Biomath 210 Homework 5

Huwenbo Shi (603-778-363) shihuwenbo@ucla.edu

November 23, 2015

Problem 6.3

Let $\pi_i(\boldsymbol{\theta}) = \frac{\exp(\mathbf{x}_i^* \boldsymbol{\theta})}{1+\exp(\mathbf{x}_i^* \boldsymbol{\theta})}$ be the success probability of the i-th trial. Then the likelihood of the data $\mathbf{x}_1, \ldots, \mathbf{x}_m$ is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{m} \left[\frac{\exp(\mathbf{x}_i^* \boldsymbol{\theta})}{1 + \exp(\mathbf{x}_i^* \boldsymbol{\theta})} \right]^{y_i} \left[\frac{1}{1 + \exp(\mathbf{x}_i^* \boldsymbol{\theta})} \right]^{(1-y_i)}, \tag{1}$$

which has the log-likelihood

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{m} [y_i \mathbf{x}_i^* \boldsymbol{\theta} - \ln(1 + \exp(\mathbf{x}_i^* \boldsymbol{\theta}))]. \tag{2}$$

Let $f_i(\boldsymbol{\theta}) = y_i \mathbf{x}_i^* \boldsymbol{\theta} - \ln(1 + \exp(\mathbf{x}_i^* \boldsymbol{\theta}))$, then

$$\nabla f_i(\boldsymbol{\theta}) = y_i \mathbf{x}_i - \frac{\exp(\mathbf{x}_i^* \boldsymbol{\theta})}{1 + \exp(\mathbf{x}_i^* \boldsymbol{\theta})} \mathbf{x}_i = [y_i - \pi_i(\boldsymbol{\theta})] \mathbf{x}_i. \tag{3}$$

Therefore, the score vector is

$$\nabla \ell(\boldsymbol{\theta}) = \sum_{i=1}^{m} [y_i - \pi_i(\boldsymbol{\theta})] \mathbf{x}_i. \tag{4}$$

To derive the observed information matrix, we find

$$\frac{\partial}{\partial \boldsymbol{\theta}_j} f_i(\boldsymbol{\theta}) = y_i x_{ij} - \frac{\exp(\mathbf{x}_i^* \boldsymbol{\theta})}{1 + \exp(\mathbf{x}_i^* \boldsymbol{\theta})} x_{ij}, \tag{5}$$

and

$$\begin{aligned} \frac{\partial^2}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k} f_i(\boldsymbol{\theta}) &= - \left[\frac{(1 + \exp(\mathbf{x}_i^* \boldsymbol{\theta})) \exp(\mathbf{x}_i^* \boldsymbol{\theta}) x_{ik} - \exp(\mathbf{x}_i^* \boldsymbol{\theta}) \exp(\mathbf{x}^* \boldsymbol{\theta}) x_{ik}}{(1 + \exp(\mathbf{x}_i^* \boldsymbol{\theta}))^2} \right] x_{ij} \\ &= - \frac{\exp(\mathbf{x}_i^* \boldsymbol{\theta})}{1 + \exp(\mathbf{x}_i^* \boldsymbol{\theta})} \left[x_{ik} - \frac{\exp(\mathbf{x}^* \boldsymbol{\theta})}{1 + \exp(\mathbf{x}_i^* \boldsymbol{\theta})} x_{ik} \right] x_{ij} \\ &= -\pi_i[\boldsymbol{\theta})(1 - \pi_i(\boldsymbol{\theta})] x_{ij} x_{ik}. \end{aligned} \tag{6}$$

So, $\partial^2 f_i(\boldsymbol{\theta}) = -\pi_i(\boldsymbol{\theta})[1 - \pi_i(\boldsymbol{\theta})] \mathbf{x}_i \mathbf{x}_i^*$, and $\partial^2 \ell(\boldsymbol{\theta}) = -\sum_{i=1}^{m} \pi_i(\boldsymbol{\theta})[1 - \pi_i(\boldsymbol{\theta})] \mathbf{x}_i \mathbf{x}_i^*$. The observed information matrix is therefore, $-\partial^2 \ell(\boldsymbol{\theta}) = \sum_{i=1}^{m} \pi_i(\boldsymbol{\theta})[1 - \pi_i(\boldsymbol{\theta})] \mathbf{x}_i \mathbf{x}_i^*$.

Problem 6.6 (?)

First, we derive the majorization for $\sqrt{u^2+\epsilon}-\sqrt{\epsilon}$. Let $f(v)=\sqrt{v+\epsilon}-\sqrt{\epsilon}$. Because $f(v)$ is a concave function, it satisfies the supporting hyperplane inequality

$$\sqrt{v+\epsilon}-\sqrt{\epsilon} \leqslant \sqrt{w+\epsilon}-\sqrt{\epsilon}+\frac{1}{2\sqrt{w+\epsilon}}(v-w). \tag{7}$$

Set $v=u^2$ and $w=u_n^2$, we have

$$\sqrt{u^2+\epsilon}-\sqrt{\epsilon} \leqslant \sqrt{u_n^2+\epsilon}-\sqrt{\epsilon}+\frac{1}{2\sqrt{u_n^2+\epsilon}}(u^2-u_n^2). \tag{8}$$

Equality holds when $u=u_n$.

