Problem 1

Let $\boldsymbol{X}=(X_1,X_2,\ldots,X_n)^T$, then the mean value of this random vector is $\bar{\boldsymbol{X}}=\left(\bar{X}_1,\bar{X}_2,\ldots,\bar{X}_n\right)^T$. The covariance matrix of \boldsymbol{X} is calculated by:

$$oldsymbol{C}_X = \left\langle \left(X - ar{X}
ight)\!\left(X - ar{X}
ight)^T
ight
angle$$

Let $oldsymbol{M}$ be the constant matrix, then :

$$oldsymbol{Y} = oldsymbol{M} \cdot oldsymbol{X}$$

Then, the covariance of \boldsymbol{Y} is calculated by:

$$egin{aligned} oldsymbol{C}_Y &= \left\langle \left(Y - ar{Y}
ight) \left(Y - ar{Y}
ight)^T
ight
angle \ &= \left\langle \left(MX - Mar{X}
ight) \left(MX - Mar{X}
ight)^T
ight
angle \ &= \left\langle M\left(X - ar{X}
ight) \left(X - ar{X}
ight)^T M^T
ight
angle \ &= M \left\langle \left(X - ar{X}
ight) \left(X - ar{X}
ight)^T
ight
angle M^T \ &= M C_X M^T \end{aligned}$$

Problem 2

Method 1

The logarithm of the PDF is:

$$\ln p = -rac{n}{2} \mathrm{ln} \left| 2\pi oldsymbol{C}
ight| - rac{1}{2} \Big[\left(oldsymbol{x} - oldsymbol{\mu} \left(oldsymbol{ heta}_{true}
ight)
ight)^T \cdot oldsymbol{C}^{-1} \cdot \left(oldsymbol{x} - oldsymbol{\mu} \left(oldsymbol{ heta}_{true}
ight)
ight) \Big]$$

Then, the first order derivative w.r.t θ is:

$$egin{aligned} rac{\partial \ln p}{\partial heta_i} = &rac{1}{2} \Bigg[\left(rac{\partial oldsymbol{\mu}}{\partial heta_i}
ight)^T \cdot oldsymbol{C}^{-1} \cdot \left(oldsymbol{x} - oldsymbol{\mu} \left(oldsymbol{ heta}_{true}
ight)
ight] \ &+ rac{1}{2} \Bigg[\left(oldsymbol{x} - oldsymbol{\mu} \left(oldsymbol{ heta}_{true}
ight)
ight)^T \cdot oldsymbol{C}^{-1} \cdot \left(rac{\partial oldsymbol{\mu}}{\partial heta_i}
ight) \Bigg] \end{aligned}$$

Then, the second derivative is:

$$egin{aligned} rac{\partial^2 \ln p}{\partial heta_i \partial heta_j} &= rac{1}{2} \left[\left(rac{\partial^2 oldsymbol{\mu}}{\partial heta_i \partial heta_j}
ight)^T \cdot oldsymbol{C}^{-1} \cdot \left(oldsymbol{x} - oldsymbol{\mu} \left(oldsymbol{ heta}_{true}
ight)
ight)
ight] \ &- rac{1}{2} \left[\left(rac{\partial oldsymbol{\mu}}{\partial heta_i}
ight)^T \cdot oldsymbol{C}^{-1} \cdot \left(rac{\partial oldsymbol{\mu}}{\partial heta_i}
ight)
ight] \ &+ rac{1}{2} \left[\left(oldsymbol{x} - oldsymbol{\mu} \left(oldsymbol{ heta}_{true}
ight)
ight)^T \cdot oldsymbol{C}^{-1} \cdot \left(rac{\partial^2 oldsymbol{\mu}}{\partial heta_i \partial heta_i}
ight)
ight] \end{aligned}$$

