1. The notation used in the solutions for the concentration (hyper) ellipsoid problems is defined in the lecture supplement on concentration ellipsoids. Note that
\[ \tilde{\theta}^T \Sigma^{-1} \tilde{\theta} = \tilde{\pi}^T \Lambda^{-1} \tilde{\pi} = k^2. \]
Thus \( k\)-level surfaces in \( \tilde{\theta} \)-space uniquely correspond to \( k\)-level surfaces in \( \tilde{\pi} \)-space and vice-versa. Note also that the canonical coordinate-axes in \( \tilde{\pi} \)-space
\[ e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^T \]
where the “1” is in the \( i \)-th position, correspond to the orthonormal directions \( u_i \) in \( \tilde{\theta} \)-space according to \( e_i \Leftrightarrow u_i, \ i = 1, \ldots, n. \)

In \( \tilde{\pi} \)-space we have
\[ \tilde{\pi}^T \Lambda^{-1} \tilde{\pi} = \frac{\tilde{\pi}^2[1]}{\lambda_1^2} + \cdots + \frac{\tilde{\pi}^2[n]}{\lambda_n^2} = k^2 \]
or
\[ \frac{\tilde{\pi}^2[1]}{(k\lambda_1)^2} + \cdots + \frac{\tilde{\pi}^2[n]}{(k\lambda_n)^2} = 1. \]
This shows that each axis \( e_i \) in \( \tilde{\pi} \)-space is a semi-major axis of a hyperellipsoid with length \( k\lambda_i \). Thus in the original \( \tilde{\theta} \)-space each direction \( u_i \) is a semi-major axis of length \( k\lambda_i \). Note that \( \lambda_i \) is the standard deviation of the error in the direction of the \( i \)-th semi-major axis and that \( k \) gives the number of error standard deviations that we might happen to be interested in.

Let \( 0 < k' < k \) and consider the \( k' \)-level surface defined by
\[ \tilde{\theta}^T \Sigma^{-1} \tilde{\theta} = k'^2 < k^2. \]
This defines a hyperellipsoid with semi-major axes of length \( k'\lambda_i, \ i = 1, \cdots, n. \) Since \( k'\lambda_i < k\lambda_i \) for all \( i \) this shows that a \( k' \)-level surface lies entirely within the \( k \)-level surface for any \( k' < k \). Also note that
\[ \left( \alpha \tilde{\theta} \right)^T \Sigma^{-1} \left( \alpha \tilde{\theta} \right) = k^2, \]
where
\[ \alpha = \frac{k}{k'} > 1. \]
Thus any vector \( \tilde{\theta} \) on the \( k' \)-level surface has to be expanded by a factor \( \alpha > 1 \) in order to lie on the \( k \)-level surface.

We have shown that the region defined by
\[ \tilde{\theta}^T \Sigma^{-1} \tilde{\theta} < k^2 \]
is the interior of the \( k \)-level surface of the hyperellipsoid.
2. (i) Note that for \( i = 1, 2 \),
\[
E_\theta \left\{ a^T \tilde{\theta}_i \right\} = a^T E_\theta \left\{ \tilde{\theta}_i \right\} = 0
\]
for all \( a \). Therefore from the fact that \( \Sigma_1 \leq \Sigma_2 \), for all \( a \) we have
\[
a^T \Sigma_1 a \leq a^T \Sigma_2 a
\]
\[
a^T E_\theta \left\{ \tilde{\theta}_1 \tilde{\theta}_1^T \right\} a \leq a^T E_\theta \left\{ \tilde{\theta}_2 \tilde{\theta}_2^T \right\} a
\]
\[
E_\theta \left\{ \left( a^T \tilde{\theta}_1 \right)^2 \right\} \leq E_\theta \left\{ \left( a^T \tilde{\theta}_2 \right)^2 \right\}
\]
\[
Cov_\theta \left\{ a^T \tilde{\theta}_1 \right\} \leq Cov_\theta \left\{ a^T \tilde{\theta}_2 \right\}.
\]

(ii) The form of the Tchebycheff inequality relevant to this problem is
\[
P \left( |e| \leq L \right) \geq 1 - \frac{E \left\{ e^2 \right\}}{L^2} \triangleq p.
\]
Setting \( L = L_i \) and \( e = a^T \tilde{\theta}, i = 1, 2 \), for a specified lower–bound probability \( p \) we have
\[
p = 1 - \frac{Cov_\theta \left\{ a^T \tilde{\theta}_i \right\}}{L_i^2},
\]
for \( i = 1, 2 \). This yields
\[
L_1 = \sqrt{\frac{Cov_\theta \left\{ a^T \tilde{\theta}_1 \right\}}{1 - p}} \leq \sqrt{\frac{Cov_\theta \left\{ a^T \tilde{\theta}_2 \right\}}{1 - p}} = L_2.
\]