Next, we derive the majorization for $\sqrt{u_+^2+\epsilon}-\sqrt{\epsilon}$. For $u_n=0$, we have the inequality $\sqrt{u_+^2+\epsilon}-\sqrt{\epsilon} \leqslant \frac{1}{2\sqrt{\epsilon}}u_+^2$ by substituting $v=u_+^2$ and $w=0$ in (7). Since $u^2 \geqslant u_+^2$, we have $\sqrt{u_+^2+\epsilon}-\sqrt{\epsilon} \leqslant \frac{1}{2\sqrt{\epsilon}}u^2$. Clearly, equality holds when $u=u_n=0$.

Problem 6.8

First, we find the optimal $\mathbf{D}$ by minimizing $\|\mathbf{S}-\mathbf{F}\mathbf{F}^*-\mathbf{D}\|_F^2$ when $\mathbf{F}$ is fixed. We notice that

$$\begin{aligned}
&\|\mathbf{S}-\mathbf{F}\mathbf{F}^*-\mathbf{D}\|_F^2 = \mathrm{tr}[(\mathbf{S}-\mathbf{F}\mathbf{F}^*-\mathbf{D})^*(\mathbf{S}-\mathbf{F}\mathbf{F}^*-\mathbf{D})] \\
&= \mathrm{tr}(\mathbf{S}\mathbf{S}) - 2\mathrm{tr}(\mathbf{S}\mathbf{F}\mathbf{F}^*) - 2\mathrm{tr}(\mathbf{S}\mathbf{D}) + 2\mathrm{tr}(\mathbf{D}\mathbf{F}\mathbf{F}^*) + \mathrm{tr}(\mathbf{D}\mathbf{D}) + \mathrm{tr}(\mathbf{F}\mathbf{F}^*\mathbf{F}\mathbf{F}^*) \\
&= -2\mathrm{tr}(\mathbf{S}\mathbf{D}) + 2\mathrm{tr}(\mathbf{D}\mathbf{F}\mathbf{F}^*) + \mathrm{tr}(\mathbf{D}\mathbf{D}) + c \\
&= -2\sum_i \mathbf{S}_{ii}\mathbf{d}_i + 2\sum_i (\mathbf{F}\mathbf{F}^*)_{ii}\mathbf{d}_i + \mathbf{d}^*\mathbf{d} + c,
\end{aligned} \tag{9}$$

where c is a constant and $\mathbf{D}=\mathrm{diag}(\mathbf{d})$. Setting each component of the gradient $(-2\mathbf{S}_{ii}+2(\mathbf{F}\mathbf{F}^*)_{ii}+2\mathbf{d}_i)$ to 0, we get $\mathbf{d}_i=\mathbf{S}_{ii}-(\mathbf{F}\mathbf{F}^*)_{ii}$. Therefore, when $\mathbf{F}$ is fixed, the optimal diagonal matrix $\mathbf{D}$ satisfies $\mathbf{D}_{ii}=\mathbf{S}_{ii}-(\mathbf{F}\mathbf{F}^*)_{ii}$.

Next, we fix $\mathbf{D}$ and minimize the objective over $\mathbf{F}$. Let $\mathbf{N}=\mathbf{F}\mathbf{F}^*$, we first minimize the objective $f(\mathbf{N})=-2\mathrm{tr}(\mathbf{S}\mathbf{N})+2\mathrm{tr}(\mathbf{D}\mathbf{N})+\mathrm{tr}(\mathbf{N}\mathbf{N})+e$, where e is a constant, over $\mathbf{N}$ with the constraint that $\mathbf{N}$ is positive semi-definite and $\mathrm{rank}(\mathbf{N}) \leqslant r$. This problem is equivalent to finding the best rank r approximation of $\mathbf{S}-\mathbf{D}$ because the objective can be written as $f(\mathbf{N})=\mathrm{tr}(\mathbf{N}\mathbf{N})-2\mathrm{tr}[(\mathbf{S}-\mathbf{D})\mathbf{N}]+\mathrm{tr}[(\mathbf{S}-\mathbf{D})^*(\mathbf{S}-\mathbf{D})]+b=\|\mathbf{N}-(\mathbf{S}-\mathbf{D})\|_F^2+b$, where b is a constant. From **Proposition 7.2.3**, an analytical solution for $\mathbf{N}$ is

$$\mathbf{N}=\mathbf{F}\mathbf{F}^*=\sum_{i=1}^{r}\max\{\sigma_i,0\}\mathbf{u}_i\mathbf{u}_i^*, \tag{10}$$

where σ_i and $\mathbf{u}_i$ are the eigenvalues and eigenvectors of the ordered spectral decomposition $\mathbf{S}-\mathbf{D}=\mathbf{U}\mathbf{\Sigma}\mathbf{U}^*$. Solution $\mathbf{F}$ to the equality in Equation (10) is not unique. One can set $\mathbf{F}=\sum_{i=1}^{r}\sqrt{\max\{\sigma_i,0\}}\mathbf{u}_i\mathbf{u}_i^*$ Alternatively, one can set the i-th column of the first r columns of $\mathbf{F}$ to be $\sqrt{\max\{\sigma_i,0\}}\mathbf{u}_i$ and $\mathbf{0}$ for the rest. Both solutions satisfy the equality in Equation (10).