Then, the i, j-th element in Fisher Matrix is the expectation value of the above equation:

$$F_{ij} = -\left\langle rac{\partial^2 \ln p}{\partial heta_i \partial heta_j}
ight
angle = 0 - rac{1}{2} \left\langle \left[\left(rac{\partial oldsymbol{\mu}}{\partial heta_i}
ight)^T \cdot oldsymbol{C}^{-1} \cdot \left(rac{\partial oldsymbol{\mu}}{\partial heta_j}
ight)
ight]
ight
angle - rac{1}{2} \left\langle \left[\left(rac{\partial oldsymbol{\mu}}{\partial heta_j}
ight)^T \cdot oldsymbol{C}^{-1} \cdot \left(rac{\partial oldsymbol{\mu}}{\partial heta_i}
ight)
ight]
ight
angle + 0$$

Since this two terms are symmetric, we then have:

$$F_{ij} = -\left\langle rac{\partial^2 \ln p}{\partial heta_i \partial heta_j}
ight
angle = -\left\langle \left\lceil \left(rac{\partial oldsymbol{\mu}}{\partial heta_i}
ight)^T \cdot oldsymbol{C}^{-1} \cdot \left(rac{\partial oldsymbol{\mu}}{\partial heta_j}
ight)
ight
ceil
angle$$

Method 2

Some properties that would be used in calculating the gradient of matrix:

$$egin{aligned} x &= \operatorname{tr}\left(x
ight) \ \operatorname{tr}\left(c_{1}oldsymbol{A} + c_{2}oldsymbol{B}
ight) &= c_{1}\operatorname{tr}\left(oldsymbol{A}
ight) + c_{2}\operatorname{tr}\left(oldsymbol{B}
ight) \ \operatorname{tr}\left(oldsymbol{A}
ight) &= \operatorname{tr}\left(oldsymbol{A}^{T}oldsymbol{B}
ight) \ \operatorname{tr}\left(oldsymbol{A}oldsymbol{B}
ight) &= \operatorname{tr}\left(oldsymbol{B}oldsymbol{A}
ight) \ \operatorname{tr}\left(oldsymbol{A}oldsymbol{B}C
ight) &= \operatorname{tr}\left(oldsymbol{C}oldsymbol{A}oldsymbol{B}
ight) \ \operatorname{tr}\left(oldsymbol{A}oldsymbol{B}C
ight) &= \operatorname{tr}\left(oldsymbol{C}oldsymbol{A}oldsymbol{B}
ight) \ \operatorname{tr}\left(oldsymbol{A}oldsymbol{B}C
ight) &= \operatorname{tr}\left(oldsymbol{C}oldsymbol{A}oldsymbol{B}\left(oldsymbol{G}\left(oldsymbol{X}\right)
ight) \ \operatorname{tr}\left(oldsymbol{A}oldsymbol{B}Coldsymbol{A}\right) &= \operatorname{tr}\left(oldsymbol{B}Coldsymbol{A}\right) \ \operatorname{tr}\left(oldsymbol{A}oldsymbol{B}Coldsymbol{A}\right) \ \operatorname{tr}\left(oldsymbol{A}oldsymbol{B}\left(oldsymbol{G}\left(oldsymbol{X}
ight)
ight) \ \operatorname{tr}\left(oldsymbol{B}Coldsymbol{A}\right) &= \operatorname{tr}\left(oldsymbol{B}Coldsymbol{A}\right) \ \operatorname{tr}\left(oldsymbol{A}oldsymbol{B}Coldsymbol{A}\right) \ \operatorname{tr}\left(oldsymbol{B}Coldsymbol{A}\right) \ \operatorname{tr}\left(oldsymbol{B}Coldsymbol{B}Coldsymbol{B}Coldsymbol{A}\right) \ \operatorname{tr}\left(oldsymbol{B}Coldsy$$

And the most important one:

$$\mathrm{tr}\left(\mathrm{d}oldsymbol{F}_{p imes q}
ight)=\mathrm{tr}\left(\left(rac{\mathrm{d}oldsymbol{F}_{p imes q}}{\mathrm{d}oldsymbol{X}}
ight)^{T}\mathrm{d}oldsymbol{X}
ight)$$