We can also determine that
\[
\frac{L_1}{L_2} = \frac{Cov_\theta \left\{ a^T \tilde{\theta}_1 \right\}}{Cov_\theta \left\{ a^T \tilde{\theta}_2 \right\}}.
\]

This shows that in any direction in \( \tilde{\theta} \)–space, the at–least probability–\( p \) upper error–bound \( L_i, i = 1, 2 \), in the estimate is smaller for \( \tilde{\theta}_1 \) than for \( \tilde{\theta}_2 \).

3. Given that \( 0 < \Sigma_1 \leq \Sigma_2 \) we want to prove that \( 0 < \Sigma_2^{-1} \leq \Sigma_1^{-1} \). We will do this by figuring out how to simultaneously diagonalized \( \Sigma_1 \) and \( \Sigma_2 \). Note that if \( \Sigma > 0 \) it is easy to show that \( \Sigma^{-1} > 0 \).

\(^1\)Convince yourself of this fact.
**Why diagonalize?** We first diagonalize because we can easily show that the desired result holds for *positive definite* diagonal matrices. First note that diagonal matrices *commute*, i.e.

\[ D_1 D_2 = D_2 D_1 \]

for all diagonal matrices \( D_1 \) and \( D_2 \). Also note that the square root of a positive definite matrix is a positive definite matrix. Thus for two positive definite diagonal matrices such that

\[ 0 < D_1 \leq D_2 \]

we have for all \( x \)

\[ 0 \leq x^T (D_2 - D_1)x = x^T D_1 (D_1^{-1} - D_2^{-1}) D_2 x \]

\[ = x^T D_1^{\frac{1}{2}} D_2^{\frac{1}{2}} (D_1^{-1} - D_2^{-1}) D_2^{\frac{1}{2}} D_1^{\frac{1}{2}} x \]

\[ = y^T (D_1^{-1} - D_2^{-1}) y \]

for all \( y = D_2^{\frac{1}{2}} D_1^{\frac{1}{2}} x \). Therefore \( 0 \leq D_1^{-1} - D_2^{-1} \) or

\[ 0 < D_2^{-1} \leq D_1^{-1} . \]

**Simultaneous Diagonalization.** We diagonalize \( \Sigma_1 \) as

\[ \Sigma_1 = U \Lambda U^T \]

which yields

\[ \Lambda^{-\frac{1}{2}} U^T \Sigma_1 U \Lambda^{-\frac{1}{2}} = I . \]  

(1)

We can then form the following positive definite matrix from \( \Sigma_2 \),

\[ \Lambda^{-\frac{1}{2}} U^T \Sigma_2 U \Lambda^{-\frac{1}{2}} , \]

which, of course, can also be diagonalized,

\[ \Lambda^{-\frac{1}{2}} U^T \Sigma_2 U \Lambda^{-\frac{1}{2}} = V \Pi V^T , \]

yielding

\[ V^T \Lambda^{-\frac{1}{2}} U^T \Sigma_2 U \Lambda^{-\frac{1}{2}} V = \Pi . \]  

(2)

Now premultiply (1) by \( V^T \) and then postmultiply the result by \( V \). This yields,

\[ V^T \Lambda^{-\frac{1}{2}} U^T \Sigma_1 U \Lambda^{-\frac{1}{2}} V = I . \]

(3)

Inspection of (2) and (3) shows that we have simultaneously diagonalized \( \Sigma_1 \) and \( \Sigma_2 \) as

\[ A^T \Sigma_1 A = I \]

\[ A^T \Sigma_2 A = \Pi \]

where \( A = U \Lambda^{-\frac{1}{2}} V \)
with \( A \) invertible (but not generally orthogonal). Note that this can be rewritten as

\[
\Sigma_1 = A^{-T}A^{-1}, \\
\Sigma_2 = A^{-T}\Pi A^{-1},
\]

which, when inverted, yields

\[
\Sigma_1^{-1} = AA^T, \\
\Sigma_2^{-1} = A\Pi^{-1}A^T.
\]

**Completion of the Proof.** We have that for all \( x \)

\[
0 \leq x^T(\Sigma_2 - \Sigma_1)x = (A^{-1}x)^T(\Pi - I)(A^{-1}x) = y^T(\Pi - I)y,
\]

for all \( y = A^{-1}x \). Therefore

\[
0 < I \leq \Pi,
\]

which yields (from our earlier results for diagonal matrices)

\[
0 < \Pi^{-1} \leq I.
\]

Final note that for all \( x \),

\[
x^T(\Sigma_1^{-1} - \Sigma_2^{-1})x = (A^T x)^T(I - \Pi^{-1})(A^T x) = y^T( I - \Pi^{-1})y \geq 0,
\]

for all \( y = A^T x \) showing that \( \Sigma_1^{-1} \geq \Sigma_2^{-1} \) as claimed.