Problem 6.13

We first write the likelihood of the data

$$L(\mathbf{M}, \mathbf{U}, \mathbf{V}) = \prod_{i=1}^{k} \frac{\exp[-\frac{1}{2}\mathrm{tr}(\mathbf{V}^{-1}(\mathbf{X}_i - \mathbf{M})^*\mathbf{U}^{-1}(\mathbf{X}_i - \mathbf{M}))]}{(2\pi)^{\frac{np}{2}} \det(\mathbf{V})^{\frac{n}{2}} \det(\mathbf{U})^{\frac{p}{2}}}, \tag{11}$$

with log-likelihood

$$\begin{aligned}\ell(\mathbf{M}, \mathbf{U}, \mathbf{V}) &= \sum_{i=1}^{k} \left\{ -\frac{1}{2}\mathrm{tr}[\mathbf{V}^{-1}(\mathbf{X}_i - \mathbf{M})^*\mathbf{U}^{-1}(\mathbf{X}_i - \mathbf{M})] \right\} \\ &\quad - k\ln[(2\pi)^{\frac{np}{2}}] - k\frac{n}{2}\ln\det(\mathbf{V}) - k\frac{p}{2}\ln\det(\mathbf{U}) \\ &= \sum_{i=1}^{k} \left[-\frac{1}{2}\mathrm{tr}(\mathbf{V}^{-1}\mathbf{X}_i^*\mathbf{U}^{-1}\mathbf{X}_i) + \mathrm{tr}(\mathbf{V}^{-1}\mathbf{X}_i^*\mathbf{U}^{-1}\mathbf{M}) - \frac{1}{2}\mathrm{tr}(\mathbf{V}^{-1}\mathbf{M}^*\mathbf{U}^{-1}\mathbf{M}) \right] \\ &\quad - k\ln[(2\pi)^{\frac{np}{2}}] - k\frac{n}{2}\ln\det(\mathbf{V}) - k\frac{p}{2}\ln\det(\mathbf{U}).\end{aligned} \tag{12}$$

We first estimate the mean $\mathbf{M}$. Taking the derivative with respect to $\mathbf{M}$ and setting it to zero, we get

$$0 = \sum_{i=1}^{k} \mathbf{V}^{-1}\mathbf{X}_i^*\mathbf{U}^{-1} - \mathbf{V}\mathbf{M}\mathbf{U}^{-1} = \mathbf{V}^{-1}\left(\sum_{i=1}^{k}\mathbf{X}_i - k\mathbf{M}\right)\mathbf{U}^{-1} \tag{13}$$

Solving for $\mathbf{M}$, we get $\mathbf{M} = \frac{1}{k}\sum_{i=1}^{k}\mathbf{X}_i$.

Next, we fix $\mathbf{V}$, and maximize the log-likelihood over $\mathbf{U}$. Taking the derivative with respect to $\mathbf{U}^{-1}$ and setting it to zero, we get

$$0 = -\sum_{i=1}^{k}\left(\frac{1}{2}\mathbf{X}_i\mathbf{V}^{-1}\mathbf{X}_i^* - \mathbf{M}\mathbf{V}^{-1}\mathbf{X}_i^* + \frac{1}{2}\mathbf{M}\mathbf{V}^{-1}\mathbf{M}\right) - \frac{kp}{2}\mathbf{U} \tag{14}$$

Solving for $\mathbf{U}$, we get $\mathbf{U} = \frac{1}{kp}\sum_{i=1}^{k}(\mathbf{X}_i - \mathbf{M})^*\mathbf{V}^{-1}(\mathbf{X}_i - \mathbf{M})$.

Finally, we fix $\mathbf{U}$, and maximize the log-likelihood over $\mathbf{V}$. Similarly, we set the derivative with respect to $\mathbf{V}^{-1}$ to 0, and get

$$0 = -\sum_{i=1}^{k}\left(\frac{1}{2}\mathbf{X}_i\mathbf{U}^{-1}\mathbf{X}_i^* - \mathbf{M}\mathbf{U}^{-1}\mathbf{X}_i^* + \frac{1}{2}\mathbf{M}\mathbf{U}^{-1}\mathbf{M}\right) - \frac{kn}{2}\mathbf{V} \tag{15}$$

Solving for $\mathbf{V}$, we get $\mathbf{V} = \frac{1}{kn}\sum_{i=1}^{k}(\mathbf{X}_i - \mathbf{M})^*\mathbf{U}^{-1}(\mathbf{X}_i - \mathbf{M})$.

Problem 6.14

For the mixture of r Gaussian distributions, the density of $\mathbf{y}$ is $\Pr(\mathbf{y}) = \sum_{j=1}^{r}\pi_j f(\mathbf{y}|\boldsymbol{\mu}_j, \boldsymbol{\Omega})$. The likelihood is therefore,

$$L(\boldsymbol{\Theta}) = \prod_{i=1}^{m}\left[\sum_{j=1}^{r}\pi_j f(\mathbf{y}_i|\boldsymbol{\mu}_j, \boldsymbol{\Omega})\right], \tag{16}$$

with log-likelihood

$$\begin{aligned}\ell(\mathbf{\Theta}) &= \sum_{i=1}^{m} \ln\left[\sum_{j=1}^{r} \pi_j f(\mathbf{y}_i|\boldsymbol{\mu}_j, \mathbf{\Omega})\right] \\ &= \sum_{i=1}^{m} \ln\left\{\sum_{j=1}^{r} \pi_j \frac{1}{(2\pi)^{\frac{p}{2}} \det(\mathbf{\Omega})^{\frac{1}{2}}} \exp[-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu}_j)^* \mathbf{\Omega}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_j)]\right\},\end{aligned} \tag{17}$$

where $\mathbf{\Theta} = (\boldsymbol{\mu}_j, \mathbf{\Omega})$ represents the parameters. We maximize the log-likelihood with the constraint $\sum_{j=1}^{r} \pi_j = 1$, which entails the Lagrangian,

$$\sum_{i=1}^{m} \ln\left\{\sum_{j=1}^{r} \pi_j \frac{1}{(2\pi)^{\frac{p}{2}} \det(\mathbf{\Omega})^{\frac{1}{2}}} \exp[-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu}_j)^* \mathbf{\Omega}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_j)]\right\} + \lambda \sum_{j=1}^{r} \pi_j - \lambda. \tag{18}$$

