STAT 601 Homework 4

Yifan Zhu

February 23, 2018

1. Let $\boldsymbol{O}$ denote the observed data, $\boldsymbol{M}$ denote the missing data. In this case, let the observed pairs be $\boldsymbol{z}_i = \begin{bmatrix} x_i & y_i \end{bmatrix}^T$, $i = 1, \ldots, m$, the pairs when x is missing be $\boldsymbol{u}_j = \begin{bmatrix} x_j^m & y_j \end{bmatrix}^T$, $j = 1, \ldots, m_x$, the pairs when y is missing be $\boldsymbol{v}_k = \begin{bmatrix} x_k & y_k^m \end{bmatrix}^T$, $k = 1, \ldots, m_y$. And then the full log likelihood with paramter $\boldsymbol{\theta}$ is

$$\ell(\boldsymbol{O}, \boldsymbol{M}; \boldsymbol{\theta}) = \log f(\boldsymbol{O}, \boldsymbol{M}; \boldsymbol{\theta}) = \sum_{i=1}^{m} \log f(\boldsymbol{z}_i; \boldsymbol{\theta}) + \sum_{j=1}^{m_x} \log f(\boldsymbol{u}_j; \boldsymbol{\theta}) + \sum_{k=1}^{m_y} \log f(\boldsymbol{v}_k; \boldsymbol{\theta})$$

And we also know for $\boldsymbol{z}$ from bivariate normal with parameter $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, i.e. $\boldsymbol{z} \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we have

$$\begin{aligned} \ell(\boldsymbol{z}; \boldsymbol{\theta}) = \log f(\boldsymbol{z}; \boldsymbol{\theta}) &= \log \left(\frac{1}{2\pi |\boldsymbol{\Sigma}|^{1/2}} \exp \left[-\frac{1}{2} (\boldsymbol{z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{z} - \boldsymbol{\mu}) \right] \right) \\ &= -\log 2\pi - \frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} (\boldsymbol{z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{z} - \boldsymbol{\mu}) \end{aligned}$$

Hence we have

$$\begin{aligned} \ell(\boldsymbol{O}, \boldsymbol{M}; \boldsymbol{\theta}) = &-n \log 2\pi - \frac{n}{2} \log |\boldsymbol{\Sigma}| \\ &- \frac{1}{2} \sum_{i=1}^{m} (\boldsymbol{z}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{z}_i - \boldsymbol{\mu}) \\ &- \frac{1}{2} \sum_{j=1}^{m_x} (\boldsymbol{u}_j - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{u}_j - \boldsymbol{\mu}) \\ &- \frac{1}{2} \sum_{k=1}^{m_y} (\boldsymbol{v}_k - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{v}_k - \boldsymbol{\mu}) \end{aligned}$$

And because of independence, we have

$$\begin{aligned} &\mathrm{E}\left((\boldsymbol{z}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{z}_i - \boldsymbol{\mu}) \middle| \boldsymbol{O}, \boldsymbol{\theta}_p \right) = (\boldsymbol{z}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{z}_i - \boldsymbol{\mu}) \\ &\mathrm{E}\left((\boldsymbol{u}_j - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{u}_j - \boldsymbol{\mu}) \middle| \boldsymbol{O}, \boldsymbol{\theta}_p \right) = \mathrm{E}\left((\boldsymbol{u}_j - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{u}_j - \boldsymbol{\mu}) \middle| y_j; \boldsymbol{\theta}_p \right) \\ &\mathrm{E}\left((\boldsymbol{v}_k - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{v}_k - \boldsymbol{\mu}) \middle| \boldsymbol{O}, \boldsymbol{\theta}_p \right) = \mathrm{E}\left((\boldsymbol{v}_k - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{v}_k - \boldsymbol{\mu}) \middle| x_k; \boldsymbol{\theta}_p \right) \end{aligned}$$

Thus

$$
\begin{aligned}
Q(\boldsymbol{\theta};\boldsymbol{\theta}_p) &= \mathrm{E}\left(\ell(\boldsymbol{O},\boldsymbol{M};\boldsymbol{\theta})\middle|\boldsymbol{O},\boldsymbol{\theta}_p\right) \\
&= -n\log 2\pi - \frac{n}{2}\log|\boldsymbol{\Sigma}| \\
&\quad - \frac{1}{2}\sum_{i=1}^{m}(\boldsymbol{z}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_i-\boldsymbol{\mu}) \\
&\quad - \frac{1}{2}\sum_{j=1}^{m_x}\mathrm{E}\left((\boldsymbol{u}_j-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{u}_j-\boldsymbol{\mu})\middle|y_j;\boldsymbol{\theta}_p\right) \\
&\quad - \frac{1}{2}\sum_{k=1}^{m_y}\mathrm{E}\left((\boldsymbol{v}_k-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{v}_k-\boldsymbol{\mu})\middle|x_k;\boldsymbol{\theta}_p\right)
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
&\mathrm{E}\left((\boldsymbol{u}_j-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{u}_j-\boldsymbol{\mu})\middle|y_j;\boldsymbol{\theta}_p\right) \\
=&\mathrm{trace}\left(\boldsymbol{\Sigma}^{-1}\,\mathrm{Var}(\boldsymbol{u}_j-\boldsymbol{\mu}|y_j;\boldsymbol{\theta}_p)\right) + \mathrm{E}(\boldsymbol{u}_j-\boldsymbol{\mu}|y_j;\boldsymbol{\theta}_p)^T\boldsymbol{\Sigma}^{-1}\,\mathrm{E}(\boldsymbol{u}_j-\boldsymbol{\mu}|y_j;\boldsymbol{\theta}_p) \\
=&\mathrm{trace}\left(\boldsymbol{\Sigma}^{-1}\,\mathrm{Var}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p)\right) + \left(\mathrm{E}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p)-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathrm{E}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p)-\boldsymbol{\mu}\right), \quad j=1,\ldots,m_x
\end{aligned}
$$