So, the derivative of the matrix is then the transpose of the left part.

$$\begin{aligned} \operatorname{tr}\left(\operatorname{d}\ln p\right) &= \operatorname{tr}\left(\frac{1}{2}\left(\frac{\partial \boldsymbol{\mu}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\operatorname{d}\boldsymbol{\theta}\right)^{T} \cdot \boldsymbol{C}^{-1} \cdot \left[\boldsymbol{X} - \boldsymbol{\mu}\left(\boldsymbol{\theta}\right)\right] + \frac{1}{2}\left[\boldsymbol{X} - \boldsymbol{\mu}\left(\boldsymbol{\theta}\right)\right]^{T} \cdot \boldsymbol{C}^{-1} \cdot \left(\frac{\partial \boldsymbol{\mu}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\operatorname{d}\boldsymbol{\theta}\right)\right) \right) \\ &= \operatorname{tr}\left(\frac{1}{2}\left(\frac{\partial \boldsymbol{\mu}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\operatorname{d}\boldsymbol{\theta}\right)^{T} \cdot \boldsymbol{C}^{-1} \cdot \left[\boldsymbol{X} - \boldsymbol{\mu}\left(\boldsymbol{\theta}\right)\right]\right) + \operatorname{tr}\left(\frac{1}{2}\left[\boldsymbol{X} - \boldsymbol{\mu}\left(\boldsymbol{\theta}\right)\right]^{T} \cdot \boldsymbol{C}^{-1} \cdot \left(\frac{\partial \boldsymbol{\mu}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\operatorname{d}\boldsymbol{\theta}\right)\right) \\ &= \frac{1}{2}\operatorname{tr}\left(\left[\boldsymbol{X} - \boldsymbol{\mu}\left(\boldsymbol{\theta}\right)\right]^{T} \cdot \left(\boldsymbol{C}^{-1}\right)^{T} \cdot \left(\frac{\partial \boldsymbol{\mu}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\operatorname{d}\boldsymbol{\theta}\right)\right) + \frac{1}{2}\operatorname{tr}\left(\left[\boldsymbol{X} - \boldsymbol{\mu}\left(\boldsymbol{\theta}\right)\right]^{T} \cdot \boldsymbol{C}^{-1} \cdot \left(\frac{\partial \boldsymbol{\mu}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\operatorname{d}\boldsymbol{\theta}\right)\right) \end{aligned}$$

So, the gradient of $\ln p$ is:

$$rac{\mathrm{d} \ln p}{\mathrm{d} oldsymbol{ heta}} = rac{1}{2} \Bigg(\left(rac{\partial oldsymbol{\mu} \left(oldsymbol{ heta}
ight)}{\partial oldsymbol{ heta}}
ight)^T \cdot \left(\left(oldsymbol{C}^{-1}
ight)^T + oldsymbol{C}^{-1}
ight) \cdot \left[oldsymbol{X} - oldsymbol{\mu} \left(oldsymbol{ heta}
ight) \Bigg] \Bigg)$$

then, take the second derivative of this:

$$d\left(\frac{d \ln p}{d \boldsymbol{\theta}}\right) = \frac{1}{2} d \left(\left(\frac{\partial \boldsymbol{\mu}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\right)^{T}\right) \cdot \left[\left(\boldsymbol{C}^{-1}\right)^{T} + \boldsymbol{C}^{-1}\right] \cdot \left[\boldsymbol{X} - \boldsymbol{\mu}\left(\boldsymbol{\theta}\right)\right] \\ - \frac{1}{2} \left(\frac{\partial \boldsymbol{\mu}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\right)^{T} \cdot \left[\left(\boldsymbol{C}^{-1}\right)^{T} + \boldsymbol{C}^{-1}\right] \cdot d \boldsymbol{\mu}\left(\boldsymbol{\theta}\right)$$