Following Example (1.3), we obtain the minorization

$$\begin{aligned}&\sum_{i=1}^{m} \ln\left[\sum_{j=1}^{r} \pi_j f(\mathbf{y}_i|\boldsymbol{\mu}_j, \mathbf{\Omega})\right] + \lambda \sum_{j=1}^{r} \pi_j - \lambda \\ &\geqslant \sum_{i=1}^{m}\sum_{j=1}^{r} w_{nij} \ln\left\{\frac{1}{w_{nij}} \pi_j \frac{1}{(2\pi)^{\frac{p}{2}} \det(\mathbf{\Omega})^{\frac{1}{2}}} \exp[-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu}_j)^* \mathbf{\Omega}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_j)]\right\} + \lambda \sum_{j=1}^{r} \pi_j - \lambda \\ &= \sum_{i=1}^{m}\sum_{j=1}^{r} w_{nij} \left[\ln \pi_j - \frac{1}{2}\ln\det(\mathbf{\Omega}) - \frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu}_j)^* \mathbf{\Omega}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_j)\right] + \lambda \sum_{j=1}^{r} \pi_j - \lambda + c_n,\end{aligned} \tag{19}$$

where c_n is a constant.

Taking derivative with respect to π_j, and setting it to 0, we get

$$0 = \frac{1}{\pi_{n+1,j}} \sum_{i=1}^{m} w_{nij} + \lambda, \text{ and } \pi_{n+1,j} = -\frac{1}{\lambda}\sum_{i=1}^{m} w_{nij}. \tag{20}$$

Summing over j we get $\sum_{i=1}^{r} \pi_j = 1 = -\frac{1}{\lambda}\sum_{i=1}^{m}\sum_{j=1}^{r} w_{nij} = -\frac{m}{\lambda}$, which entails $\lambda = -m$ and that $\pi_{n+1,j} = \frac{1}{m}\sum_{i=1}^{m} w_{nij}$.

We update $\boldsymbol{\mu}_j$ and $\mathbf{\Omega}$ through block descent. The objective in Equation (19) can be written as

$$\begin{aligned}&\sum_{i=1}^{m}\sum_{j=1}^{r} w_{nij} \ln \pi_j - \frac{m}{2}\ln\det(\mathbf{\Omega}) + \lambda \sum_{j=1}^{r} \pi_j - \lambda + c_n \\ &- \frac{1}{2}\sum_{j=1}^{r}\left[(\sum_{i=1}^{m} w_{nij}\mathbf{y}_i - \boldsymbol{\mu}_j \sum_{i=1}^{m} w_{nij})^* \mathbf{\Omega}^{-1}(\sum_{i=1}^{m} w_{nij}\mathbf{y}_i - \boldsymbol{\mu}_j \sum_{i=1}^{m} w_{nij})\right].\end{aligned} \tag{21}$$

The last term of the above equation entails the update for $\boldsymbol{\mu}_{n+1,j}$,

$$\boldsymbol{\mu}_{n+1,j} = \frac{1}{\sum_{i=1}^{m} w_{nij}} \sum_{i=1}^{m} w_{nij}\mathbf{y}_i, \tag{22}$$

at which the last term of Equation (21) attains its minimum 0.

Next, we fix $\boldsymbol{\mu}_j = \boldsymbol{\mu}_{n+1,j}$. Taking the derivative with respect to $\mathbf{\Omega}^{-1}$, and setting it to 0 yields

$$0 = \frac{m}{2}\mathbf{\Omega} - \frac{1}{2}\sum_{j=1}^{r}\sum_{i=1}^{m} w_{nij}(\mathbf{y}_i - \boldsymbol{\mu}_{n+1,j})(\mathbf{y}_i - \boldsymbol{\mu}_{n+1,j})^*, \tag{23}$$

which entails the update

$$\mathbf{\Omega}_{n+1,j} = \frac{1}{m}\sum_{j=1}^{r}\sum_{i=1}^{m} w_{nij}(\mathbf{y}_i - \boldsymbol{\mu}_{n+1,j})(\mathbf{y}_i - \boldsymbol{\mu}_{n+1,j})^*. \tag{24}$$

Problem 6.15

Imposing the inverse Wishart prior, we obtain the log-likelihood

$$\begin{aligned}\ell(\mathbf{\Theta}) = &\sum_{i=1}^{m} \ln\left\{\sum_{j=1}^{r}\pi_j \frac{1}{(2\pi)^{\frac{p}{2}}\det(\mathbf{\Omega}_j)^{\frac{1}{2}}}\exp[-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu}_j)^*\mathbf{\Omega}_j^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_j)]\right\} \\ &- \sum_{j=1}^{r}\left[\frac{a}{2}\ln\det\mathbf{\Omega}_j + \frac{b}{2}\mathrm{tr}(\mathbf{\Omega}_j^{-1}\mathbf{S}_j)\right].\end{aligned} \tag{25}$$