4. (i) This was essentially done in Problem (1) above.

(ii) The form of the Tchebycheff inequality most convenient to this problem is

\[
P (y \leq \epsilon) \geq 1 - \frac{E\{y\}}{\epsilon}, \quad y \geq 0, \quad \epsilon > 0.
\]

In particular we take \( \epsilon = k^2 \) and \( y = \tilde{\theta}^T\Sigma^{-1}\tilde{\theta} \). Note that

\[
E_{\theta}\{y\} = E_{\theta}\{\tilde{\theta}^T\Sigma^{-1}\tilde{\theta}\} = E_{\theta}\{\text{trace} \Sigma^{-1} \tilde{\theta}\tilde{\theta}^T\} = \text{trace} \Sigma^{-1} \Sigma = \text{trace} I = n.
\]

Then

\[
P \left( \tilde{\theta}^T\Sigma^{-1}\tilde{\theta} \leq k^2 \right) \geq 1 - \frac{n}{k^2} \triangleq p.
\]

If we set a value for the probability level \( p \), this sets a value for \( k \) of

\[
k = \sqrt{\frac{n}{1-p}}.
\]
(iii) Note that $\Sigma_2^{-1} \leq \Sigma_1^{-1}$. For the same vector $\tilde{\theta}$, define $k > 0$ and $k' > 0$ by

$$k^2 \triangleq \tilde{\theta}^T \Sigma_1^{-1} \tilde{\theta}.$$ 

and

$$k'^2 \triangleq \tilde{\theta}^T \Sigma_2^{-1} \tilde{\theta}.$$ 

We have

$$k'^2 = \tilde{\theta}^T \Sigma_2^{-1} \tilde{\theta} \leq \tilde{\theta}^T \Sigma_1^{-1} \tilde{\theta} = k^2.$$

This shows that

$$\left(\alpha \tilde{\theta}\right)^T \Sigma_2^{-1} \left(\alpha \tilde{\theta}\right) = k^2$$

where

$$\alpha = \frac{k}{k'} \geq 1.$$ 

This shows that every $\tilde{\theta}$ on a $k$–level surface of $\Sigma_1$ has to be “grown” by a factor of $\alpha \geq 1$ in order to then lie on the $k$–level surface (same $k$) of $\Sigma_2$. Thus the $k$–level surface of $\Sigma_1$ lies entirely within the $k$-level surface (same $k$) of $\Sigma_2$. An equivalent statement is that along any direction in error space, $k$ standard deviations of the error $\tilde{\theta}_1$ is less than $k$ standard deviations of the error $\tilde{\theta}_2$.

(iv) If two error covariances can be ordered as $\Sigma_1 \leq \Sigma_2$ then the $k$–level error-concentration ellipsoids of $\Sigma_1$ lies entirely within the $k$–level (same $k$) error concentration ellipsoids of $\Sigma_2$. If the two error covariances cannot be ordered in this manner, then their $k$–level (same $k$) concentration ellipsoids cannot be properly nested one inside the other.

5. (a) i. $M$ has $n$ eigenvector-eigenvalue pairs $(\lambda_i, x_i)$, $M x_i = \lambda_i x_i$, $i = 1, \ldots, n$, where the $n$ eigenvectors are linearly independent.\(^{2}\) Thus,

$$M \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} \lambda_1 \\ 0 \\ 0 \\ \vdots \\ \lambda_n \end{bmatrix}$$

\(^{2}\) This is a nontrivial assumption. A sufficient condition for it to be true are that all $n$ eigenvalues are distinct. Another is that the matrix $M$ be normal, $MM^H = M^HM$. Since a hermitian matrix, $M = M^H$ is normal, an $n \times n$ hermitian matrix has $n$ linearly independent eigenvectors. Note that this implies that a real symmetric matrix has $n$ linearly independent eigenvectors. Normalizing the eigenvectors to have unit length, it is not hard to show that a hermitian matrix also has real eigenvalues: $\lambda_i = \lambda_i x_i^H x_i = x_i^H M x_i = x_i^T M^T x_i = x_i^H M^H x_i = x_i^H M x_i = \lambda_i$. Although we don’t prove it, it is also the case that the $n$ eigenvectors of a hermitian matrix can be taken to be mutually orthonormal.
or \( MX = X \Lambda \). With the assumption that the \( n \) eigenvectors are linearly independent, the so-called modal matrix \( X \) is invertible so that \( \Lambda = X^{-1} MX \).