and similarly

$$
\begin{aligned}
&\mathrm{E}\left((\boldsymbol{v}_k-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{v}_k-\boldsymbol{\mu})\middle|x_k;\boldsymbol{\theta}_p\right) \\
=&\mathrm{trace}\left(\boldsymbol{\Sigma}^{-1}\,\mathrm{Var}(\boldsymbol{v}_k-\boldsymbol{\mu}|x_k;\boldsymbol{\theta}_p)\right) + \mathrm{E}(\boldsymbol{v}_k-\boldsymbol{\mu}|x_k;\boldsymbol{\theta}_p)^T\boldsymbol{\Sigma}^{-1}\,\mathrm{E}(\boldsymbol{v}_k-\boldsymbol{\mu}|x_k;\boldsymbol{\theta}_p) \\
=&\mathrm{trace}\left(\boldsymbol{\Sigma}^{-1}\,\mathrm{Var}(\boldsymbol{v}_k|x_k;\boldsymbol{\theta}_p)\right) + \left(\mathrm{E}(\boldsymbol{v}_k|x_k;\boldsymbol{\theta}_p)-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathrm{E}(\boldsymbol{v}_k|x_k;\boldsymbol{\theta}_p)-\boldsymbol{\mu}\right), \quad k=1,\ldots,m_y
\end{aligned}
$$

In order to maximize $Q(\boldsymbol{\theta};\boldsymbol{\theta}_p)$, we take the matrix derivative wrt $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. We have

$$
\begin{aligned}
&\frac{\partial}{\partial\boldsymbol{\mu}}\mathrm{E}\left((\boldsymbol{u}_j-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{u}_j-\boldsymbol{\mu})\middle|y_j;\boldsymbol{\theta}_p\right) \\
=&\frac{\partial}{\partial\boldsymbol{\mu}}\mathrm{trace}\left(\boldsymbol{\Sigma}^{-1}\,\mathrm{Var}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p)\right) + \frac{\partial}{\partial\boldsymbol{\mu}}\left(\mathrm{E}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p)-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathrm{E}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p)-\boldsymbol{\mu}\right) \\
=&0+\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-\mathrm{E}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p)\right)+\boldsymbol{\Sigma}^{-T}\left(\boldsymbol{\mu}-\mathrm{E}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p)\right) \\
=&2\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-\mathrm{E}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p)\right)
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial}{\partial\boldsymbol{\Sigma}}\mathrm{E}\left((\boldsymbol{v}_k-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{v}_k-\boldsymbol{\mu})\middle|x_k;\boldsymbol{\theta}_p\right) \\
=&\frac{\partial}{\partial\boldsymbol{\Sigma}}\mathrm{trace}\left(\boldsymbol{\Sigma}^{-1}\,\mathrm{Var}(\boldsymbol{v}_k|x_k;\boldsymbol{\theta}_p)\right) + \frac{\partial}{\partial\boldsymbol{\Sigma}}\left(\mathrm{E}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p)-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathrm{E}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p)-\boldsymbol{\mu}\right) \\
=&-\left(\boldsymbol{\Sigma}^{-1}\,\mathrm{Var}\left(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p\right)\boldsymbol{\Sigma}^{-1}\right)^T-\boldsymbol{\Sigma}^{-T}\left(\mathrm{E}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p)-\boldsymbol{\mu}\right)\left(\mathrm{E}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p)-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-T} \\
=&-\boldsymbol{\Sigma}^{-T}\left[\mathrm{Var}\left(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p\right)+\left(\mathrm{E}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p)-\boldsymbol{\mu}\right)\left(\mathrm{E}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p)-\boldsymbol{\mu}\right)^T\right]\boldsymbol{\Sigma}^{-T}
\end{aligned}
$$