With the constraint $\sum_{j=1}^{r}\pi_j = 1$, the Lagrangian becomes

$$\begin{aligned}&\sum_{i=1}^{m} \ln\left\{\sum_{j=1}^{r}\pi_j \frac{1}{(2\pi)^{\frac{p}{2}}\det(\mathbf{\Omega}_j)^{\frac{1}{2}}}\exp[-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu}_j)^*\mathbf{\Omega}_j^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_j)]\right\} \\ &- \sum_{j=1}^{r}\left[\frac{a}{2}\ln\det\mathbf{\Omega}_j + \frac{b}{2}\mathrm{tr}(\mathbf{\Omega}_j^{-1}\mathbf{S}_j)\right] + \lambda\sum_{j=1}^{r}\pi_j - \lambda.\end{aligned} \tag{26}$$

Similarly, we have the minorization

$$\begin{aligned}&\sum_{i=1}^{m}\sum_{j=1}^{r} w_{nij}\left[\ln\pi_j - \frac{1}{2}\ln\det(\mathbf{\Omega}_j) - \frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu}_j)^*\mathbf{\Omega}_j^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_j)\right] \\ &- \sum_{j=1}^{r}\left[\frac{a}{2}\ln\det\mathbf{\Omega}_j + \frac{b}{2}\mathrm{tr}(\mathbf{\Omega}_j^{-1}\mathbf{S}_j)\right] + \lambda\sum_{j=1}^{r}\pi_j - \lambda + c_n\end{aligned} \tag{27}$$

The derivative of the surrogate function with respect to π_j doesn't change, so

$$0 = \frac{1}{\pi_{n+1,j}}\sum_{i=1}^{m} w_{nij} + \lambda, \text{ and } \pi_{n+1,j} = -\frac{1}{\lambda}\sum_{i=1}^{m} w_{nij}. \tag{28}$$

Similar to Equation (21), the Lagrangian can also be written as

$$\sum_{i=1}^{m}\sum_{j=1}^{r} w_{nij}\ln\pi_j - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{r} w_{nij}\ln\det(\mathbf{\Omega}_j) + \lambda\sum_{j=1}^{r}\pi_j - \lambda + c_n - \sum_{j=1}^{r}\left[\frac{a}{2}\ln\det\mathbf{\Omega}_j + \frac{b}{2}\mathrm{tr}(\mathbf{\Omega}_j^{-1}\mathbf{S}_j)\right]$$
$$-\frac{1}{2}\sum_{j=1}^{r}\left[(\sum_{i=1}^{m} w_{nij}\mathbf{y}_i - \boldsymbol{\mu}_j\sum_{i=1}^{m} w_{nij})^*\mathbf{\Omega}_j^{-1}(\sum_{i=1}^{m} w_{nij}\mathbf{y}_i - \boldsymbol{\mu}_j\sum_{i=1}^{m} w_{nij})\right]. \tag{29}$$

The last term of the equation above entails the same update for $\boldsymbol{\mu}_j$,

$$\boldsymbol{\mu}_{n+1,j} = \frac{1}{\sum_{i=1}^{m} w_{nij}}\sum_{i=1}^{m} w_{nij}\mathbf{y}_i, \tag{30}$$

Taking $\mathbf{S}_j = \mathbf{S}$ for all j and $\boldsymbol{\mu}_j = \boldsymbol{\mu}_{n+1,j}$, and setting the derivative of the Lagrangian with respect to $\mathbf{\Omega}_j^{-1}$ to 0, we get

$$0 = \frac{1}{2}\sum_{i=1}^{m} w_{nij}\mathbf{\Omega}_j + \frac{a}{2}\mathbf{\Omega}_j - \frac{b}{2}\mathbf{S}_j - \frac{1}{2}\sum_{i=1}^{m} w_{nij}(\mathbf{y}_i - \boldsymbol{\mu}_{n+1,j})(\mathbf{y}_i - \boldsymbol{\mu}_{n+1,j})^*. \tag{31}$$

Solving for $\mathbf{\Omega}_j$, we obtain

$$\begin{aligned}\mathbf{\Omega}_j &= \frac{b\mathbf{S}}{\sum_{i=1}^{m} w_{nij} + a} + \frac{\sum_{i=1}^{m} w_{nij}(\mathbf{y}_i - \boldsymbol{\mu}_{n+1,j})(\mathbf{y}_i - \boldsymbol{\mu}_{n+1,j})^*}{\sum_{i=1}^{m} w_{nij} + a} \\ &= \frac{a}{\sum_{i=1}^{m} w_{nij} + a}\left(\frac{b}{a}\mathbf{S}\right) + \frac{\sum_{i=1}^{m} w_{nij}}{\sum_{i=1}^{m} w_{nij} + a}\tilde{\mathbf{\Omega}}_{n+1,j},\end{aligned} \tag{32}$$

where

$$\tilde{\mathbf{\Omega}}_{n+1,j} = \frac{1}{\sum_{i=1}^{m} w_{nij}}\sum_{i=1}^{m} w_{nij}(\mathbf{y}_i - \boldsymbol{\mu}_{n+1,j})(\mathbf{y}_i - \boldsymbol{\mu}_{n+1,j})^*. \tag{33}$$