It is also evident that \( M = X \Lambda X^{-1} \). Thus,

\[ \text{tr } M = \text{tr } (X \Lambda X^{-1}) = \text{tr } (X \Lambda^{-1} X) = \text{tr } \Lambda = \sum_{i=1}^{n} \lambda_i. \tag{4} \]

Because a real \( n \times n \) can be viewed as a special type of a complex matrix (one for which all elements have zero imaginary parts), the result (4) also holds for real, symmetric matrices \( M = M^H = M^T \).

ii. As mentioned in Footnote 2, an \( n \times n \) hermitian matrix is guaranteed to have a full set of \( n \) linearly independent eigenvectors \( x_i, i = 1, \ldots, n \). Therefore, given the result (4) which holds for such matrices, to show that a complex \( n \times n \) matrix \( M \) that is hermitian and positive semidefinite satisfies

\[ \text{tr } M = \sum_{i=1}^{n} \lambda_i \geq 0, \tag{5} \]

it is enough to show that each eigenvalue \( \lambda_i \) is real and nonnegative.

This is straightforward to do because, by definition of being positive semidefinite, we have

\[ \mathbb{R} \ni x^H M x \geq 0 \text{ for all } x \in \mathbb{C}^n. \]

If in particular we take \( x = x_i \), with \( x_i \) normalized, we have

\[ 0 \leq x_i^H M x_i = \lambda_i x_i^H x_i = \lambda_i. \]

The result (5) also holds for real symmetric positive semidefinite matrices as such matrices can be view as special forms of hermitian semidefinite matrices.

(b) From the definition \( \text{MSE}_\theta(\hat{\theta}) = E_\theta \{ \tilde{\theta} \tilde{\theta}^T \} \), \( \tilde{\theta} = \hat{\theta} - \theta \), the matrix \( \text{MSE}_\theta(\hat{\theta}) \) is symmetric positive semidefinite for all parameter vectors \( \theta \) and for estimators \( \hat{\theta} \). Therefore

\[ \text{tr MSE}_\theta(\hat{\theta}) = E_\theta \{ \tilde{\theta}^T \tilde{\theta} \} = E_\theta \{ \| \tilde{\theta} \|^2 \} = \text{mse}_\theta(\hat{\theta}) \geq 0, \]

as expected from our nonnegativity result for the trace of a symmetric positive semidefinite derived in Part (a) of this problem.

From the definition of the positive semidefinite partial ordering, we have

\[ \text{MSE}_\theta(\hat{\theta}_*) \leq \text{MSE}_\theta(\hat{\theta}) \iff M_\theta \triangleq \text{MSE}_\theta(\hat{\theta}) - \text{MSE}_\theta(\hat{\theta}_*) \geq 0. \]

---

3This shows that an \( n \times n \) matrix with \( n \) linearly independent eigenvalues is diagonalizable. In particular hermitian and symmetric matrices are diagonalizable.

4The matrix \( M \) being hermitian is sufficient for the quadratic form \( x^H M x \) to be real (just show that \( x^H M x = \overline{x^H M x} \)). \( M \) being positive semidefinite imposes the addition condition that this quadratic form be nonnegative.
Note that $M_{\theta}$ is real, symmetric, and positive semidefinite. Therefore, from our results in Part (a),

$$0 \leq \text{tr } M_{\theta} = \text{tr } \text{MSE}_{\theta}(\hat{\theta}) - \text{tr } \text{MSE}_{\theta}(\hat{\theta}_*) = \text{mse}_{\theta}(\hat{\theta}) - \text{mse}_{\theta}(\hat{\theta}_*).$$

Thus we have shown that

$$\text{MSE}_{\theta}(\hat{\theta}_*) \leq \text{MSE}_{\theta}(\hat{\theta}) \implies \text{mse}_{\theta}(\hat{\theta}_*) \leq \text{mse}_{\theta}(\hat{\theta}).$$

6. Note that the Gaussian Linear Model meets the requirements of the Gauss-Markov Theorem. Also (ignoring minor notational differences) note from the solution to Homework Problem 5.1 and the class lecture on the BLUE, that the BLUE and the MLE under the Gaussian assumption are identical, thus trivially the MLE is the BLUE by inspection.