And similarly we have

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\mu}} \operatorname{E}\left((\boldsymbol{v}_k - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{v}_k - \boldsymbol{\mu}) \middle| x_k; \boldsymbol{\theta}_p\right) \\
=& \frac{\partial}{\partial \boldsymbol{\mu}} \operatorname{trace}\left(\boldsymbol{\Sigma}^{-1} \operatorname{Var}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p)\right) + \frac{\partial}{\partial \boldsymbol{\mu}} \left(\operatorname{E}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p) - \boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1} \left(\operatorname{E}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p) - \boldsymbol{\mu}\right) \\
=& 0 + \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu} - \operatorname{E}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p)\right) + \boldsymbol{\Sigma}^{-T} \left(\boldsymbol{\mu} - \operatorname{E}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p)\right) \\
=& 2\boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu} - \operatorname{E}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p)\right)
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\Sigma}} \operatorname{E}\left((\boldsymbol{v}_k - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{v}_k - \boldsymbol{\mu}) \middle| x_k; \boldsymbol{\theta}_p\right) \\
=& \frac{\partial}{\partial \boldsymbol{\Sigma}} \operatorname{trace}\left(\boldsymbol{\Sigma}^{-1} \operatorname{Var}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p)\right) + \frac{\partial}{\partial \boldsymbol{\Sigma}} \left(\operatorname{E}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p) - \boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1} \left(\operatorname{E}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p) - \boldsymbol{\mu}\right) \\
=& -\left(\boldsymbol{\Sigma}^{-1} \operatorname{Var}\left(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p\right) \boldsymbol{\Sigma}^{-1}\right)^T - \boldsymbol{\Sigma}^{-T} \left(\operatorname{E}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p) - \boldsymbol{\mu}\right) \left(\operatorname{E}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p) - \boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-T} \\
=& -\boldsymbol{\Sigma}^{-T} \left[\operatorname{Var}\left(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p\right) + \left(\operatorname{E}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p) - \boldsymbol{\mu}\right) \left(\operatorname{E}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p) - \boldsymbol{\mu}\right)^T\right] \boldsymbol{\Sigma}^{-T}
\end{aligned}
$$

And also

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\Sigma}} \log |\Sigma| = \boldsymbol{\Sigma}^{-T} \\
&\frac{\partial}{\partial \boldsymbol{\mu}} (\boldsymbol{z}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{z}_i - \boldsymbol{\mu}) = 2\boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{z}_i) \\
&\frac{\partial}{\partial \boldsymbol{\Sigma}} (\boldsymbol{z}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{z}_i - \boldsymbol{\mu}) = -\boldsymbol{\Sigma}^{-T} (\boldsymbol{z}_i - \boldsymbol{\mu}) (\boldsymbol{z}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-T}
\end{aligned}
$$

Hence

$$
\frac{\partial Q(\boldsymbol{\theta}; \boldsymbol{\theta}_p)}{\partial \boldsymbol{\mu}} = 2\boldsymbol{\Sigma}^{-1} \left(n\boldsymbol{\mu} - \sum_{i=1}^{m} \boldsymbol{z}_i - \sum_{j=1}^{m_x} \operatorname{E}(\boldsymbol{u}_j | y_j; \boldsymbol{\theta}_p) - \sum_{k=1}^{m_y} \operatorname{E}(v_k | x_k; \boldsymbol{\theta}_p)\right)
$$

and

$$
\begin{aligned}
\frac{\partial Q(\boldsymbol{\theta}; \boldsymbol{\theta}_p)}{\partial \boldsymbol{\Sigma}} =& -\frac{n}{2} \boldsymbol{\Sigma}^{-T} \\
&+ \frac{1}{2} \boldsymbol{\Sigma}^{-T} \left\{ \sum_{i=1}^{m} (\boldsymbol{z}_i - \boldsymbol{\mu})(\boldsymbol{z}_i - \boldsymbol{\mu})^T \right. \\
&+ \sum_{j=1}^{m_x} \left[\operatorname{Var}(\boldsymbol{u}_j | y_j; \boldsymbol{\theta}_p) + \left(\operatorname{E}(\boldsymbol{u}_j | y_j; \boldsymbol{\theta}_p) - \boldsymbol{\mu}\right) \left(\operatorname{E}(\boldsymbol{u}_j | y_j; \boldsymbol{\theta}_p) - \boldsymbol{\mu}\right)^T\right] \\
&\left. + \sum_{k=1}^{m_y} \left[\operatorname{Var}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p) + \left(\operatorname{E}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p) - \boldsymbol{\mu}\right) \left(\operatorname{E}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p) - \boldsymbol{\mu}\right)^T\right] \right\} \boldsymbol{\Sigma}^{-T}
\end{aligned}
$$

Set

$$
\begin{cases}
\frac{\partial Q(\boldsymbol{\theta}; \boldsymbol{\theta}_p)}{\partial \boldsymbol{\mu}} = 0 \\
\frac{\partial Q(\boldsymbol{\theta}; \boldsymbol{\theta}_p)}{\partial \boldsymbol{\Sigma}} = 0
\end{cases}
$$

