Fisher Information Matrix of Multivariate Gaussian random effect model

Dongyue Xie

1 Model

Assume

$$\boldsymbol{x}_i \sim N(\boldsymbol{0}, \boldsymbol{U} + \boldsymbol{D} + \boldsymbol{V}_i), \tag{1}$$

where $\boldsymbol{V}_i \in \mathbb{R}^{R\times R}$ is a known matrix, $\boldsymbol{D}$ is a diagonal matrix and $\boldsymbol{U}$ is an unstructured covariance matrix. The goal is to derive the Fisher information matrix of parameters in $\boldsymbol{U}, \boldsymbol{D}$.

Denote $\boldsymbol{S}_i = \boldsymbol{U} + \boldsymbol{D} + \boldsymbol{V}_i$, and $\boldsymbol{\theta} = (\boldsymbol{u}, \boldsymbol{\sigma}^2)^T$, where $\boldsymbol{u} = vec(\boldsymbol{U})$ and contains only unique elements so $\boldsymbol{u} \in \mathbb{R}^{R(R+1)/2}$, and $\boldsymbol{\sigma}^2 = (\sigma_1^2, ..., \sigma_R^2)$. Note that elements of $\boldsymbol{S}_i$ are functions of $\boldsymbol{\theta}$ and for simplicity we write $\boldsymbol{S}_i$ instead of $\boldsymbol{S}_i(\boldsymbol{\theta})$.

The log likelihood of $\boldsymbol{x}_i$ is

$$l(\boldsymbol{\theta}; \boldsymbol{x}_i) = -\frac{1}{2}\log|\boldsymbol{S}_i| - \frac{1}{2}\boldsymbol{x}_i^T \boldsymbol{S}_i^{-1}\boldsymbol{x}_i. \tag{2}$$

The gradient of $l(\boldsymbol{\theta}; \boldsymbol{x}_i)$ with respect to the jth entry of $\boldsymbol{\theta}$ gives the score function

$$\frac{\partial l(\boldsymbol{\theta}; \boldsymbol{x}_i)}{\partial \theta_j} = -\frac{1}{2}\operatorname{Tr}(\boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{S}_i}{\partial \theta_j}) + \frac{1}{2}\boldsymbol{x}_i^T \boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{S}_i}{\partial \theta_j}\boldsymbol{S}_i^{-1}\boldsymbol{x}_i. \tag{3}$$

The second order derivative of $l(\boldsymbol{\theta}; \boldsymbol{x}_i)$ with respect to $\boldsymbol{\theta}$ gives the Hessian,

$$\begin{aligned}\frac{\partial^2 l(\boldsymbol{\theta}; \boldsymbol{x}_i)}{\partial \theta_j \partial \theta_k} = &-\frac{1}{2}\operatorname{Tr}\left(\boldsymbol{S}_i^{-1}\frac{\partial^2 \boldsymbol{S}_i}{\partial \theta_j \partial \theta_k} - \boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{S}_i}{\partial \theta_j}\boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{S}_i}{\partial \theta_k}\right) \\ &+ \frac{1}{2}\boldsymbol{x}_i^T \boldsymbol{S}_i^{-1}\left(\frac{\partial^2 \boldsymbol{S}_i}{\partial \theta_j \partial \theta_k} - 2\frac{\partial \boldsymbol{S}_i}{\partial \theta_j}\boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{S}_i}{\partial \theta_k}\right)\boldsymbol{S}_i^{-1}\boldsymbol{x}_i.\end{aligned} \tag{4}$$

The Fisher information matrix is given by the expectation of the negative Hessian,

$$-\mathbb{E}(\frac{\partial^2 l(\boldsymbol{\theta}; \boldsymbol{x}_i)}{\partial \theta_j \partial \theta_k}) = \frac{1}{2}\operatorname{Tr}\left(\boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{S}_i}{\partial \theta_j}\boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{S}_i}{\partial \theta_k}\right). \tag{5}$$