Problem 6.16

We consider the model

$$\mathbf{y} \sim N\left(\begin{pmatrix}\mathbf{x}_1^* \\ \vdots \\ \mathbf{x}_m^*\end{pmatrix}\boldsymbol{\beta}, \sigma_1^2\mathbf{I} + \sigma_2^2\begin{pmatrix} t_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & t_m\end{pmatrix}\right). \tag{34}$$

To create simulations, I draw $\mathbf{x}_i \in \mathbf{R}^p$ and $\boldsymbol{\beta} \in \mathbf{R}^p$ randomly from the uniform distribution (i.e. the rand function in Matlab), and $\mathbf{y}$ from the normal distribution shown above. I set $p = 10$, $m = 100$, and $t_i = 0.01 * i$. I created 100 set of simulations in total, and I report the mean and standard errors of $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, and $\frac{\hat{\sigma}_1^2}{\hat{\sigma}_1^2+\hat{\sigma}_1^2}$. I also include a figure showing the convergence of the objective. Overall, I observe that the algorithm consistently underestimates σ_1. But the estimation for σ_2 and the ratio $\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}$ is less biased.

σ_1^2	$\hat{\sigma}_1^2$ (s.e.)	σ_2^2	$\hat{\sigma}_2^2$ (s.e.)	$\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}$	$\frac{\hat{\sigma}_1^2}{\hat{\sigma}_1^2+\hat{\sigma}_2^2}$ (s.e.)
2.0	1.579 (0.066)	2.0	2.320 (0.173)	0.500	0.472 (0.027)
2.0	1.537 (0.086)	4.0	4.360 (0.203)	0.333	0.295 (0.020)
2.0	1.559 (0.104)	6.0	6.197 (0.291)	0.250	0.246 (0.021)
2.0	1.591 (0.118)	8.0	8.327 (0.347)	0.200	0.198 (0.019)
4.0	3.286 (0.120)	2.0	2.648 (0.238)	0.667	0.611 (0.029)
6.0	4.977 (0.165)	2.0	2.680 (0.285)	0.750	0.700 (0.029)
8.0	6.329 (0.216)	2.0	3.660 (0.407)	0.800	0.692 (0.030)

Table 1: Simulated and estimated variance components

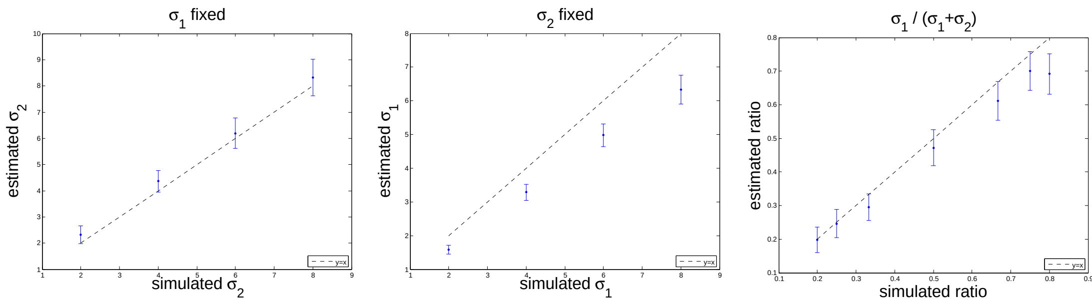

Figure 1: Estimated variance components vs. simulated variance components. Error bars show 2 × s.e. on each side.

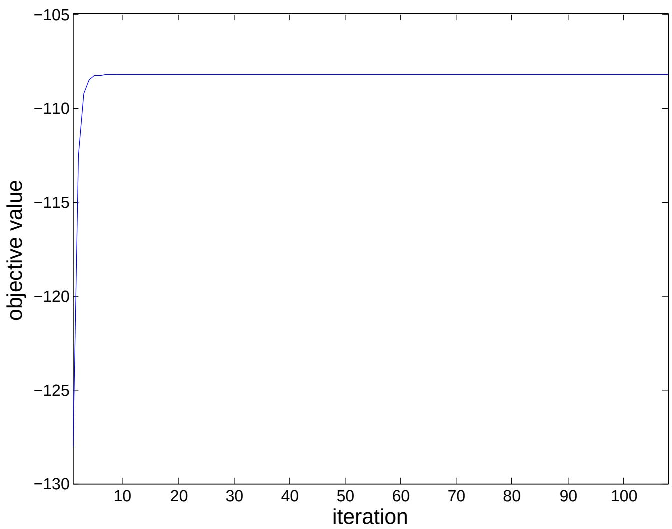

Figure 2: Convergence of the objective

lmm.m (implements the update 6.8 and 6.10)

```
function [beta, s1, s2, obj_val] = lmm(y, A, V1, V2)
    % initialization
    eps = 10^-4; max_iter = 1000;

    s1 = 1.0; s2 = 1.0;
    omega = s1*V1+s2*V2; omega_inv = pinv(omega);
    beta = pinv(A'*omega_inv*A)*A'*omega_inv*y;

    obj_val = zeros(max_iter, 1);

    % iterate until convergence
    for i=1:max_iter

        % compute objval
        obj_val(i) = -0.5*log(det(omega))...
            -0.5*(y-A*beta)'*omega_inv*(y-A*beta);

        % compute new sigma
        u = omega_inv*(y-A*beta);
        s1_next = s1*sqrt(u'*V1*u/trace(omega_inv*V1));
        s2_next = s2*sqrt(u'*V2*u/trace(omega_inv*V2));

        % compute new omega
        omega = s1_next*V1+s2_next*V2;
        omega_inv = pinv(omega);

        % compute new beta
        beta_next = pinv(A'*omega_inv*A)*A'*omega_inv*y;

        % check convergence
        dist = norm([s1 s2 beta']-[s1_next s2_next beta']);
        if(dist < eps), break; end

        % update parameters
        s1 = s1_next; s2 = s2_next; beta = beta_next;
    end

    obj_val = obj_val(1:i);
end
```