We then have

$$
\hat{\boldsymbol{\mu}} = \frac{1}{n} \left\{ \sum_{i=1}^{m} \boldsymbol{z}_i + \sum_{j=1}^{m_x} \operatorname{E}(\boldsymbol{u}_j | y_j; \boldsymbol{\theta}_p) + \sum_{k=1}^{m_y} \operatorname{E}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p) \right\}
$$

and

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\Bigg\{ \sum_{i=1}^{m}(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}})^T$$
$$+ \sum_{j=1}^{m_x} \Big[\mathrm{Var}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p) + (\mathrm{E}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p) - \hat{\boldsymbol{\mu}})\,(\mathrm{E}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p) - \hat{\boldsymbol{\mu}})^T\Big]$$
$$+ \sum_{k=1}^{m_y} \Big[\mathrm{Var}(\boldsymbol{v}_k|x_k;\boldsymbol{\theta}_p) + (\mathrm{E}(\boldsymbol{v}_k|x_k;\boldsymbol{\theta}_p) - \hat{\boldsymbol{\mu}})\,(\mathrm{E}(\boldsymbol{v}_k|x_k;\boldsymbol{\theta}_p) - \hat{\boldsymbol{\mu}})^T\Big] \Bigg\}$$

Thus in the M-step of EM algorithom

$$\boldsymbol{\theta}_{p+1} = (\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$$

And suppose

$$\boldsymbol{\theta}_p = \left(\boldsymbol{\mu}_p = \begin{bmatrix}\mu_{x,p}\\ \mu_{y,p}\end{bmatrix}, \boldsymbol{\Sigma}_p = \begin{bmatrix}\sigma_{xx,p} & \sigma_{xy,p}\\ \sigma_{yx,p} & \sigma_{yy,p}\end{bmatrix}\right)$$

Then

$$\mathrm{E}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p) = \begin{bmatrix}\mathrm{E}(x_j^m|y_j;\boldsymbol{\theta}_p)\\ y_j\end{bmatrix} = \begin{bmatrix}\mu_{x,p} + \frac{\sigma_{xy,p}}{\sigma_{yy,p}}(y_j - \mu_{y,p})\\ y_j\end{bmatrix}$$

and

$$\mathrm{Var}(\boldsymbol{u}_j|y_j;\boldsymbol{\theta}_p) = \begin{bmatrix}\mathrm{Var}(x_j^m|y_j;\boldsymbol{\theta}_p) & 0\\ 0 & 0\end{bmatrix} = \begin{bmatrix}\sigma_{xx,p} - \frac{\sigma_{xy,p}^2}{\sigma_{yy,p}} & 0\\ 0 & 0\end{bmatrix}$$

Similarly, we have

$$\mathrm{E}(\boldsymbol{v}_k|x_k;\boldsymbol{\theta}_p) = \begin{bmatrix}x_k\\ \mathrm{E}(y_k^m|x_k;\boldsymbol{\theta}_p)\end{bmatrix} = \begin{bmatrix}x_k\\ \mu_{y,p} + \frac{\sigma_{yx,p}}{\sigma_{xx,p}}(x_k - \mu_{x,p})\end{bmatrix}$$

and

$$\mathrm{Var}(\boldsymbol{v}_k|x_k;\boldsymbol{\theta}_p) = \begin{bmatrix}0 & 0\\ 0 & \mathrm{Var}(y_k^m|x_k;\boldsymbol{\theta}_p)\end{bmatrix} = \begin{bmatrix}0 & 0\\ 0 & \sigma_{yy,p} - \frac{\sigma_{yx,p}^2}{\sigma_{xx,p}}\end{bmatrix}$$

R codes and estimates in each iteration (Table 1) is in the appendix. The final result of parameter estimates are

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix}\hat{\mu}_x\\ \hat{\mu}_y\end{bmatrix} = \begin{bmatrix}19.61405\\ 29.52332\end{bmatrix}$$

and

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix}\hat{\sigma}_{xx} & \hat{\sigma}_{xy}\\ \hat{\sigma}_{yx} & \hat{\sigma}_{yy}\end{bmatrix} = \begin{bmatrix}2.810984 & 2.146136\\ 2.146136 & 3.568150\end{bmatrix}$$

2. With only the observations where x and y are both observed, i.e. $\{\boldsymbol{z}_i\}_{i=1}^m$, the MLE is

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{m} \boldsymbol{z}_i$$

and

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{m}(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}})^T$$

So the estimate of parameters are

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix} \hat{\mu}_x \\ \hat{\mu}_y \end{bmatrix} = \begin{bmatrix} 19.88877 \\ 29.84538 \end{bmatrix}$$

and

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} \hat{\sigma}_{xx} & \hat{\sigma}_{xy} \\ \hat{\sigma}_{yx} & \hat{\sigma}_{yy} \end{bmatrix} = \begin{bmatrix} 1.6404591 & 0.4093769 \\ 0.4093769 & 0.8555870 \end{bmatrix}$$

