right]$$

$$= \sigma^2 - \frac{\sigma^2}{N} \, tr \left[\underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T \right]$$

$$= \sigma^2 - \sigma^2/N \; tr \left[\underbrace{(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{H}}_{\underline{I} \; (p \times p)} \right]$$

$$= \sigma^2 - \sigma^2 p/N = \sigma^2 \frac{N-p}{N}$$

$\hat{\sigma}^2$ 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)
$$J = (\underline{x} - \underline{H}\underline{\theta})^T \underline{W} (\underline{x} - \underline{H}\underline{\theta})$$

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

$$= (\underbrace{\underline{D}\underline{x}}_{\underline{x}'} - \underbrace{\underline{D}\underline{H}}_{\underline{H}'}\underline{\theta})^T (\underline{D}\underline{x} - \underline{D}\underline{H}\underline{\theta})$$

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

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

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

$$J_{MIN} = \underline{x}'^T (\underline{I} - \underline{H}' (\underline{H}'^T \underline{H}')^{-1} \underline{H}'^T) \underline{x}'$$

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

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

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) $\underline{x}_1 = \underline{z}_1 + \underline{z}_1^\perp \qquad \underline{x}_2 = \underline{z}_2 + \underline{z}_2^\perp$

$$\ell = \underline{x}_1^T \underline{P}\underline{x}_2 - \underline{x}_2^T \underline{P}\underline{x}_1 = (\underline{z}_1 + \underline{z}_1^\perp)^T \underline{P}(\underline{z}_2 + \underline{z}_2^\perp) - (\underline{z}_2 + \underline{z}_2^\perp)^T \underline{P}(\underline{z}_1 + \underline{z}_1^\perp)$$

$$= (\underline{z}_1 + \underline{z}_1^\perp)^T \underline{P}\underline{z}_2 - (\underline{z}_2 + \underline{z}_2^\perp)^T \underline{P}\underline{z}_1$$

since $\underline{P}\underline{z}^\perp = \underline{0}$

$$\ell = (\underline{z}_1 + \underline{z}_1^\perp)^T \underline{z}_2 - (\underline{z}_2 + \underline{z}_2^\perp)^T \underline{z}_1$$

since $\underline{P}\underline{z} = \underline{z}$

$$\ell = \underline{z}_1^T \underline{z}_2 - \underline{z}_2^T \underline{z}_1 = 0 \qquad \text{since } \underline{z}_i^{\perp T} \underline{z}_j = 0$$

Now let $\underline{x}_1 = \underline{e}_i = [0\,0 \ldots 0\,1\,0 \ldots 0]^T$

$\underline{x}_2 = \underline{e}_j \qquad \uparrow i^{\text{th}}$ 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 $= p$