In-Class Exercise 10

1. Consider a single output signal y with known inputs x_1 and x_2, which has the form of

$$y = a_1 x_1 + a_2 x_2 + b + w$$

where a_1, a_2 and b are unknown constants while $w \sim \mathcal{N}(0, \sigma^2)$.

(a) Write down the probability density function (PDF) of y.

(b) Given y, x_1 and x_2, is it possible to find a_1, a_2 and b? Explain your answer.

2. Without expanding $(\boldsymbol{y}-\boldsymbol{A}\tilde{\boldsymbol{x}})^T\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{A}\tilde{\boldsymbol{x}})$, apply chain rule to compute the differentiation of $(\boldsymbol{y}-\boldsymbol{A}\tilde{\boldsymbol{x}})^T\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{A}\tilde{\boldsymbol{x}})$ with respect to $\tilde{\boldsymbol{x}}$ to get the least squares solution $\hat{\boldsymbol{x}}=(\boldsymbol{A}^T\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{W}\boldsymbol{y}$.

Note that for $\boldsymbol{x}\in\mathbb{R}^n$, $\boldsymbol{a}\in\mathbb{R}^n$, $\boldsymbol{y}\in\mathbb{R}^m$ and $\boldsymbol{A}\in\mathbb{R}^{m\times n}$:

$$\frac{d\boldsymbol{x}^T\boldsymbol{a}}{d\boldsymbol{x}}=\frac{d\boldsymbol{a}^T\boldsymbol{x}}{d\boldsymbol{x}}=\boldsymbol{a}$$

$$\frac{d\boldsymbol{A}\boldsymbol{x}}{d\boldsymbol{x}}=\boldsymbol{A}^T,\quad \frac{d\boldsymbol{x}^T\boldsymbol{A}}{d\boldsymbol{x}}=\boldsymbol{A}$$

$$\frac{d\boldsymbol{y}^T\boldsymbol{A}\boldsymbol{x}}{d\boldsymbol{x}}=\boldsymbol{A}^T\boldsymbol{y},\quad \frac{d\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{y}}{d\boldsymbol{x}}=\boldsymbol{A}\boldsymbol{y}$$

$$\frac{d\boldsymbol{x}^T\boldsymbol{x}}{d\boldsymbol{x}}=2\boldsymbol{x},\quad \frac{d\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{d\boldsymbol{x}}=(\boldsymbol{A}+\boldsymbol{A}^T)\boldsymbol{x}=2\boldsymbol{A}\boldsymbol{x}\ \text{if}\ \boldsymbol{A}=\boldsymbol{A}^T$$

3. Given N samples of y_n which has the form of:

$$y_n = \cos(\omega n) + e_n, \quad n = 1, \cdots, N$$

where $\omega \in (0, \pi)$ in the unknown frequency of a sinusoid, and $\{e_1, \cdots, e_N\}$ are independent with $e_n \sim \mathcal{N}(0, \sigma^2)$.

(a) Is $\{e_1, \cdots, e_N\}$ a white noise sequence?

(b) Is the model linear or nonlinear?

(c) Let $\boldsymbol{y} = [y_1 \ \cdots \ y_N]^T$. Write down the joint probability density function (PDF) of $\boldsymbol{y}$ parameterized by ω.

(d) Construct the least squares (LS) cost function $J(\tilde{\omega})$ to estimate ω. Use identity matrix as the weighting matrix.

(e) Can the LS estimate $\hat{\omega}_{\mathrm{LS}}$ be easily determined? Explain your answer.

(f) Is $\hat{\omega}_{\mathrm{LS}}$ equal to the maximum likelihood estimate $\hat{\omega}_{\mathrm{ML}}$? Explain your answer.

4. Given N samples of y_n which has the form of:

$$y_n = e^{A+n} + w_n, \quad n = 1, \cdots, N$$

where A is an unknown constant, and $\{w_1, \cdots, w_N\}$ are independent with $w_n \sim \mathcal{N}(0, \sigma^2)$.