We denote the Fisher information matrix as $\boldsymbol{I}(\boldsymbol{\theta})$, and it's j,kth element is

$$\boldsymbol{I}_{jk}(\boldsymbol{\theta}) = \frac{1}{2}\sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{S}_i}{\partial \theta_j}\boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{S}_i}{\partial \theta_k}\right). \tag{6}$$

The asymptotic variance of $\hat{\boldsymbol{\theta}}^{mle}$ is given by $I(\boldsymbol{\theta})^{-1}$, and can be estimated by $I(\hat{\boldsymbol{\theta}})^{-1}$. We can partition the information matrix into the following four blocks

$$\begin{aligned}\boldsymbol{I}(\boldsymbol{\theta}) &= \left[\begin{array}{c|c} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{u} \partial \boldsymbol{u}^T} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{u} \partial (\boldsymbol{\sigma}^2)^T} \\ \hline \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\sigma^2}^T \partial \boldsymbol{u}} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\sigma^2}^T \partial \boldsymbol{\sigma^2}} \end{array}\right] \\ &= \left[\begin{array}{c|c} \boldsymbol{A} & \boldsymbol{B} \\ \hline \boldsymbol{B}^T & \boldsymbol{C} \end{array}\right]\end{aligned} \tag{7}$$

We order the elements in $\boldsymbol{u}$ as the unique elements in each row of matrix $\boldsymbol{U}$, i.e.

$$\boldsymbol{u} = (u_{11}, u_{12}, ..., u_{1R}, u_{22}, ..., u_{2R}, ..., u_{RR}). \tag{8}$$

The matrix $\boldsymbol{A}$ of dimension $\frac{R(R+1)}{2} \times \frac{R(R+1)}{2}$ is

$$\boldsymbol{A} = \begin{pmatrix} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial^2 u_{11}} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{11} \partial u_{12}} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{11} \partial u_{RR}} \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{12} \partial u_{11}} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial^2 u_{12}} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{12} \partial u_{RR}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{RR} \partial u_{11}} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{RR} \partial u_{12}} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial^2 u_{RR}} \end{pmatrix}. \tag{9}$$

The matrix $\boldsymbol{B}$ of dimension $\frac{R(R+1)}{2} \times R$ is

$$\boldsymbol{B} = \begin{pmatrix} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{11} \partial \sigma_1^2} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{11} \partial \sigma_2^2} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{11} \partial \sigma_R^2} \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{12} \partial \sigma_1^2} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{12} \partial \sigma_2^2} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{12} \partial \sigma_R^2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{RR} \partial \sigma_1^2} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{RR} \partial \sigma_2^2} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{RR} \partial \sigma_R^2} \end{pmatrix}. \tag{10}$$

The matrix $\boldsymbol{C}$ of dimension $R \times R$ is

$$\boldsymbol{C} = \begin{pmatrix} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial^2 \sigma_1^2} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma_1^2 \partial \sigma_2^2} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma_1^2 \partial \sigma_R^2} \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma_2^2 \partial \sigma_1^2} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial^2 \sigma_2^2} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma_2^2 \partial \sigma_R^2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma_R^2 \partial \sigma_1^2} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma_R^2 \partial \sigma_2^2} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial^2 \sigma_R^2} \end{pmatrix}. \tag{11}$$

Let $\boldsymbol{e}_j$ denote a vector that the jth entry is 1 and others are 0. The following derivatives hold

$$\begin{aligned}
\frac{\partial \boldsymbol{S}_i}{\partial u_{rr}} &= \frac{\partial \boldsymbol{U}}{\partial u_{rr}} = \boldsymbol{e}_r \boldsymbol{e}_r^T, \\
\frac{\partial \boldsymbol{S}_i}{\partial u_{r_1 r_2}} &= \frac{\partial \boldsymbol{U}}{\partial u_{r_1 r_2}} = \boldsymbol{e}_{r_1} \boldsymbol{e}_{r_2}^T + \boldsymbol{e}_{r_2} \boldsymbol{e}_{r_1}^T, \\
\frac{\partial \boldsymbol{S}_i}{\partial \sigma_r^2} &= \frac{\partial \boldsymbol{D}}{\partial \sigma_r^2} = \boldsymbol{e}_r \boldsymbol{e}_r^T.
\end{aligned} \tag{12}$$

1.1 Calculation of matrix $\boldsymbol{A}$:

Denote $\tilde{\boldsymbol{S}}_i = \boldsymbol{S}_i^{-1}$, with column vectors $\tilde{\boldsymbol{s}}_{i1}, ..., \tilde{\boldsymbol{s}}_{iR}$.