The solution in either case is $\hat{x} = A^+ y$ where,

$$A^+ = (A^T C^{-1} A)^{-1} A^T C^{-1},$$

is a weighted–norm pseudoinverse of $A$ which, because $A$ is assumed full column rank, is also a left–inverse, $A^+ A = I$. As shown in the solution to HP 5.1, this results in $\hat{x}$ being an absolutely unbiased estimator of $x$.

Under the Gaussian assumption it is readily shown that

$$\ln p_x(y) \sim -\frac{1}{2} \|y - Ax\|_2^2,$$

ignoring (as usual) terms independent of $x$. Taking the derivative of this with respect to $x$ (using the row vector definition),

$$\frac{\partial}{\partial x} \ln p_x(y) = (y - Ax)^T C^{-1} A.$$

Taking the transpose of this yields the score function,

$$S(x) = A^T C^{-1}(y - Ax).$$

(Note that the score has zero mean for all $x$, as required.) The Fisher information matrix of an unbiased estimator is the covariance of the score function,

$$J_x = \mathbb{E}_x \{S(x)S^T(x)\}$$

$$= A^T C^{-1} \mathbb{E}_x \{ (y - Ax)(y - Ax)^T \} C^{-1} A$$

$$= A^T C^{-1} \mathbb{E}_x \{ nn^T \} C^{-1} A$$

$$= A^T C^{-1} CC^{-1} A = A^T C^{-1} A,$$
which is this instance happens to be independent of \( x \). The Cramér–Rao lower bound is then,
\[
CRLB_x = J_x^{-1} = (A^T C^{-1} A)^{-1},
\]
which in this instance is independent of \( x \).

Because \( A^+ \) is a left inverse, it is easy to show that,
\[
\tilde{x} = \hat{x} - x = A^+ n,
\]
where \( \tilde{x} \) has zero mean. We have,
\[
Cov_x \{ \hat{x} \} = Cov_x \{ \tilde{x} \} = A^+ E \{ nn \} A^+ T = A^+ C A^+ T = (A^T C^{-1} A)^{-1}.
\]
Thus the MLE attains the CRLB and is therefore efficient.

Alternatively, note that
\[
S(x) = A^T C^{-1} (y - Ax)
\]
\[
= A^T C^{-1} (A (A^T C^{-1} A)^{-1} A^T C^{-1} y - Ax)
\]
\[
= J_x (A^+ y - x),
\]
which is the necessary and sufficient condition for the unbiased estimator \( A^+ y \) to attain the CRLB.

7. (a) This problem is an easy application of Homework Problem 5.6 once it is recognized that \( \hat{\theta}(y) \) is an unbiased estimator of \( \theta + b(\theta) \).

With \( E_\theta \{ \hat{\theta} \} = b(\theta) \) we have
\[
\text{MSE}_\theta(\hat{\theta}) \triangleq E_\theta \{ \tilde{\theta} \hat{\theta}^T \}
\]
\[
= E_\theta \{ \left[ (\tilde{\theta} - b(\theta)) + b(\theta) \right] \left[ (\tilde{\theta} - b(\theta)) + b(\theta) \right]^T \}
\]
\[
= \text{Cov}_\theta(\tilde{\theta}) + b(\theta) b^T(\theta)
\]

To complete the rest of the problem we will use the result of Homework Prob. 5.6.

To do so take
\[
g(\theta) \triangleq \theta + b(\theta)
\]
and note that the Jacobian matrix of \( g(\theta) \) is given by
\[
g'(\theta) = I + b'(\theta)
\]

Also note that
\[
\hat{g}(y) \triangleq \hat{\theta}(y)
\]

8
is a uniformly unbiased estimate of $g(\theta)$

$$E_{\theta} \{ \hat{g}(y) \} = E_{\theta} \left\{ \hat{\theta}(y) \right\} = \theta + b(\theta) = g(\theta)$$

We have

$$\text{Cov}_{\theta}(\hat{\theta}) = E_{\theta} \left\{ \left( \hat{\theta} - b(\theta) \right) \left( \hat{\theta} - b(\theta) \right)^T \right\}$$

$$= E_{\theta} \left\{ \left[ \hat{\theta} - (\theta + b(\theta)) \right] \left[ \hat{\theta} - (\theta + b(\theta)) \right]^T \right\}$$

$$= \text{MSE}_{\theta}(\hat{g}) = \text{Cov}_{\theta}(\hat{g}) = \text{Cov}_{\theta}(\tilde{g})$$

$$\geq g'(\theta) J^{-1}(\theta) g'(\theta)$$  \hspace{1cm} \text{(from Homework Problem 5.6)}

or

$$\text{Cov}_{\theta}(\hat{\theta}) \geq (I + b'(\theta))^T J^{-1}(\theta) (I + b'(\theta))^T$$

yielding the result to be proved. This shows that

$$\text{MSE}_{\theta}(\hat{\theta}) \geq (I + b'(\theta))^T J^{-1}(\theta) (I + b'(\theta))^T + b(\theta)b^T(\theta)$$

Note that the right hand side of this inequality depends upon the bias of the particular estimator $\hat{\theta}$ whereas within the class of unbiased estimators the CRLB is independent of the (unbiased) estimator.