(a) Construct the least squares (LS) cost function $J(\tilde{A})$ to estimate A. Use the weighting matrix corresponding to the maximum likelihood estimate.

(b) Determine the LS estimate $\hat{A}_{\mathrm{LS}}$.

5. Given $N \geq 3$ sets of outputs $\{y_1, \cdots, y_N\}$ and inputs $\{x_{1,1}, \cdots, x_{N,1}\}$ and $\{x_{1,2}, \cdots, x_{N,2}\}$, modelled as:

$$\begin{aligned}
y_1 &= x_{1,1}a_1 + x_{1,2}a_2 + b + w_1 \\
y_2 &= x_{2,1}a_1 + x_{2,2}a_2 + b + w_2 \\
&\vdots \qquad \vdots \qquad \vdots \qquad \vdots \\
y_N &= x_{N,1}a_1 + x_{N,2}a_2 + b + w_N
\end{aligned}$$

where a_1, a_2 and b are unknown constants while $\{w_1, \cdots, w_N\}$ are independent with $w_n \sim \mathcal{N}(0, \sigma^2)$.

By assigning suitable vector and matrix, determine the maximum likelihood estimate of $\boldsymbol{\alpha} = [a_1\ a_2\ b]^T$, denoted as $\hat{\boldsymbol{\alpha}} = [\hat{a}_1\ \hat{a}_2\ \hat{b}]^T$, in vector-matrix form.

6. Consider a measurement vector $\boldsymbol{r} = [r_1\ r_2\ r_3]^T$, and its elements are expressed as:

$$r_1 = A + w_1, \quad r_2 = A + w_2, \quad r_3 = A + w_3$$

where A is the constant to be estimated, while $w_1 \sim \mathcal{N}(0, \sigma_1^2)$, $w_2 \sim \mathcal{N}(0, \sigma_2^2)$ and $w_3 \sim \mathcal{N}(0, \sigma_3^2)$ are noise components and they are independent of each other.

(a) Write down the probability density function (PDF) of r_1.
(b) Compute the covariance matrix of $\boldsymbol{r}$.
(c) Write down the joint PDF of $\boldsymbol{r}$.
(d) Find the maximum likelihood (ML) estimate of A, $\hat{A}$.
(e) Determine the mean and variance of $\hat{A}$.
(f) Suppose $\boldsymbol{r} = [5.1\ 6.2\ 7.3]^T$ while $\sigma_1^2 = 0.1$, $\sigma_2^2 = 1$ and $\sigma_3^2 = 5$. Compute $\hat{A}$ and the variance of $\hat{A}$.

7. Given N measurements of r_n:

$$r_n = \alpha \sin(\omega n + \phi) + w_n, \quad n = 1, \cdots, N$$

where α is the unknown constant to be estimated, while ω and ϕ are known constants, and $w_n \sim \mathcal{N}(0, \sigma^2)$, $n = 1, \cdots, N$, are independent.

(a) Write down the joint probability density function (PDF) of $\boldsymbol{r} = [r_1 \ \cdots \ r_N]^T$.

(b) Find the maximum likelihood (ML) estimate of α, $\hat{\alpha}$.

(c) Determine the mean and variance of $\hat{\alpha}$.

Solution

1.(a)

The PDF of y is:

$$p(y)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}\left(y-a_1x_1-a_2x_2-b\right)^2}$$

1.(b)

It is not possible because there is one equation but with 3 unknowns a_1, a_2 and b. Even without noise, it is clear that there are infinite sets of $\{a_1,a_2,b\}$ satisfying $y=a_1x_1+a_2x_2+b$ when y, x_1 and x_2 are given.