Denote the (r_1, r_2)th entry of matrix $\tilde{\boldsymbol{S}}^{-1}$ as $\tilde{s}_{i,r_1 r_2}$.

We denote the entries of matrix $\boldsymbol{A}$ as a_{rj_r,lk_l}, for $r, l \in [R]$, $j_r = r, r+1, ..., R-r+1$, and $k_l = l, l+1, ..., R-l+1$. Each entry is given by

$$a_{rj_r,lk_l} = \frac{1}{2} \sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1} \frac{\partial \boldsymbol{S}_i^{-1}}{\partial u_{rj_r}} \boldsymbol{S}_i^{-1} \frac{\partial \boldsymbol{S}_i^{-1}}{\partial u_{lk_l}} \right). \tag{13}$$

The matrix $\boldsymbol{A}$ is symmetric since $a_{rj_r,lk_l} = a_{lk_l,rj_r}$. We consider the following cases:

- When $r = j_r$, $l = k_l$,

$$\begin{aligned}
a_{rr,ll} &= \frac{1}{2} \sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1} \frac{\partial \boldsymbol{U}}{\partial u_{rr}} \boldsymbol{S}_i^{-1} \frac{\partial \boldsymbol{U}}{\partial u_{ll}} \right) \\
&= \frac{1}{2} \sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1} \boldsymbol{e}_r \boldsymbol{e}_r^T \boldsymbol{S}_i^{-1} \boldsymbol{e}_l \boldsymbol{e}_l^T \right) \\
&= \frac{1}{2} \sum_i \boldsymbol{e}_l^T \left(\boldsymbol{S}_i^{-1} \boldsymbol{e}_r \boldsymbol{e}_r^T \boldsymbol{S}_i^{-1} \right) \boldsymbol{e}_l \\
&= \frac{1}{2} \sum_i (\tilde{s}_{i,rl})^2.
\end{aligned} \tag{14}$$

- When $r = j_r, l \neq k_l$,

$$\begin{aligned}
a_{rr,lk_l} &= \frac{1}{2} \sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1} \frac{\partial \boldsymbol{U}}{\partial u_{rr}} \boldsymbol{S}_i^{-1} \frac{\partial \boldsymbol{U}}{\partial u_{lk_l}} \right) \\
&= \frac{1}{2} \sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1} \boldsymbol{e}_r \boldsymbol{e}_r^T \boldsymbol{S}_i^{-1} (\boldsymbol{e}_l \boldsymbol{e}_{k_l}^T + \boldsymbol{e}_{k_l} \boldsymbol{e}_l^T) \right) \\
&= \frac{1}{2} \sum_i \left(\boldsymbol{e}_{k_l}^T \left(\boldsymbol{S}_i^{-1} \boldsymbol{e}_r \boldsymbol{e}_r^T \boldsymbol{S}_i^{-1} \right) \boldsymbol{e}_l + \boldsymbol{e}_l^T \left(\boldsymbol{S}_i^{-1} \boldsymbol{e}_r \boldsymbol{e}_r^T \boldsymbol{S}_i^{-1} \right) \boldsymbol{e}_{k_l} \right) \\
&= \sum_i \tilde{s}_{i,rl} \tilde{s}_{i,rk_l}.
\end{aligned} \tag{15}$$