main.m (generates simulations and tests lmm.m)

```
% simulate data
n = 100;                    % number of simulations
m = 100;                    % number of samples
p = 10;                     % number of dimensions
sigma_1 = 2.0;              % simulated sigma_1
sigma_2 = 4.0;              % simulated sigma_2
sim_rec = zeros(n, 5);      % record simulation result

% simulate data and estimate sigma_1 and sigma_2
for i=1:n

    % simulate y
    beta = rand(p, 1);       % beta vector for simulation
    A  = rand(m, p);         % measurement matrix
    V1 = eye(m);             % variance component 1
    V2 = diag(1:m)/100;      % variance component 2 (time)
    y  =  mvnrnd(A*beta, sigma_1*V1+sigma_2*V2)';

    % esimate beta, sigma_1, and sigma_2
    [beta_est, s1_est, s2_est, obj_val] = lmm(y, A, V1, V2);
    if(i == 1)
        figure('visible', 'off'); plot(obj_val, 'b-');
        xlim([1 size(obj_val,1)]); xlabel('iteration');
        ylabel('objective value'); set(gca,'FontSize',12)
        print('prob_16_obj','-depsc','-r0');
    end

    % record result
    sim_rec(i,:) = [sigma_1 s1_est sigma_2 s2_est norm(beta-beta_est)];
end

% print summary
fprintf('s1: %.3f\n', sigma_1);
fprintf('estimated s1: %.3f %.3f\n', mean(sim_rec(:,2)), ...
    sqrt(var(sim_rec(:,2))/n));

fprintf('\n');

fprintf('s2: %.2f\n', sigma_2);
fprintf('estimated s2: %.3f %.3f\n', mean(sim_rec(:,4)), ...
    sqrt(var(sim_rec(:,4))/n));
```

```
fprintf('\n');

fprintf('s1/(s1+s2): %.3f\n', sigma_1/(sigma_1+sigma_2));
fprintf('estimated s1/(s1+s2): %.3f %.3f\n', ...
    mean(sim_rec(:,2)./(sim_rec(:,2)+sim_rec(:,4))),...
    sqrt(var(sim_rec(:,2)./(sim_rec(:,2)+sim_rec(:,4)))/n));
```

Problem 6.17

In this problem, we assume that $\mathbf{A}$ is a positive definite matrix. Let

$$f(\mathbf{x}, \mathbf{A}) = \sup_{\mathbf{y}} \left[\mathbf{x}^* \mathbf{y} - \frac{1}{2} \mathbf{y}^* \mathbf{A} \mathbf{y} \right]. \tag{35}$$

Since $\mathbf{A}$ is positive definite, the function inside the supremum is strictly convex. Based on Example 3.4.2, we know that the maximum is attained at $\mathbf{y} = \mathbf{A}^{-1}\mathbf{x}$, which yields $f(\mathbf{x}, \mathbf{A}) = \frac{1}{2}\mathbf{x}^*\mathbf{A}^{-1}\mathbf{x}$. Because the Fenchel conjugate is convex, we conclude that the map $(\mathbf{x}, \mathbf{A}) \to \frac{1}{2}\mathbf{x}^*\mathbf{A}^{-1}\mathbf{x}$ is convex.

Problem 6.19

First, we show that $\mathbf{v}_j$ lie on the surface of the unit sphere in $\mathbf{R}^{c-1}$. For $j = 1$, $\|\mathbf{v}_1\| = (c-1)^{-\frac{1}{2}}\|\mathbf{1}\| = (c-1)^{-\frac{1}{2}}(c-1)^{\frac{1}{2}} = 1$. For $2 \leqslant j \leqslant c$,

$$\begin{aligned} \|\mathbf{v}_j\| &= [(r\mathbf{1} + s\mathbf{e}_{j-1})^*(r\mathbf{1} + s\mathbf{e}_{j-1})]^{\frac{1}{2}} = [r^2(c-1) + s^2 + 2rs]^{\frac{1}{2}} \\ &= \left[\frac{(1+\sqrt{c})^2}{(c-1)^2} + \frac{c}{c-1} - 2\frac{\sqrt{c}}{\sqrt{c-1}} \frac{1+\sqrt{c}}{(c-1)^{\frac{3}{2}}} \right]^{\frac{1}{2}} \\ &= \left[\frac{c^2 - 2c + 1}{(c-1)^2} \right]^{\frac{1}{2}} = 1. \end{aligned} \tag{36}$$

Therefore, all $\mathbf{v}_j$ lie on the surface of the unit sphere in $\mathbf{R}^{c-1}$.