Then take the expectation value of this equation, the first term is 0, so, we have:

$$\left\langle d \left(\frac{d \ln p}{d \boldsymbol{\theta}} \right) \right\rangle = \left\langle -\frac{1}{2} \left(\frac{\partial \boldsymbol{\mu} \left(\boldsymbol{\theta} \right)}{\partial \boldsymbol{\theta}} \right)^{T} \cdot \left[\left(\boldsymbol{C}^{-1} \right)^{T} + \boldsymbol{C}^{-1} \right] \cdot d \boldsymbol{\mu} \left(\boldsymbol{\theta} \right) \right\rangle$$

$$= \left\langle -\frac{1}{2} \left(\frac{\partial \boldsymbol{\mu} \left(\boldsymbol{\theta} \right)}{\partial \boldsymbol{\theta}} \right)^{T} \cdot \left[\left(\boldsymbol{C}^{-1} \right)^{T} + \boldsymbol{C}^{-1} \right] \cdot \frac{\partial \boldsymbol{\mu} \left(\boldsymbol{\theta} \right)}{\partial \boldsymbol{\theta}} d \boldsymbol{\theta} \right\rangle$$

Therefore:

$$\left\langle \mathrm{d} \left(\frac{\mathrm{d}^2 \ln p}{\mathrm{d} oldsymbol{ heta}^2}
ight)
ight
angle = \left\langle - \frac{1}{2} \left(\frac{\partial oldsymbol{\mu} \left(oldsymbol{ heta}
ight)}{\partial oldsymbol{ heta}}
ight)^T \cdot \left[\left(oldsymbol{C}^{-1}
ight)^T + oldsymbol{C}^{-1}
ight] \cdot \frac{\partial oldsymbol{\mu} \left(oldsymbol{ heta}
ight)}{\partial oldsymbol{ heta}}
ight
angle$$

Since the covariance matrix is symmetric, the final result is then:

$$\left\langle \frac{\mathrm{d}^{2}\ln p}{\mathrm{d}\boldsymbol{\theta}^{2}}\right\rangle = -\left\langle \left(\frac{\partial\boldsymbol{\mu}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right)^{T}\cdot\boldsymbol{C}^{-1}\cdot\frac{\partial\boldsymbol{\mu}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right\rangle$$

Then, we have:

$$oldsymbol{F}_{ij} = \left(rac{\partial oldsymbol{\mu}\left(oldsymbol{ heta}
ight)}{\partial heta_i}
ight)^T \cdot oldsymbol{C}^{-1} \cdot rac{\partial oldsymbol{\mu}\left(oldsymbol{ heta}
ight)}{\partial heta_j}$$

When n=1=m,

$$F = \left(rac{\partial \mu}{\partial heta}
ight)^2 rac{1}{\sigma^2}$$

So, higher value of F means that is contains more information. There are two explanations:

1. when the derivative is large, that means μ is sensitive to θ , a small deviation from the true value could lead to a big difference.

2. when the σ variance is small, that means the distribution is more localized around the mean value.

Problem 3

Take the derivative of $\chi^{2}\left(a\right)$:

$$egin{aligned} oldsymbol{\chi}^2\left(a
ight) &= \left[oldsymbol{x} - a oldsymbol{v}
ight]^T oldsymbol{C}^{-1} \left[oldsymbol{x} - a oldsymbol{v}
ight] + \left[oldsymbol{x} - a oldsymbol{v}
ight]^T oldsymbol{C}^{-1} \left[oldsymbol{x} - a oldsymbol{v}
ight] + \left[oldsymbol{x} - a oldsymbol{v}
ight]^T oldsymbol{C}^{-1} \left[oldsymbol{x} - a oldsymbol{v}
ight] \mathrm{d} a \ &= -2 \left[oldsymbol{v}^T oldsymbol{C}^{-1} \left[oldsymbol{x} - a oldsymbol{v}
ight] \right] \ &= -2 \left[oldsymbol{v}^T oldsymbol{C}^{-1} \left[oldsymbol{x} - a oldsymbol{v}
ight] \end{aligned}$$

Since the transpose of a scalar is still itself, we can combine the two term in the third line.