- When $r \neq j_r, l \neq k_l$,

$$\begin{aligned}
a_{rj_r,lk_l} &= \frac{1}{2}\sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{U}}{\partial u_{rj_r}}\boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{U}}{\partial u_{lk_l}}\right) \\
&= \frac{1}{2}\sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1}(\boldsymbol{e}_r\boldsymbol{e}_{j_r}^T + \boldsymbol{e}_{j_r}\boldsymbol{e}_r^T)\boldsymbol{S}_i^{-1}(\boldsymbol{e}_l\boldsymbol{e}_{k_l}^T + \boldsymbol{e}_{k_l}\boldsymbol{e}_l^T)\right) \qquad (16) \\
&= \sum_i (\tilde{s}_{i,lj_r}\tilde{s}_{i,rk_l} + \tilde{s}_{i,rl}\tilde{s}_{i,j_rk_l}).
\end{aligned}$$

1.2 Calculation of matrix $\boldsymbol{B}$

Denote the entries of matrix $\boldsymbol{B}$ as $b_{rj_r,l}$, for $r, l \in [R]$, $j_r = r, r+1, ..., R-r+1$. Each entry is given by

$$b_{rj_r,l} = \frac{1}{2}\sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{U}}{\partial u_{rj_r}}\boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{D}}{\partial \sigma_l^2}\right). \qquad (17)$$

We consider the following cases:

- When $r = j_r$,

$$\begin{aligned}
b_{rr,l} &= \frac{1}{2}\sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{U}}{\partial u_{rr}}\boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{D}}{\partial \sigma_l^2}\right) \\
&= \frac{1}{2}\sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1}(\boldsymbol{e}_r\boldsymbol{e}_r^T)\boldsymbol{S}_i^{-1}(\boldsymbol{e}_l\boldsymbol{e}_l^T)\right) \qquad (18) \\
&= \frac{1}{2}\sum_i (\tilde{s}_{i,rl})^2.
\end{aligned}$$

- When $r \neq j_r$,

$$\begin{aligned}
b_{rj_r,l} &= \frac{1}{2}\sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{U}}{\partial u_{rj_r}}\boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{D}}{\partial \sigma_l^2}\right) \\
&= \frac{1}{2}\sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1}(\boldsymbol{e}_r\boldsymbol{e}_{j_r}^T + \boldsymbol{e}_{j_r}\boldsymbol{e}_r^T)\boldsymbol{S}_i^{-1}(\boldsymbol{e}_l\boldsymbol{e}_l^T)\right) \qquad (19) \\
&= \sum_i (\tilde{s}_{i,rl}\tilde{s}_{i,j_rl}).
\end{aligned}$$

In the special case where $\boldsymbol{D} = \sigma^2\boldsymbol{I}$, $\boldsymbol{B} = \boldsymbol{b}$ is a vector of length $R(R+1)/2$, and

- When $r = j_r$,

$$\begin{aligned}
b_{rr} &= \frac{1}{2}\sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{U}}{\partial u_{rr}}\boldsymbol{S}_i^{-1}\right) \\
&= \frac{1}{2}\sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1}(\boldsymbol{e}_r\boldsymbol{e}_r^T)\boldsymbol{S}_i^{-1}\right) \qquad (20) \\
&= \frac{1}{2}\sum_i \sum_{l=1}^{R} (\tilde{s}_{i,rl})^2.
\end{aligned}$$

- When $r \neq j_r$,

$$
\begin{aligned}
b_{rj_r} &= \frac{1}{2}\sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{U}}{\partial u_{rj_r}}\boldsymbol{S}_i^{-1}\right)\\
&= \frac{1}{2}\sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1}(\boldsymbol{e}_r\boldsymbol{e}_{j_r}^T + \boldsymbol{e}_{j_r}\boldsymbol{e}_r^T)\boldsymbol{S}_i^{-1}\right)\\
&= \sum_i \sum_{l=1}^{R} (\tilde{s}_{i,rl}\tilde{s}_{i,j_r l}).
\end{aligned}
\tag{21}
$$