2.

$$\frac{d(\boldsymbol{y}-\boldsymbol{A}\tilde{\boldsymbol{x}})^T\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{A}\tilde{\boldsymbol{x}})}{d\tilde{\boldsymbol{x}}} = \frac{d(\boldsymbol{y}-\boldsymbol{A}\tilde{\boldsymbol{x}})}{d\tilde{\boldsymbol{x}}} \cdot \frac{d(\boldsymbol{y}-\boldsymbol{A}\tilde{\boldsymbol{x}})^T\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{A}\tilde{\boldsymbol{x}})}{d(\boldsymbol{y}-\boldsymbol{A}\tilde{\boldsymbol{x}})}$$
$$= -\boldsymbol{A}^T \cdot 2\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{A}\tilde{\boldsymbol{x}})$$

Setting the expression to zero with $\tilde{\boldsymbol{x}} = \hat{\boldsymbol{x}}$, we obtain:

$$\boldsymbol{A}^T \cdot 2\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{A}\hat{\boldsymbol{x}}) = \boldsymbol{0} \Rightarrow \boldsymbol{A}^T\boldsymbol{W}\boldsymbol{y} = \boldsymbol{A}^T\boldsymbol{W}\boldsymbol{A}\hat{\boldsymbol{x}} \Rightarrow \hat{\boldsymbol{x}} = \left(\boldsymbol{A}^T\boldsymbol{W}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{W}\boldsymbol{y}$$

3.(a)

Yes, $\{e_1, \cdots, e_N\}$ is a white sequence as the random variables are independent with zero mean and identical power or variance of σ^2.

3.(b)

As ω is nonlinear in the measurements, it is a nonlinear model.

3.(c)

Grouping all noise components as $\boldsymbol{e} = [e_1 \ \cdots \ e_N]^T$, we know that (see Example 3.20) $\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_N)$. Similar to (3.38), we can write:

$$p(\boldsymbol{y};\omega) = \frac{1}{(2\pi)^{N/2}\sigma^N} e^{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{\mu})^T(\boldsymbol{y}-\boldsymbol{\mu})} = \frac{1}{(2\pi)^{N/2}\sigma^N} e^{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(y_n - \cos(\omega n))^2}$$

where $\boldsymbol{\mu} = [\cos(\omega) \ \cdots \ \cos(\omega N)]^T$.

3.(d)

Setting the weighting matrix as $\boldsymbol{I}_N$, the LS cost function is:

$$J(\tilde{\omega}) = (\boldsymbol{y} - \tilde{\boldsymbol{\mu}})^T \boldsymbol{W} (\boldsymbol{y} - \tilde{\boldsymbol{\mu}}) = (\boldsymbol{y} - \tilde{\boldsymbol{\mu}})^T (\boldsymbol{y} - \tilde{\boldsymbol{\mu}}) = \sum_{n=1}^{N} (y_n - \cos(\tilde{\omega} n))^2$$

where $\tilde{\boldsymbol{\mu}} = [\cos(\tilde{\omega}) \ \cdots \ \cos(\tilde{\omega} N)]^T$.

3.(e)

To find the solution, we can apply differentiation:

$$\left. \frac{dJ(\tilde{\omega})}{d\tilde{\omega}} \right|_{\tilde{\omega}=\hat{\omega}} = 0 \Rightarrow \sum_{n=1}^{N} n \sin(\hat{\omega} n)(y_n - \cos(\hat{\omega} n)) = 0$$