To get the a_{best} that makes the χ^2 minimized, make the derivative 0:

$$rac{\mathrm{d}\left(\chi^{2}\left(a
ight)
ight)}{\mathrm{d}a}=-2\left[oldsymbol{v}^{T}oldsymbol{C}^{-1}\left[oldsymbol{x}-aoldsymbol{v}
ight]
ight]=0 \ a=rac{oldsymbol{v}^{T}oldsymbol{C}^{-1}oldsymbol{x}}{oldsymbol{v}^{T}oldsymbol{C}^{-1}oldsymbol{v}}$$

For a random variable A_{best} , its variance is :

$$egin{aligned} A_{best} &= rac{oldsymbol{X}^T oldsymbol{C}^{-1} oldsymbol{V}}{oldsymbol{V}^T oldsymbol{C}^{-1} oldsymbol{V}} \ \operatorname{Cov}\left(A_{best}
ight) &= \operatorname{Cov}\left(rac{oldsymbol{X}^T oldsymbol{C}^{-1} oldsymbol{V}}{oldsymbol{V}^T oldsymbol{C}^{-1} oldsymbol{V}}
ight) \ &= rac{1}{oldsymbol{(V}^T oldsymbol{C}^{-1} oldsymbol{V})^2} \operatorname{Cov}\left(oldsymbol{V}^T oldsymbol{C}^{-1} oldsymbol{X}
ight) \ &= rac{1}{oldsymbol{(V}^T oldsymbol{C}^{-1} oldsymbol{V})^2} oldsymbol{(V}^T oldsymbol{C}^{-1} oldsymbol{V}^T oldsymbol{V}^T oldsymbol{C}^{-1} oldsymbol{V}^T oldsymbol{C}^{-1} oldsymbol{V}^T oldsymbol{C}^{-1} oldsymbol{V}^T oldsymbol{C}^{-1} oldsymbol{V}^T oldsymbol{C}^{-1} oldsymbol{V}^T oldsymbol{V}^$$

So, the variance of A_{best} is then $\mathrm{Var}\left(A_{best}
ight) = m{V}^{-1}m{C}m{\left(m{V}^T
ight)}^{-1}$.

Fisher matrix for the gaussian distribution is:

$$oldsymbol{F} = \left(rac{\partial oldsymbol{\mu} \left(ec{ heta}
ight)}{\partial ec{ heta}}
ight)^T \cdot oldsymbol{C}^{-1} \cdot \left(rac{\partial oldsymbol{\mu} \left(ec{ heta}
ight)}{\partial ec{ heta}}
ight)$$

Here, the variable is a_{best} , so the Fisher matrix becomes:

$$egin{aligned} m{F} &= \left(rac{\partial am{v}}{\partial a}
ight)^T \cdot m{C}^{-1} \cdot \left(rac{\partial am{v}}{\partial a}
ight) \ &= m{v}^T \cdot m{C}^{-1} \cdot m{v} \end{aligned}$$

So, the inverse of Fisher matrix is:

$$\boldsymbol{F}^{-1} = \boldsymbol{V}^{-1} \cdot \boldsymbol{C} \cdot \left(\boldsymbol{V}^T\right)^{-1}$$

So, this is the same as the inverse of Fisher matrix.

Cramer-Rao-bound:

$$\operatorname{var}\left(\hat{ heta}
ight) \geq rac{1}{F\left(heta
ight)}$$

The precision of any unbiased estimation is at most the Fisher information matrix.

The estimator is statistically efficient, making the best use of the available information in estimating the parameter of interest.