1.3 Calculation of matrix $\boldsymbol{C}$

Denote the entries of $\boldsymbol{C}$ as c_{rl}, for $r, l \in [R]$. Each entry of $\boldsymbol{C}$ is given by

$$
\begin{aligned}
c_{rl} &= \frac{1}{2}\sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{D}}{\partial \sigma_r^2}\boldsymbol{S}_i^{-1}\frac{\partial \boldsymbol{D}}{\partial \sigma_l^2}\right)\\
&= \frac{1}{2}\sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1}(\boldsymbol{e}_r\boldsymbol{e}_r^T)\boldsymbol{S}_i^{-1}(\boldsymbol{e}_l\boldsymbol{e}_l^T)\right)\\
&= \frac{1}{2}\sum_i \tilde{s}_{i,rl}^2
\end{aligned}
\tag{22}
$$

In the special case where $\boldsymbol{D} = \sigma^2\boldsymbol{I}$, $\boldsymbol{C} = c$ is a scalar, and

$$
\begin{aligned}
c &= \frac{1}{2}\sum_i \operatorname{Tr}\left(\boldsymbol{S}_i^{-1}\boldsymbol{S}_i^{-1}\right)\\
&= \frac{1}{2}\sum_i\sum_r\sum_l \tilde{s}_{i,rl}^2.
\end{aligned}
\tag{23}
$$

2 Inverse Fisher Information Matrix

We are interested in the variance of $\boldsymbol{\sigma}^2$, which is given by

$$(\boldsymbol{C} - \boldsymbol{B}^T\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}. \tag{24}$$

Let's consider the case $\boldsymbol{D} = \sigma^2\boldsymbol{I}$ so $\operatorname{var}(\hat{\sigma}^2) = \frac{1}{c - \boldsymbol{b}^T\boldsymbol{A}^{-1}\boldsymbol{b}}$. The difficulty of obtaining $\operatorname{var}(\hat{\sigma}^2)$ is from formulating $\boldsymbol{b}$, $\boldsymbol{A}$ and get the inverse of $\boldsymbol{A}$.

Since the Fisher information matrix is always positive semidefinite and we have assumed it's nonsingular, we have $\boldsymbol{b}^T\boldsymbol{A}^{-1}\boldsymbol{b} > 0$, and

$$\operatorname{var}(\hat{\sigma}^2) = \frac{1}{c - \boldsymbol{b}^T\boldsymbol{A}^{-1}\boldsymbol{b}} > \frac{1}{c}. \tag{25}$$

We can use $1/c$ as a lower bound of $\operatorname{var}(\hat{\sigma}^2)$ in practice(if not evaluating $\boldsymbol{A}$).

For a general $\boldsymbol{D}$, we are interested in the diagonal of $\operatorname{cov}(\hat{\boldsymbol{\sigma}}^2) = (\boldsymbol{C} - \boldsymbol{B}^T\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}$.

Lemma 2.1. *Let $\boldsymbol{X}$ be a positive definite matrix, then $(\boldsymbol{X}^{-1})_{jj} \geq \frac{1}{\boldsymbol{X}_{jj}}$.*

Proof. The eigen-decompostion of $\boldsymbol{X}$ is $\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$. Then

$$\boldsymbol{X}_{jj} = \boldsymbol{q}_j^T \boldsymbol{\Lambda} \boldsymbol{q}_j = \sum_k q_{jk}^2 \lambda_k,$$

and

$$(\boldsymbol{X})_{jj}^{-1} = \boldsymbol{q}_j^T \boldsymbol{\Lambda}^{-1} \boldsymbol{q}_j = \sum_k q_{jk}^2 / \lambda_k.$$

□

By Cauchy–Schwarz inequality,

$$\boldsymbol{X}_{jj}(\boldsymbol{X})_{jj}^{-1} \geq (\sum_k q_{jk})^2 = 1.$$

Since by assumption $\boldsymbol{C} - \boldsymbol{B}^T\boldsymbol{A}^{-1}\boldsymbol{B}$ is positive definite, by lemma (2.1), we have

$$(\boldsymbol{C} - \boldsymbol{B}^T\boldsymbol{A}^{-1}\boldsymbol{B})_{rr}^{-1} \geq 1/(\boldsymbol{C} - \boldsymbol{B}^T\boldsymbol{A}^{-1}\boldsymbol{B})_{rr} = \frac{1}{\boldsymbol{C}_{rr} - \boldsymbol{b}_r^T\boldsymbol{A}^{-1}\boldsymbol{b}_r} \geq \frac{1}{\boldsymbol{C}_{rr}}. \tag{26}$$

We summarize the results below.