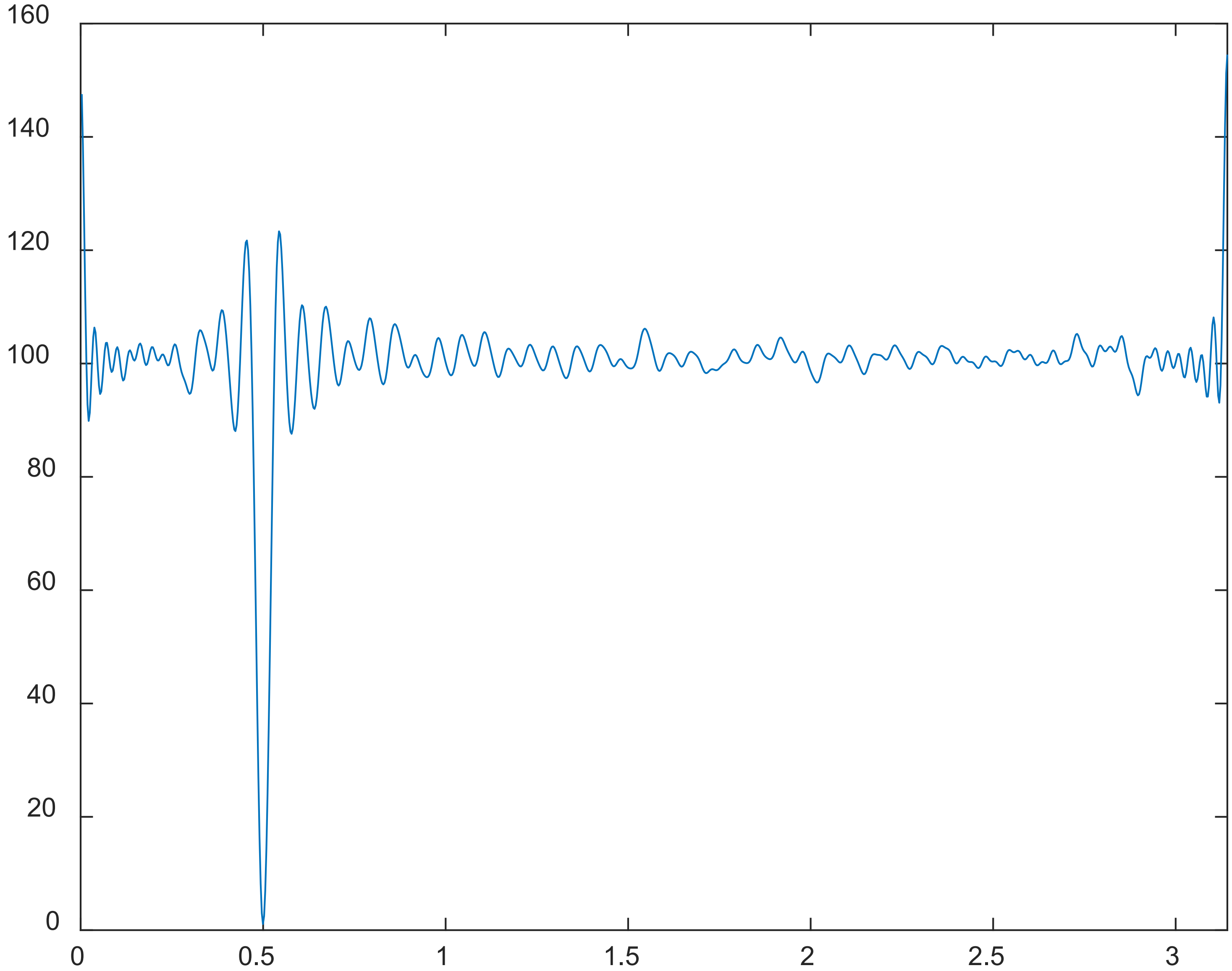
However, it is not easy to find $\hat{\omega}$ because the equation is nonlinear.

An illustration with $\omega = 0.5$, $N = 100$ and noise power 0.01:

```
w=0.5;
N=100;
n=1:N;
y=cos(w.*n)+0.1*randn(1,N);
for i=1:1000;
f(i)=sum((y-cos(i.*pi./1000.*n)).^2);
end
plot((1:1000)./1000.*pi,f)
axis([0 pi, 0 160])
```

Our task is to find the global minimum. However, because of the nonlinearity of the cost function, there are other local minima. We can also see that there are many solutions satisfying

$$\sum_{n=1}^{N} n \sin(\hat{\omega} n)(y_n - \cos(\hat{\omega} n)) = 0$$



160
140
120
100
80
60
40
20
0
0
0.5
1
1.5
2
2.5
3

3.(f)