To prove equidistance, we first show that the distance between $\mathbf{v}_1$ and all other $\mathbf{v}_j$ is the same. We compute the distance between $\mathbf{v}_1$ and an arbitrary $\mathbf{v}_k$,

$$\begin{aligned} \|\mathbf{v}_1 - \mathbf{v}_k\|^2 &= \|\mathbf{v}_1\|^2 + \|\mathbf{v}_k\|^2 - 2\mathbf{v}_1^*\mathbf{v}_k = 2 - 2\mathbf{v}_1^*\mathbf{v}_k \\ &= 2 - 2[r(c-1)^{-\frac{1}{2}}(c-1) + (c-1)^{-\frac{1}{2}}s] \\ &= 2 - 2\left[-\frac{1+\sqrt{c}}{c-1} + \frac{\sqrt{c}}{c-1} \right] = 2 + \frac{2}{c-1}. \end{aligned} \tag{37}$$

Therefore, the distance between $\mathbf{v}_1$ and all other $\mathbf{v}_j$ is $\sqrt{2 + \frac{2}{c-1}}$. Next, we show that the distance between $\mathbf{v}_j$ and $\mathbf{v}_k$ for any arbitrary $c \geqslant j \geqslant 2$, $c \geqslant k \geqslant 2$, and $k \neq j$ are the same. We compute the distance between $\mathbf{v}_j$ and $\mathbf{v}_k$,

$$\begin{aligned}\|\mathbf{v}_j - \mathbf{v}_k\|^2 &= \|\mathbf{v}_j\|^2 + \|\mathbf{v}_k\|^2 - 2\mathbf{v}_j^*\mathbf{v}_k = 2 - 2\mathbf{v}_j^*\mathbf{v}_k \\ &= 2 - 2[r^2(c-2) + 2rs] \\ &= 2 - 2\left[\frac{1+c+2\sqrt{c}}{(c-1)^2} - \frac{2(1+\sqrt{c})\sqrt{c}}{(c-1)^2}\right] \\ &= 2 - 2\left[\frac{1-c}{(c-1)^2}\right] = 2 + \frac{2}{c-1}.\end{aligned} \tag{38}$$

Therefore, the distance between $\mathbf{v}_j$ and $\mathbf{v}_k$ are the same.

In conclusion, $\mathbf{v}_j$ lie on the surface of the unit sphere in $\mathbf{R}^{c-1}$, and that all vertex pairs are equidistant.

Problem 6.20

We show that for $\mathbf{R}^{c-1}$ the maximum number of pairwise equidistant points that can be situated is c. Let $\mathbf{x}_0, \cdots, \mathbf{x}_{c-1} \in \mathbf{R}^{c-1}$ be the set of pairwise equidistant points such that $\|\mathbf{x}_i - \mathbf{x}_j\| = d$ for all pairs of $i \neq j$. Without loss of generality, we subtract $\mathbf{x}_0$ from each $\mathbf{x}_i$, resulting in the new set of points $\mathbf{0}, \mathbf{x}_1 - \mathbf{x}_0, \cdots, \mathbf{x}_{c-1} - \mathbf{x}_0$. Since $\|\mathbf{x}_i - \mathbf{x}_j\| = \|(\mathbf{x}_i - \mathbf{x}_0) - (\mathbf{x}_j - \mathbf{x}_0)\| = d$, the new set of points are still pairwise equidistant. Let $\mathbf{y}_0 = \mathbf{0}, \cdots, \mathbf{y}_{c-1} = \mathbf{x}_{c-1} - \mathbf{x}_0$ denote the new set of points. For each $i \geqslant 1$, we have $\|\mathbf{y}_i\| = \|\mathbf{y}_i - \mathbf{0}\| = \sqrt{\mathbf{y}_i^*\mathbf{y}_i} = d$. And for each $i \neq j$, we have $\|\mathbf{y}_i - \mathbf{y}_j\| = d$. Since $\|\mathbf{y}_i - \mathbf{y}_j\|^2 = d^2 + d^2 - 2\mathbf{y}_i^*\mathbf{y}_j = d^2$ for $i \neq j$, we also have the equality $\mathbf{y}_i^*\mathbf{y}_j = \frac{d^2}{2}$. Now we show that the vectors $\mathbf{y}_1, \cdots, \mathbf{y}_{c-1}$ are linear independent. Let $a_1, \cdots, a_{c-1}$ be scalars such that $\sum_{i=1}^{c-1} a_i\mathbf{y}_i = 0$. Taking the inner product of $\mathbf{y}_j$ and $\sum_{i=1}^{c-1} a_i\mathbf{y}_i$ yields the equality $\mathbf{y}_j^*(\sum_{i=1}^{c-1} a_i\mathbf{y}_i) = \sum_{i=1}^{c-1} a_i\mathbf{y}_i^*\mathbf{y}_j = a_jd^2 + \sum_{i=1,i\neq j}^{c-1} a_i\frac{d^2}{2} = 0$. Since d is positive, the equality is satisfied only if $a_1 = \cdots = a_{c-1} = 0$. Therefore, the vectors $\mathbf{y}_1, \cdots, \mathbf{y}_{c-1}$ are linear independent and form a basis for $\mathbf{R}^{c-1}$. In other words, the problem of finding pairwise equidistant points in $\mathbf{R}^{c-1}$ is equivalent to finding the basis of $\mathbf{R}^{c-1}$ on the sphere $\{\mathbf{x} : \|x\| = d\}$ that satisfies $\|\mathbf{x}_i - \mathbf{x}_j\| = d$ for $i \neq j$ and then including the point $\mathbf{0}$, which can then be shifted and rotated. Since the number of vectors in the basis for $\mathbf{R}^{c-1}$ is $c-1$, the number of pairwise equidistant points that can be situated is c. And it's impossible to situate $c+1$ points in $\mathbf{R}^{c-1}$.