3. In each iteration, we use the conditional mean to replace the missing data, and update the parameter estimates with the MLE of multivariate normal. Thus with the notatoin we use in part 1, the updated $\boldsymbol{\theta}_{p+1} = (\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$, and

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\left\{\sum_{i=1}^{m} \boldsymbol{z}_i + \sum_{j=1}^{m_x} \mathrm{E}(\boldsymbol{u}_j | y_j; \boldsymbol{\theta}_p) + \sum_{k=1}^{m_y} \mathrm{E}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p)\right\}$$

and

$$\begin{aligned}\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\Bigg\{ &\sum_{i=1}^{m}(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}})^T \\ &+ \sum_{j=1}^{m_x} \left(\mathrm{E}(\boldsymbol{u}_j | y_j; \boldsymbol{\theta}_p) - \hat{\boldsymbol{\mu}}\right)\left(\mathrm{E}(\boldsymbol{u}_j | y_j; \boldsymbol{\theta}_p) - \hat{\boldsymbol{\mu}}\right)^T \\ &+ \sum_{k=1}^{m_y} \left(\mathrm{E}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p) - \hat{\boldsymbol{\mu}}\right)\left(\mathrm{E}(\boldsymbol{v}_k | x_k; \boldsymbol{\theta}_p) - \hat{\boldsymbol{\mu}}\right)^T \Bigg\}\end{aligned}$$

R codes and estimates in each iteration (Table 2) is in the appendix. The final result of parameter estimates are

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix} \hat{\mu}_x \\ \hat{\mu}_y \end{bmatrix} = \begin{bmatrix} 19.57659 \\ 29.52319 \end{bmatrix}$$

and

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} \hat{\sigma}_{xx} & \hat{\sigma}_{xy} \\ \hat{\sigma}_{yx} & \hat{\sigma}_{yy} \end{bmatrix} = \begin{bmatrix} 3.243723 & 2.867498 \\ 2.867498 & 3.261116 \end{bmatrix}$$

4. Since both estimators are maximizer of a marginal likelihood with the same full model, so when the data are missing at random, the marginals can go through the whole data, and I anticipate that their bias are close. And since we ultilize more information in the estimator of part 1 (EM), I anticipate that the variance of estimator of part 1 is smaller, or the precision higher.