Yes. The $\hat{\omega}_{\mathrm{LS}}$ equal to $\hat{\omega}_{\mathrm{ML}}$. It is because maximizing

$$\frac{1}{(2\pi)^{N/2}\sigma^N}e^{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(y_n-\cos(\tilde{\omega}n))^2}$$

is equal to maximizing

$$-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(y_n-\cos(\tilde{\omega}n))^2 \text{ or } -\sum_{n=1}^{N}(y_n-\cos(\tilde{\omega}n))^2$$

or minimizing

$$\sum_{n=1}^{N}(y_n-\cos(\tilde{\omega}n))^2$$

4.(a)

For IID and Gaussian distributed $\{w_1, \cdots, w_N\}$, we know that the covariance matrix corresponds to an identity matrix. Hence the weighting matrix corresponding to the maximum likelihood estimate is simply the identity matrix.

As a result, the LS cost function is:

$$J(\tilde{A}) = \sum_{n=1}^{N} \left(y_n - e^{\tilde{A}+n} \right)^2$$

4.(b)

$$\left.\frac{d\sum_{n=1}^{N}\left(y_n - e^{\tilde{A}+n}\right)^2}{d\tilde{A}}\right|_{\tilde{A}=\hat{A}_{\mathrm{LS}}} = \left.\sum_{n=1}^{N} 2\left(y_n - e^{\tilde{A}+n}\right)\left(-e^{\tilde{A}+n}\right)\right|_{\tilde{A}=\hat{A}_{\mathrm{LS}}} = 0$$

$$\Rightarrow \sum_{n=1}^{N} y_n e^{\hat{A}_{\mathrm{LS}}+n} = \sum_{n=1}^{N} e^{2\hat{A}_{\mathrm{LS}}+2n}$$

$$\Rightarrow e^{\hat{A}_{\mathrm{LS}}} \sum_{n=1}^{N} y_n e^{n} = e^{2\hat{A}_{\mathrm{LS}}} \sum_{n=1}^{N} e^{2n}$$

$$\Rightarrow e^{\hat{A}_{\mathrm{LS}}} = \frac{\sum_{n=1}^{N} y_n e^{n}}{\sum_{n=1}^{N} e^{2n}} \Rightarrow \hat{A}_{\mathrm{LS}} = \ln\left(\frac{\sum_{n=1}^{N} y_n e^{n}}{\sum_{n=1}^{N} e^{2n}}\right)$$

5.

As discussed in Question 4, for IID and Gaussian distributed $\{w_1, \cdots, w_N\}$, the weighting matrix corresponding to the maximum likelihood estimate is simply the identity matrix.

We assign:

$$\boldsymbol{y} = [y_1 \ \cdots \ y_N]^T \ \text{ and } \ \boldsymbol{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & 1 \\ \vdots & \vdots & \vdots \\ x_{N,1} & x_{N,2} & 1 \end{bmatrix}$$

Apply (6.7) yields:

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y}$$

6.(a)
The PDF of r_1 is:

$$p(r_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(r_1 - A)^2}$$

6.(b)
The covariance matrix is:

$$\boldsymbol{C} = \mathbb{E}\{(\boldsymbol{r} - \mathbf{1}_3 A)(\boldsymbol{r} - \mathbf{1}_3 A)^T\} = \mathbb{E}\{\boldsymbol{w}\boldsymbol{w}^T\} = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}$$

where

$$\boldsymbol{w} = [w_1 \ w_2 \ w_3]^T$$

6.(c)

The joint PDF of $\boldsymbol{r}$ has the form of:

$$p(\boldsymbol{r}) = \frac{1}{(2\pi)^{3/2}|\boldsymbol{C}|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{r}-\boldsymbol{1}_3 A)^T \boldsymbol{C}^{-1}(\boldsymbol{r}-\boldsymbol{1}_3 A)}$$