1. When $\boldsymbol{V}_i = \boldsymbol{I}$,
   - if $\boldsymbol{D} = \sigma^2\boldsymbol{I}$, $\mathrm{var}(\hat{\sigma}^2) \geq \frac{2}{N\,\mathrm{Tr}((\boldsymbol{U}+\boldsymbol{I})^{-1}(\boldsymbol{U}+\boldsymbol{I})^{-1})}$.
   - if $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{\sigma}^2)$, $\mathrm{var}(\hat{\sigma}_r^2) \geq \frac{2}{N((\boldsymbol{U}+\boldsymbol{I})_{rr}^{-1})^2}$.
2. When $\boldsymbol{V}_i = \boldsymbol{V}$,
   - if $\boldsymbol{D} = \sigma^2\boldsymbol{I}$, $\mathrm{var}(\hat{\sigma}^2) \geq \frac{2}{N\,\mathrm{Tr}((\boldsymbol{U}+\boldsymbol{V})^{-1}(\boldsymbol{U}+\boldsymbol{V})^{-1})}$.
   - if $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{\sigma}^2)$, $\mathrm{var}(\hat{\sigma}_r^2) \geq \frac{2}{N((\boldsymbol{U}+\boldsymbol{V})_{rr}^{-1})^2}$.
3. When $\boldsymbol{V}_i$ varies with samples,
   - if $\boldsymbol{D} = \sigma^2\boldsymbol{I}$, $\mathrm{var}(\hat{\sigma}^2) \geq \frac{2}{\sum_i \mathrm{Tr}((\boldsymbol{U}+\boldsymbol{V}_i)^{-1}(\boldsymbol{U}+\boldsymbol{V}_i)^{-1})}$.
   - if $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{\sigma}^2)$, $\mathrm{var}(\hat{\sigma}_r^2) \geq \frac{2}{\sum_i((\boldsymbol{U}+\boldsymbol{V}_i)_{rr}^{-1})^2}$.

Recall in one dimensional case $x_i \sim N(0, \sigma^2 + v_i^2)$, $\mathrm{var}(\hat{\sigma}^2) = 2(\sum_i w_i^2)^{-1}$ where $w_i = \frac{1}{v_i+\sigma^2}$. So when $\hat{\sigma}^2 = 0$, we have $\hat{\mathrm{var}}(\hat{\sigma}^2) = 2(\sum_i v_i^{-2})^{-1}$.

Now in multivariate case, if $\boldsymbol{V}_i$ is diagonal and $\boldsymbol{U} = 0$, then $\mathrm{var}(\hat{\sigma}_r^2) \geq 2(\sum_i \boldsymbol{V}_{i,rr}^{-2})^{-1}$. In this case the lower bound is tight.

3 When U is rank 1

Let $\boldsymbol{U} = \boldsymbol{u}\boldsymbol{u}^T$, $\boldsymbol{D} = \sigma^2 \boldsymbol{I}$ and $\boldsymbol{\theta} = (\boldsymbol{u}, \sigma^2)$, the first order derivatives are

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{u}} &= -(\boldsymbol{S}_i)^{-1}\boldsymbol{u} + \boldsymbol{x}_i^T(\boldsymbol{S}_i)^{-1}\boldsymbol{u}(\boldsymbol{S}_i)^{-1}\boldsymbol{x}, \\ \frac{\partial l(\boldsymbol{\theta})}{\partial \sigma^2} &= -\frac{1}{2}\operatorname{Tr}(\boldsymbol{S}_i^{-1}) + \frac{1}{2}\boldsymbol{x}_i^T\boldsymbol{S}_i^{-1}\boldsymbol{S}_i^{-1}\boldsymbol{x}_i. \end{aligned} \tag{27}$$