Appendix

Table 1: Estimates of paramters in each iteration with EM in part 1

iteration	μ_x	μ_y	σ_{xx}	σ_{xy}	σ_{yy}
1	19.6653238614613	29.6158278987521	2.47591305559001	1.4549260732653	2.90720611314166
2	19.6247983107412	29.5492852891321	2.63346939876761	1.78196148746505	3.41666635776087
3	19.6203090856511	29.5333131326926	2.66865423476446	1.89941644431462	3.53972174506091
4	19.6201846332483	29.527965996642	2.68625921485882	1.96009048829923	3.57147491095681
5	19.6199809770128	29.5255499857348	2.70315390515685	2.00120816495289	3.5802057795633
6	19.6194342940863	29.5243148394793	2.71991823107853	2.03218570447614	3.58206733804075
7	19.6187272019664	29.5236694024572	2.73544094686566	2.0562614156736	3.58150489952278
8	19.6179997034887	29.5233409935733	2.74908865929165	2.07514533488358	3.58005583151427
9	19.6173243360867	29.523185427252	2.76070259080081	2.09000861640648	3.57835992349057
10	19.6167308870376	29.5231230486707	2.77038304068624	2.10172952595671	3.57670288666912
11	19.6162261398478	29.5231094195019	2.77834229539182	2.1109845105604	3.57520644849858
12	19.6158056255238	29.5231198454911	2.78482532793017	2.11829952342381	3.57391164495919
13	19.6154600751066	29.5231407882356	2.79007102293983	2.12408553377033	3.57282008195611
14	19.6151788012436	29.5231650133307	2.7942951539743	2.12866476346562	3.57191532393874
15	19.6149513732057	29.5231888303091	2.79768457316405	2.13229050675776	3.57117402593657
16	19.6147683661534	29.5232105183805	2.80039696314812	2.13516227260385	3.57057159248771
17	19.6146216216967	29.5232294288835	2.80256315667669	2.13743744721381	3.57008489221991
18	19.6145042613809	29.5232454756034	2.80429045284818	2.13924033410125	3.56969339941055
19	19.6144105848894	29.5232588479649	2.80566612567061	2.14066919707609	3.56937951381353
20	19.6143359233686	29.5232698523371	2.80676073510366	2.14180176929873	3.5691284709019
21	19.6142764841905	29.5232788267733	2.80763107365279	2.14269957585522	3.56892806701104
22	19.6142292047186	29.5232860975475	2.80832269916719	2.14341133350333	3.56876831938941
23	19.6141916225136	29.5232919591583	2.8088720638857	2.14397562953187	3.56864112295253
24	19.6141617640812	29.5232966671989	2.80930827716414	2.14442303558597	3.56853993329234
25	19.614138051612	29.5233004379995	2.8096545493921	2.14477777746206	3.56845948790336
26	19.6141192259617	29.5233034515734	2.80992936563241	2.14505905526141	3.56839556827367
27	19.6141042837009	29.5233058559322	2.81014743412692	2.14528208779849	3.56834480083004
28	19.6140924260462	29.5233077717261	2.81032044956619	2.14545893928106	3.56830449265239
29	19.6140830176543	29.5233092966746	2.81045770529655	2.14559917391165	3.56827249720016
30	19.614075553508	29.5233105095461	2.81056658313501	2.14571037439039	3.56824710534719
31	19.6140696323858	29.523311473602	2.81065294450492	2.14579855271522	3.56822695742277
32	19.6140649356552	29.5233122395085	2.81072144230499	2.14586847571328	3.5682109724925
33	19.6140612103502	29.5233128477567	2.81077576928557	2.14592392304096	3.56819829167337
34	19.6140582556875	29.5233133306519	2.8108188556781	2.14596789168618	3.56818823280406
35	19.6140559123327	29.5233137139355	2.81085302633352	2.14600275808083	3.5681802542617
36	19.6140540538624	29.5233140180973	2.81088012559526	2.14603040661945	3.56817392611995
37	19.6140525799787	29.5233142594341	2.81090161648814	2.14605233155159	3.56816890718438
38	19.6140514111174	29.5233144508996	2.81091865947824	2.14606971776782	3.56816492672312
39	19.6140504841671	29.5233146027852	2.81093217499429	2.14608350485378	3.56816176994272
40	19.6140497490695	29.5233147232639	2.8109428930476	2.14609443787916	3.56815926644742
41	19.6140491661218	29.5233148188243	2.81095139260617	2.14610310766957	3.56815728107311
42	19.6140487038353	29.5233148946167	2.81095813283422	2.14610998273971	3.56815570660946
43	19.6140483372372	29.5233149547282	2.81096347787518	2.14611543461484	3.56815445802303
44	19.6140480465223	29.5233150024015	2.81096771651074	2.14611975790965	3.56815346787253
45	19.6140478159841	29.5233150402095	2.81097107775445	2.14612318625015	3.56815268267103
46	19.6140476331668	29.5233150701931	2.81097374321981	2.14612590489928	3.56815205999964
47	19.6140474881927	29.5233150939713	2.81097585692972	2.14612806076877	3.56815156621791
48	19.6140473732283	29.5233151128281	2.81097753309626	2.14612977035827	3.56815117464757
49	19.6140472820618	29.5233151277819	2.81097886229068	2.14613112605103	3.56815086413192
50	19.6140472097671	29.5233151396405	2.81097991633654	2.14613220110621	3.56815061789318
51	19.6140471524376	29.5233151490445	2.81098075219015	2.14613305361777	3.56815042262631
52	19.6140471069756	29.523315156502	2.81098141501795	2.14613372965376	3.56815026778022
53	19.6140470709245	29.5233151624158	2.81098194063693	2.14613426574583	3.56815014498781
54	19.6140470423361	29.5233151671055	2.8109823574499	2.14613469086333	3.56815004761394
55	19.6140470196657	29.5233151708243	2.81098268798022	2.14613502797871	3.56814997039691
56	19.6140470016881	29.5233151737734	2.81098295008883	2.14613529530901	3.56814990916419
57	19.6140469874321	29.523315176112	2.8109831579394	2.1461355073002	3.56814986060696
58	19.6140469761271	29.5233151779666	2.81098332276363	2.14613567540783	3.56814982210135
59	19.6140469671623	29.5233151794372	2.81098345346824	2.14613580871609	3.56814979156663
60	19.6140469600533	29.5233151806034	2.81098355711618	2.14613591442868	3.56814976735277
61	19.6140469544159	29.5233151815282	2.81098363930835	2.14613599825808	3.56814974815132
62	19.6140469499454	29.5233151822615	2.81098370448622	2.14613606473425	3.56814973292469
63	19.6140469464004	29.5233151828431	2.81098375617186	2.14613611744942	3.56814972085007
64	19.6140469435892	29.5233151833042	2.81098379715824	2.14613615925221	3.56814971127497
65	19.61404694136	29.5233151836699	2.81098382966019	2.14613619240156	3.56814970368197
66	19.6140469395922	29.5233151839599	2.81098385543402	2.14613621868877	3.56814969766077
67	19.6140469381904	29.5233151841899	2.8109838758725	2.14613623953437	3.568149692886
68	19.6140469370787	29.5233151843723	2.81098389208009	2.14613625606479	3.56814968909964
69	19.6140469361972	29.5233151845169	2.8109839049326	2.1461362691733	3.56814968609708
70	19.6140469354981	29.5233151846316	2.81098391512456	2.14613627956827	3.56814968371606
71	19.6140469349438	29.5233151847225	2.81098392320672	2.14613628781142	3.56814968182794