(b) For a scalar parameter $\theta$, an efficient biased estimator $\hat{\theta}$ will be uniformly better in the mean square error sense than an efficient unbiased estimator if uniformly in $\theta$ we have

$$\frac{(1 + b'(\theta))^2}{J(\theta)} + b^2(\theta) \leq \frac{1}{J(\theta)}$$  \hspace{1cm} (6)

with strict inequality for at least one value of $\theta$. Condition (6) is satisfied (with strict inequality for all $\theta$) if for all $\theta$

$$0 < b^2(\theta) < \frac{1 - (1 + b'(\theta))^2}{J(\theta)}$$  \hspace{1cm} (7)

The Assumption that

$$-1 < b'(\theta) < -\epsilon < 0$$

for all $\theta$ yields

$$1 - (1 - \epsilon)^2 < 1 - (1 + b'(\theta))^2$$

for all $\theta$. Therefore a sufficient condition for (7) (and hence (6)) to hold is

$$0 < b^2(\theta) < \frac{1 - (1 - \epsilon)^2}{J(\theta)}$$
for all $\theta$. For example, if $\epsilon = 0.2$ the sufficient condition becomes

$$b^2(\theta) < \frac{0.36}{J(\theta)}$$

for all $\theta$.

8. First consider a single measurement $y$. The random variable $y$ has a Poisson distribution with parameter $\lambda$. It is straightforward to determine that $E_\lambda \{y\} = \lambda$. For a single measurement, $y$, one can easily compute the quantities

$$S(y; \lambda) = \frac{\partial}{\partial \lambda} \ln P(y; \lambda) = \frac{y}{\lambda} - 1$$

$$- \frac{\partial}{\partial \lambda} S(y; \lambda) = \frac{y}{\lambda^2}$$

This yields the single measurement Fisher Information, $J_{(1)}(\lambda)$

$$J_{(1)}(\lambda) = -E_\lambda \left\{ \frac{\partial}{\partial \lambda} S(y; \lambda) \right\} = \frac{E_\lambda \{y\}}{\lambda^2} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}.$$ 

Now consider the $m$ iid measurements $Y^m = \{y_1, \cdots, y_m\}$. It can easily be ascertained that

$$S(Y^m; \lambda) = \sum_{i=1}^{m} S(y_i; \lambda)$$

$$- \frac{\partial}{\partial \lambda} S(Y^m; \lambda) = - \sum_{i=1}^{m} \frac{\partial}{\partial \lambda} S(y_i; \lambda).$$

From this the full measurement Fisher Information is determined to be

$$J_{(m)}(\lambda) = -E_\lambda \left\{ \frac{\partial}{\partial \lambda} S(Y^m; \lambda) \right\} = - \sum_{i=1}^{m} E_\lambda \left\{ \frac{\partial}{\partial \lambda} S(y_i; \lambda) \right\} = m J_{(1)}(\lambda) = \frac{m}{\lambda}.$$ 

This yields the Cramér–Rao lower bound,

$$\text{CRLB}(m) = J_{(m)}^{-1}(\lambda) = \frac{1}{m} J_{(1)}^{-1}(\lambda) = \frac{\lambda}{m}.$$ 

Note that $\text{CRLB}(m) \to 0$ as $m \to \infty$, showing that an efficient unbiased estimator $\hat{\lambda}(m)$ (if it exists) of $\lambda$ is consistent.\footnote{It is also straightforward to compute $E_\lambda \{y(y-1)\}$, from which one can readily determine that $\text{Var}_\lambda \{y\} = \lambda$, although we do not use this fact here. \footnote{Why is this fact true?}}
9. Kay 4.1 and 6.1. Except for specifying the precise form of the noise distribution in 4.1, the two problems are essentially identical in structure (viewing the known parameters \(p\) and \(r_i\) as TBD for each specific problem). As a consequence the estimator in both cases is the BLUE and has exactly the same form. Therefore, we only have to solve for the general form of the BLUE once, and then apply the solution for a specific choice of \(p\) and \(r_i\). However, in 6.1 the estimator can be claimed to be the BLUE only (and generally is not a MVUE) because the noise is not assumed to be Gaussian. In 4.1, the BLUE is also a MVUE because here the noise is also taken to be Gaussian. For arbitrary \(p\) the problem can be recast as