The second order derivatives are

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{u}\partial \boldsymbol{u}^T} =& \boldsymbol{u}^T\boldsymbol{S}_i^{-1}\boldsymbol{u}\boldsymbol{S}_i^{-1} + \boldsymbol{S}_i^{-1}\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{S}_i^{-1} - \boldsymbol{S}_i^{-1} \\ &+ \boldsymbol{S}_i^{-1}\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{S}_i^{-1} - (\boldsymbol{u}^T\boldsymbol{S}_i^{-1}\boldsymbol{u}\boldsymbol{S}_i^{-1}\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{S}_i^{-1} + \boldsymbol{x}^T\boldsymbol{S}_i^{-1}\boldsymbol{u}\boldsymbol{S}_i^{-1}\boldsymbol{x}\boldsymbol{u}^T\boldsymbol{S}_i^{-1}) \\ &- (\boldsymbol{x}^T\boldsymbol{S}_i^{-1}\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{S}_i^{-1}\boldsymbol{x}\boldsymbol{S}_i^{-1} + \boldsymbol{x}^T\boldsymbol{S}_i^{-1}\boldsymbol{u}\boldsymbol{S}_i^{-1}\boldsymbol{u}\boldsymbol{x}^T\boldsymbol{S}_i^{-1}). \end{aligned} \tag{28}$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial^2 \sigma^2} = \frac{1}{2}\operatorname{Tr}(\boldsymbol{S}_i^{-1}\boldsymbol{S}_i^{-1}) - \boldsymbol{x}_i^T\boldsymbol{S}_i^{-1}\boldsymbol{S}_i^{-1}\boldsymbol{S}_i^{-1}\boldsymbol{x}_i. \tag{29}$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{u}\partial \sigma^2} = \boldsymbol{S}_i^{-1}\boldsymbol{S}_i^{-1}\boldsymbol{u} - \boldsymbol{x}_i^T\boldsymbol{S}_i^{-1}\boldsymbol{S}_i^{-1}\boldsymbol{u}\boldsymbol{S}_i^{-1}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{S}_i^{-1}\boldsymbol{u}\boldsymbol{S}_i^{-1}\boldsymbol{S}_i^{-1}\boldsymbol{x}. \tag{30}$$

Taking expectation of second order derivatives gives

$$\begin{aligned} \mathbb{E}(\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{u}\partial \boldsymbol{u}^T}) &= -\boldsymbol{S}_i^{-1}\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{S}_i^{-1} - \boldsymbol{u}^T\boldsymbol{S}_i^{-1}\boldsymbol{u}\boldsymbol{S}_i^{-1}, \\ \mathbb{E}(\frac{\partial^2 l(\boldsymbol{\theta})}{\partial^2 \sigma^2}) &= -\frac{1}{2}\operatorname{Tr}(\boldsymbol{S}_i^{-1}\boldsymbol{S}_i^{-1}), \\ \mathbb{E}(\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{u}\partial \sigma^2}) &= -\boldsymbol{S}_i^{-1}\boldsymbol{S}_i^{-1}\boldsymbol{u}. \end{aligned} \tag{31}$$

The variance of $\hat{\sigma}^2$ is given by the inverse of

$$\frac{1}{2}\sum_i \operatorname{Tr}(\boldsymbol{S}_i^{-1}\boldsymbol{S}_i^{-1}) - \boldsymbol{u}^T\sum_i(\boldsymbol{S}_i^{-1}\boldsymbol{S}_i^{-1})\left(\sum_i(\boldsymbol{S}_i^{-1}\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{S}_i^{-1} + \boldsymbol{u}^T\boldsymbol{S}_i^{-1}\boldsymbol{u}\boldsymbol{S}_i^{-1})\right)^{-1}\sum_i(\boldsymbol{S}_i^{-1}\boldsymbol{S}_i^{-1})\boldsymbol{u}. \tag{32}$$