72	19.6140469345042	29.5233151847946	2.81098392961582	2.14613629434818	3.56814968033066

Table 2: Estimates of parameters in each iteration with conditional mean method in part 3

iteration	μ_x	μ_y	σ_{xx}	σ_{xy}	σ_{yy}
1	19.6653238614613	29.6158278987521	1.99438555582682	1.4549260732653	2.73140650305178
2	19.6209059162327	29.5383782757245	2.1267798849798	1.96088528170276	3.07569024945733
3	19.6098854037994	29.515064952074	2.40609073493248	2.34485586878782	3.31424119183557
4	19.6050614376452	29.5086358286186	2.61925962697159	2.54295927291646	3.38930804704025
5	19.6013286233277	29.5084638378349	2.76077552037522	2.63743698942437	3.38423596039412
6	19.5976572536855	29.5102322221899	2.86200587265019	2.69196531483534	3.36173696857118
7	19.5941382054047	29.5123635073115	2.93951799324605	2.72986476338103	3.34052024754429
8	19.59097738578	29.5143508512481	3.00055757298729	2.75851490783122	3.32349761556987
9	19.5882680818543	29.5160556926393	3.04910425473098	2.78083333074826	3.3102346894989
10	19.5860115173055	29.51746564317	3.08786364629005	2.79842213072747	3.29991412376291
11	19.5841638168771	29.5186105590927	3.11886372862242	2.81236111821885	3.29184875583214
12	19.5826660492764	29.5195313829889	3.14368082951538	2.82344373007246	3.28551636922103
13	19.5814590849636	29.520268299553	3.16355830734672	2.83227406077027	3.28052464543915
14	19.5804897697376	29.5208566201528	3.17948382974084	2.83932014777311	3.27657688228718
15	19.5797127934829	29.5213258553694	3.19224499554977	2.84494835436602	3.27344655891366
16	19.5790906180631	29.5217000469449	3.20247127124072	2.8494473763937	3.27095921150359
17	19.5785926388693	29.5219985205878	3.21066639993872	2.85304574505642	3.2689794653179
18	19.5781941317431	29.5222367043481	3.21723381603393	2.85592494218921	3.26740162857806
19	19.5778752266906	29.5224268747324	3.22249675434288	2.85822941365161	3.26614276862094
20	19.5776200005298	29.5225787888165	3.22671424930474	2.86007431581876	3.26513754529848
21	19.5774157118167	29.5227002008397	3.23009390683921	2.86155156472562	3.26433430909073
22	19.5772521719042	29.5227972766253	3.23280212012317	2.86273459150553	3.26369212378253
23	19.5771212351508	29.5228749229196	3.23497225017313	2.86368209887824	3.26317847467932
24	19.5770163889416	29.5229370479932	3.23671118079068	2.86444103894907	3.26276749154482
25	19.5769324253783	29.5229867675559	3.23810457258364	2.86504897981017	3.26243856224508
26	19.5768651789339	29.5230265674914	3.23922107482549	2.86553598997757	3.26217524575255
27	19.5768113170115	29.5230584325958	3.2401157019864	2.86592614103768	3.2619644163489
28	19.5767681727756	29.5230839485645	3.24083254054855	2.86623870714409	3.26179558757979
29	19.5767336117049	29.523104382908	3.24140691883745	2.86648912330849	3.26166037674126
30	19.5767059250257	29.5231207492436	3.24186714629725	2.86668975147132	3.26155207974085
31	19.5766837445792	29.5231338584478	3.24223590756282	2.86685049320716	3.2614653329764
32	19.5766659747809	29.5231443594019	3.24253137978137	2.86697927995367	3.26139584403035
33	19.57665173822	29.5231527714821	3.24276812808008	2.86708246536248	3.26134017692258
34	19.576640332149	29.5231595104898	3.2429578232	2.86716513938735	3.26129558070674
35	19.5766311936718	29.5231649093601	3.24310981659186	2.86723137976671	3.26125985255808
36	19.5766238718856	29.5231692347116	3.24323160127116	2.86728445340869	3.26123122834511
37	19.5766180055804	29.5231727000785	3.24332918111659	2.86732697767701	3.26120829512502
38	19.5766133053862	29.5231754764942	3.24340736679487	2.86736104957398	3.26118992114297
39	19.5766095394775	29.5231777009602	3.2434700128863	2.86738834921788	3.2611751998161
40	19.576606522125	29.5231794832235	3.24352020789138	2.86741022273471	3.26116340489701
41	19.5766041045266	29.5231809112023	3.24356042647992	2.86742774866256	3.26115395457829
42	19.5766021674637	29.5231820553308	3.24359265148446	2.86744179114995	3.26114638275042
43	19.5766006154182	29.5231829720373	3.2436184716499	2.86745304257577	3.26114031598444
44	19.5765993718598	29.5231837065306	3.24363915995683	2.86746205769503	3.26113545509724
45	19.5765983754715	29.5231842950312	3.24365573637897	2.86746928099597	3.26113156038589
46	19.5765975771246	29.5231847665589	3.24366901816806	2.86747506861848	3.26112843979987
47	19.5765969374557	29.5231851443647	3.24367966014484	2.86747970591524	3.26112593946625
48	19.5765964249259	29.5231854470775	3.24368818698247	2.86748342152199	3.26112393609931
49	19.5765960142649	29.5231856896233	3.24369501907349	2.86748639863094	3.26112233091971
50	19.5765956852255	29.523185883961	3.24370049325594	2.86748878402328	3.26112104478269
51	19.5765954215847	29.5231860396726	3.24370487941997	2.86749069530627	3.26112001427504
52	19.5765952103441	29.5231861644354	3.24370839381402	2.86749222671189	3.26111918858791
53	19.5765950410887	29.5231862644007	3.24371120970622	2.86749345374289	3.26111852701146
54	19.5765949054737	29.5231863444973	3.24371346592694	2.86749443689535	3.26111799692737
55	19.576594796813	29.5231864086743	3.24371527371341	2.86749522464142	3.26111757220058
56	19.5765947097491	29.5231864600957	3.24371672219399	2.86749585581913	3.26111723189062
57	19.5765946399896	29.5231865012968	3.24371788278236	2.86749636154721	3.26111695921911
58	19.5765945840952	29.5231865343089	3.24371881269841	2.86749676675941	3.26111674074249
59	19.57659453931	29.5231865607598	3.24371955778933	2.86749709143375	3.26111656568923
60	19.5765945034261	29.5231865819534	3.24372015478998	2.86749735157753	3.26111642542869
61	19.5765944746744	29.5231865989346	3.24372063313397	2.86749756001649	3.26111631304563
62	19.5765944516371	29.5231866125408	3.24372101640486	2.86749772702723	3.26111622299925
63	19.5765944331786	29.5231866234426	3.24372132349889	2.8674978608438	3.26111615085003
64	19.5765944183889	29.5231866321777	3.24372156955654	2.8674979680637	3.26111609304081
65	19.5765944065387	29.5231866391766	3.24372176670909	2.86749805397314	3.26111604672145
66	19.5765943970437	29.5231866447845	3.24372192467666	2.86749812280767	3.26111600960829
67	19.5765943894359	29.5231866492778	3.24372205124743	2.86749817796102	3.26111597987154
68	19.5765943833403	29.523186652878	3.24372215266167	2.86749822215238	3.26111595604511
69	19.5765943784561	29.5231866557626	3.24372223391936	2.8674982575605	3.2611159369543
70	19.5765943745427	29.5231866580739	3.2437222990267	2.8674982859311	3.26111592165788