Since:

$$|\boldsymbol{C}| = \sigma_1^2\sigma_2^2\sigma_3^2 \quad \text{and} \quad \boldsymbol{C}^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & 0 \\ 0 & \frac{1}{\sigma_2^2} & 0 \\ 0 & 0 & \frac{1}{\sigma_3^2} \end{bmatrix}$$

Hence, we have:

$$p(\boldsymbol{r}) = \frac{1}{(2\pi)^{3/2}\sigma_1\sigma_2\sigma_3} e^{-\frac{1}{2}\left[\frac{(r_1-A)^2}{\sigma_1^2}+\frac{(r_2-A)^2}{\sigma_2^2}+\frac{(r_3-A)^2}{\sigma_3^2}\right]}$$

6.(d)

Maximizing the likelihood function means minimizing:

$$\frac{(r_1-\tilde{A})^2}{\sigma_1^2}+\frac{(r_2-\tilde{A})^2}{\sigma_2^2}+\frac{(r_3-\tilde{A})^2}{\sigma_3^2} \quad \text{or} \quad \left(\boldsymbol{r}-\mathbf{1}_3\tilde{A}\right)^T \boldsymbol{C}^{-1}\left(\boldsymbol{r}-\mathbf{1}_3\tilde{A}\right)$$

From (6.10), the solution is:

$$\hat{A}=\frac{\mathbf{1}_3^T\boldsymbol{C}^{-1}\boldsymbol{r}}{\mathbf{1}_3^T\boldsymbol{C}^{-1}\mathbf{1}_3}=\frac{\frac{r_1}{\sigma_1^2}+\frac{r_2}{\sigma_2^2}+\frac{r_3}{\sigma_3^2}}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}+\frac{1}{\sigma_3^2}}$$

6.(e)

According to (6.11) and (6.12), we have:

$$\mathbb{E}\{\hat{A}\} = A$$

$$\text{var}\left(\hat{A}\right)=(\mathbf{1}_3^T\boldsymbol{C}^{-1}\mathbf{1}_3)^{-1}=\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}+\frac{1}{\sigma_3^2}\right)^{-1}$$

6.(f)
Based on the results in 6.(d) and 6.(e), we have

$$\hat{A} = \frac{\mathbf{1}_3^T \boldsymbol{C}^{-1} \boldsymbol{r}}{\mathbf{1}_3^T \boldsymbol{C}^{-1} \mathbf{1}_3} = \frac{\dfrac{r_1}{\sigma_1^2} + \dfrac{r_2}{\sigma_2^2} + \dfrac{r_3}{\sigma_3^2}}{\dfrac{1}{\sigma_1^2} + \dfrac{1}{\sigma_2^2} + \dfrac{1}{\sigma_3^2}} = \frac{\dfrac{5.1}{0.1} + \dfrac{6.2}{1} + \dfrac{7.3}{5}}{\dfrac{1}{0.1} + \dfrac{1}{1} + \dfrac{1}{5}} = 5.24$$

and

$$\mathrm{var}\left(\hat{A}\right) = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \frac{1}{\sigma_3^2}\right)^{-1} = \left(\frac{1}{0.1} + \frac{1}{1} + \frac{1}{5}\right)^{-1} = 0.0893$$

It can be seen that the variance should be less than the variance of each of the r_1, r_2 and r_3.

Note that if $\sigma_2^2 \to \infty$ and $\sigma_3^2 \to \infty$, then $\mathrm{var}\left(\hat{A}\right) \to \sigma_1^2 = 0.1$.