When $\boldsymbol{S}_i = \boldsymbol{S}$,

$$(\boldsymbol{S}^{-1}\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{S}^{-1} + \boldsymbol{u}^T\boldsymbol{S}^{-1}\boldsymbol{u}\boldsymbol{S}^{-1})^{-1} = \frac{\boldsymbol{S}}{\boldsymbol{u}^T\boldsymbol{S}^{-1}\boldsymbol{u}} - \frac{\boldsymbol{u}\boldsymbol{u}^T}{2(\boldsymbol{u}^T\boldsymbol{S}^{-1}\boldsymbol{u})^2}. \tag{33}$$

Let $\boldsymbol{S} = \boldsymbol{u}\boldsymbol{u}^T + \boldsymbol{V}$, i.e. $\hat{\sigma}^2 = 0$, then $\boldsymbol{S}^{-1} = \boldsymbol{V}^{-1} - \frac{\boldsymbol{V}^{-1}\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{V}^{-1}}{1+\boldsymbol{u}^T\boldsymbol{V}^{-1}\boldsymbol{u}}$, and

$$\boldsymbol{S}^{-1}\boldsymbol{S}^{-1} = (\boldsymbol{V}^{-1} - \frac{\boldsymbol{V}^{-1}\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{V}^{-1}}{1+\boldsymbol{u}^T\boldsymbol{V}^{-1}\boldsymbol{u}})(\boldsymbol{V}^{-1} - \frac{\boldsymbol{V}^{-1}\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{V}^{-1}}{1+\boldsymbol{u}^T\boldsymbol{V}^{-1}\boldsymbol{u}}) \tag{34}$$

4 Fisher scoring for weighted likelihood

Consider estimating $\boldsymbol{\theta} = \{\boldsymbol{U}, \boldsymbol{D}\}$ in the model

$$\boldsymbol{x}_i \sim N(\boldsymbol{0}, \boldsymbol{U} + \boldsymbol{D} + \boldsymbol{V}_i). \tag{35}$$

Let w_i be the weight of sample i, and $\sum_i w_i = 1$, then the weighted log-likelihood is

$$l(\boldsymbol{\theta}) = \sum_i \gamma_i(-\frac{1}{2}\log|\boldsymbol{U} + \boldsymbol{D} + \boldsymbol{V}_i| - \frac{1}{2}\boldsymbol{x}_i^T(\boldsymbol{U} + \boldsymbol{D} + \boldsymbol{V}_i)^{-1}\boldsymbol{x}_i). \tag{36}$$

Denote the score and Fisher information as $\boldsymbol{s}(\boldsymbol{\theta})$ and $\boldsymbol{I}(\boldsymbol{\theta})$ respectively, then the Fisher scoring interation is

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} + \boldsymbol{I}(\boldsymbol{\theta}^{(t)})^{-1}\boldsymbol{s}(\boldsymbol{\theta}^{(t)}). \tag{37}$$

This can be applied to the M-step in multivariate deconvolution problem.

5 Appendix

Matrix calculus: let $\boldsymbol{X}$ be a nonsingular matrix.

$$\partial a\boldsymbol{X} = a\partial\boldsymbol{X}.$$

$$\partial \operatorname{Tr}(\boldsymbol{X}) = \operatorname{Tr}(\partial\boldsymbol{X}).$$

$$\partial \log|\boldsymbol{X}| = \operatorname{Tr}(\boldsymbol{X}^{-1}\partial\boldsymbol{X}).$$

$$\partial\boldsymbol{X}^{-1} = -\boldsymbol{X}^{-1}\partial\boldsymbol{X}\partial\boldsymbol{X}^{-1}.$$

$$\partial(\boldsymbol{X}\boldsymbol{Y}) = \partial(\boldsymbol{X})\boldsymbol{Y} + \boldsymbol{X}\partial(\boldsymbol{Y}).$$

Lemma 5.1. *Let $\boldsymbol{X}$ be a positive defnite matrix, and $\boldsymbol{\Omega}$ be a diagonal matrix, then the solution to*

$$\min_{\boldsymbol{\Omega}} ||\boldsymbol{X}\boldsymbol{\Omega} - \boldsymbol{I}||_F^2$$

is given by

$$\boldsymbol{\Omega}_{jj} = \frac{\boldsymbol{X}_{jj}}{||\boldsymbol{x}_j||^2}.$$