\[
x = H\theta + w
\]

where

\[
x = \begin{pmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{pmatrix}, \quad H = \begin{pmatrix} 1 & \cdots & 1 \\ r_1 & \cdots & r_p \\ \vdots & \cdots & \vdots \\ r_1^{N-1} & \cdots & r_p^{N-1} \end{pmatrix}, \quad \theta = \begin{pmatrix} A_1 \\ \vdots \\ A_p \end{pmatrix}, \quad w = \begin{pmatrix} w[0] \\ w[1] \\ \vdots \\ w[N-1] \end{pmatrix}
\]

with

\[
E_\theta \{w\} = 0 \quad \text{and} \quad \text{Cov}_\theta \{w\} = \sigma^2 I.
\]

The general form of the BLUE is given by,

\[
\hat{\theta} = (H^T H)^{-1} H^T x \quad \text{with} \quad \text{Cov}_\theta \{\hat{\theta}\} = \sigma^2 (H^T H)^{-1}.
\]

**Problem 4.1.** Here, we take \(p = 2\), \(r_1 = 1\), \(r_2 = -1\), and \(N\) even. It is easily shown that for this case,

\[
H^T H = N \cdot I_{2 \times 2},
\]

and

\[
H^T x = \begin{pmatrix} \sum_{n=0}^{N-1} x[n] \\ \sum_{n=0}^{N-1} (-1)^n x[n] \end{pmatrix}
\]

yielding

\[
\hat{\theta} = \begin{pmatrix} \hat{A}_1 \\ \hat{A}_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{N} \sum_{n=0}^{N-1} x[n] \\ \frac{1}{N} \sum_{n=0}^{N-1} (-1)^n x[n] \end{pmatrix}
\]

\[\text{Specifically, we take } p = 2 \text{ in the second part of 4.1 and } p = 1 \text{ in 6.1. In the second part of 4.1, we set } r_1 = 1 \text{ and } r_2 = -1. \text{ In 6.1, } r = r_1 \text{ is left unspecified but assumed known.}\]
with

\[ \text{Cov}_{\theta} \{ \hat{\theta} \} = \left( \begin{array}{cc} \sigma^2 & 0 \\ \frac{\sigma^2}{N} & \frac{\sigma^2}{N} \end{array} \right). \]

In this (the Gaussian) case, BLUE = MVUE. Furthermore, we see that the estimator is consistent.

**Problem 6.1.** In this case we have \( p = 1 \) and \( r = r_1 = \) unspecified, yielding

\[ H^T H = \sum_{n=0}^{N-1} r^{2n} \quad \text{and} \quad \sum_{n=0}^{N-1} r^n x[n]. \]

The resulting estimator is

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

with

\[ \text{Cov}_A \{ \hat{A} \} = \frac{\sigma^2}{N-1} \sum_{n=0}^{N-1} r^{2n}. \]

Note that if \( r^2 < 1 \), then in the limit as \( N \to \infty \) we have

\[ \sum_{n=0}^{\infty} r^{2n} = \frac{1}{1 - r^2} < \infty \]

showing that in this case we do not have a consistent estimator. However, if \( r^2 \geq 1 \) this sum is divergent, and the estimator is consistent. In either case the BLUE cannot be claimed to be also a MVUE because we do not know if the noise is Gaussian.

10. Kay 4.13. Assuming that \( \text{Cov}_{\theta} \{ w \} = \sigma^2 I \), setting \( K = (H^T H)^{-1} H^T \),

\[ \hat{\theta} = Kx = \theta + Kw \]

\[ \text{E}_{\theta} \{ \hat{\theta} \} = \theta + \text{E}_{\theta} \{ Kw \} = \theta + B(\theta) \]

where the bias \( B(\theta) = \text{E}_{\theta} \{ Kw \} \). If \( H \) and \( w \) are independent then

\[ B(\theta) = \text{E}_{\theta} \{ K \} \text{E}_{\theta} \{ w \} = 0 \]

and the estimator is unbiased in this case. If \( H \) and \( w \) are not independent, then generally the estimator is biased. Furthermore in the latter case the computation
of the mse or the variance is much more complex. On the other hand, assuming independence we have that $\text{MSE}_\theta = \text{Cov}_\theta \{ \hat{\theta} \}$ where

$$\text{Cov}_\theta \{ \hat{\theta} \} = E_\theta \{ \tilde{\theta} \tilde{\theta}^T \} = E_\theta \{ KE_\theta \{ w w^T \} K \} = E_\theta \{ K \sigma^2 \{ K \} = \sigma^2 E_\theta \{ \sigma^2 (H^T H)^{-1} \} \} .$$

11. Kay 6.12. For the model

$$x = H\theta + w$$

with

$$E_\theta \{ w \} = 0 \quad \text{and} \quad \text{Cov}_\theta \{ w \} = C$$

the BLUE of $\theta$ is given by $\hat{\theta} = Kx$ where

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

is a left inverse of $H$, $KH = I$.