7.(a)

Since $\{w_n\}$ are IID with variance σ^2, the covariance matrix is:

$$\boldsymbol{C} = \sigma^2 \boldsymbol{I}_N \Rightarrow |\boldsymbol{C}| = \sigma^{2N} \quad \text{and} \quad \boldsymbol{C}^{-1} = \sigma^{-2} \boldsymbol{I}_N$$

Let

$$\boldsymbol{a} = [\sin(\omega+\phi) \ \ \sin(2\omega+\phi) \ \ \cdots \ \ \sin(N\omega+\phi)]^T \text{ and } \boldsymbol{w} = [w_1 \ \ \cdots \ \ w_N]^T$$

In matrix form, we have:

$$\boldsymbol{r} = \boldsymbol{a}\alpha + \boldsymbol{w}$$

The joint PDF of $\boldsymbol{r}$ is then:

$$p(\boldsymbol{r}) = \frac{1}{(2\pi)^{N/2}\sigma^N} e^{-\frac{1}{2\sigma^2}(\boldsymbol{r}-\boldsymbol{a}\alpha)^T(\boldsymbol{r}-\boldsymbol{a}\alpha)}$$

7.(b)
Applying (6.10), we obtain:

$$\hat{\alpha} = \frac{\boldsymbol{a}^T \boldsymbol{C}^{-1} \boldsymbol{r}}{\boldsymbol{a}^T \boldsymbol{C}^{-1} \boldsymbol{a}} = \frac{\boldsymbol{a}^T \boldsymbol{r}}{\boldsymbol{a}^T \boldsymbol{a}}$$

Alternatively, the solution can be obtained in scalar form by first constructing the least squares cost function:

$$J(\tilde{\alpha}) = (\boldsymbol{r} - \boldsymbol{a}\tilde{\alpha})^T (\boldsymbol{r} - \boldsymbol{a}\tilde{\alpha}) = \sum_{n=1}^{N} (r_n - \tilde{\alpha} \sin(\omega n + \phi))^2$$

Differentiating it with respect to $\tilde{\alpha}$ and then setting the resultant expression to zero, we have:

$$\left.\frac{dJ(\tilde{\alpha})}{d\tilde{\alpha}}\right|_{\tilde{\alpha}=\hat{\alpha}} = 0$$

$$\Rightarrow 2\sum_{n=1}^{N}(r_n - \hat{\alpha}\sin(\omega n+\phi))\cdot -\sin(\omega n+\phi) = 0$$

$$\Rightarrow \sum_{n=1}^{N} r_n \sin(\omega n+\phi) = \hat{\alpha}\sum_{n=1}^{N}\sin^2(\omega n+\phi)$$

$$\Rightarrow \hat{\alpha} = \frac{\sum_{n=1}^{N} r_n \sin(\omega n+\phi)}{\sum_{n=1}^{N}\sin^2(\omega n+\phi)}$$

which gives the same result in scalar form.

7.(c)

According to (6.11) and (6.12), we have:

$$\mathbb{E}\{\hat{\alpha}\} = \alpha$$

$$\mathrm{var}(\hat{\alpha}) = (\boldsymbol{a}^T \boldsymbol{C}^{-1} \boldsymbol{a})^{-1} = (\sigma^{-2} \boldsymbol{a}^T \boldsymbol{a})^{-1} = \frac{\sigma^2}{\boldsymbol{a}^T \boldsymbol{a}} = \frac{\sigma^2}{\sum_{n=1}^{N} \sin^2(\omega n + \phi)}$$

Note that substituting $r_n = \alpha \sin(\omega n + \phi) + w_n$ into $\hat{\alpha}$, we get:

$$\hat{\alpha} = \frac{\displaystyle\sum_{n=1}^{N} [\alpha \sin(\omega n + \phi) + w_n] \sin(\omega n + \phi)}{\displaystyle\sum_{n=1}^{N} \sin^2(\omega n + \phi)} = \alpha + \frac{\displaystyle\sum_{n=1}^{N} w_n \sin(\omega n + \phi)}{\displaystyle\sum_{n=1}^{N} \sin^2(\omega n + \phi)}$$

Taking the expected value of $\hat{\alpha}$, we only need to change w_n to $\mathbb{E}\{w_n\} = 0$, also resulting in:

$$\mathbb{E}\{\hat{\alpha}\} = \alpha$$