Now consider the reparameterization

$$\alpha = B\theta + B$$

with $B$ invertible. Setting

$$\theta = B^{-1} (\alpha - b)$$

and substituting into our original model yields

$$x + HB^{-1} b = HB^{-1} \alpha + w .$$

Defining

$$A = HB^{-1} \quad \text{and} \quad y = x + Ab$$

we have recast our original model as

$$y = A\alpha + w$$

with

$$E_{\alpha} \{ w \} = 0 \quad \text{and} \quad \text{Cov}_{\alpha} \{ w \} = C .$$

The BLUE of $\alpha$ is given by $\hat{\alpha} = Ly$ where

$$L = (A^T C^{-1} A)^{-1} A^T C^{-1} = B (H^T C^{-1} H)^{-1} H^T C^{-1} = BK$$

is a left inverse of $A$, $LA = I$. We have that

$$\hat{\alpha} = Ly = L(x + Ab) = Lx + b = BKx + b = B\hat{\theta} + b .$$
12. We have \( m \) iid measurements \( y_k \sim U[0, \theta], \ k = 1, \cdots, m, \) for unknown \( 0 < \theta < \infty \) and we want to find the BLUE of \( \theta \). This is an interesting problem because, as we know from Kay Problem 3.1, this distribution does not have a Cramér–Rao lower bound.

Note that for all \( k = 1, \cdots, m \) we have \( \mathbb{E}\{y_k\} = \frac{\theta}{2} \) and \( \text{Cov}_\theta\{y_k\} = \frac{\theta^2}{12} \). We choose to model this as

\[
y_k = \frac{\theta}{2} + w_k, \quad k = 1, \cdots, m, \]

where

\[
\mathbb{E}\{w_k\} = 0 \quad \text{and} \quad \text{Cov}_\theta\{w_k\} = \frac{\theta^2}{12}.
\]

Defining

\[
y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad A = \begin{pmatrix} \frac{1}{2} \\ \vdots \\ \frac{1}{2} \end{pmatrix}, \quad w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}
\]

this can be equivalently recast as,

\[
y = A\theta + w
\]

with

\[
\mathbb{E}\{w\} = 0 \quad \text{and} \quad \text{Cov}_\theta\{w\} = \sigma^2(\theta) \cdot I
\]

where

\[
\sigma^2(\theta) = \frac{\theta^2}{12}.
\]

(a) Although \( \text{Cov}_\theta\{w\} \) is functionally dependent upon the unknown parameter \( \theta \), it is off the form \( c(\theta) \cdot M \) where \( c(\theta) \) is a scalar and \( M \) is a known matrix independent of \( \theta \). Thus the BLUE is realizable and given by,

\[
\hat{\theta} = (A^T A)^{-1} A^T y = \frac{2}{m} \sum_{k=1}^{m} y_k = 2 \langle y_k \rangle = 2 m_1
\]

where \( \langle y_k \rangle = m_1 \) are two alternative ways to indicate the sample mean of the measurements \( y_k \). It is easily demonstrated that \( \hat{\theta} \) is indeed absolutely unbiased. The variance of \( \hat{\theta} \) is

\[
\text{Cov}_\theta\left\{\hat{\theta}\right\} = \sigma^2(\theta) (A^T A)^{-1} = \frac{4 \sigma^2(\theta)}{m} = \frac{4 \theta^2}{12 m} = \frac{\theta^2}{3 m}.
\]

(b) Note that the BLUE is consistent. However, it is not very efficient relative to the true MVUE. The MVUE can be found from the RBLS procedure and has variance \( \frac{\theta^2}{m(m+2)} \). The ratio of the BLUE variance over the MVUE variance is \( \frac{m+2}{3m} \to \infty \) as \( m \to \infty \). Note that for \( m = 1 \) the BLUE and MVUE estimation errors have the same standard deviation (SD), for \( m = 10 \) the BLUE error has twice the SD of the MVUE error, while for \( m = 46 \) the BLUE error has four times the SD of the MVUE error. Despite this relative inefficiency the BLUE is superior to any other linear unbiased estimator of \( \theta \).