71	19.5765943714071	29.5231866599259	3.24372235119366	2.8674983086629	3.26111590940169
72	19.5765943688947	29.5231866614097	3.24372239299219	2.86749832687665	3.26111589958148
73	19.5765943668817	29.5231866625986	3.24372242648307	2.86749834147034	3.26111589171308
74	19.5765943652688	29.5231866635513	3.24372245331748	2.86749835316346	3.26111588540856
75	19.5765943639764	29.5231866643145	3.24372247481842	2.86749836253252	3.26111588035709
76	19.5765943629409	29.5231866649261	3.24372249204595	2.86749837003943	3.26111587630963
77	19.5765943621112	29.5231866654162	3.24372250584943	2.86749837605431	3.26111587306662
78	19.5765943614465	29.5231866658088	3.2437225169094	2.86749838087371	3.26111587046817
79	19.5765943609138	29.5231866661234	3.24372252577115	2.86749838473522	3.26111586838618

R codes:

```
data <- read.table("./bivnormdat.txt", head = T)
z <- data[!(is.na(data["x"])|is.na(data["y"])),]
u <- data[is.na(data["x"]),]
v <- data[is.na(data["y"]),]

z <- t(z)

yj <- u[,"y"]
xk <- v[,"x"]

#2 MLE with only zi
mu_2 <- apply(z, MARGIN = 1, mean)

quardra <- function(x){
  return(x%*%t(x))
}

quadra_sum <- function(X){
  s <- apply(apply(X, MARGIN = 2, quardra), MARGIN = 1, sum)
  return(matrix(s, nrow = 2))
}

zminusmu <- z - matrix(rep(mu_2, ncol(z)), nrow = 2)

Sigma_2 <- quadra_sum(zminusmu)/ncol(zminusmu)

## the estimate with this method is
theta_2 <- list(mu = mu_2, Sigma = Sigma_2)

theta_2

#1 Using EM Algorithm
Eu <- function(y_j, mu_p, Sigma_p){
  Exjm <- mu_p[1] + (Sigma_p[1,2]/Sigma_p[2,2])*(y_j - mu_p[2])
  return(c(Exjm, y_j))
}

Ev <- function(x_k, mu_p, Sigma_p){
  Eykm <- mu_p[2] + (Sigma_p[2,1]/Sigma_p[1,1])*(x_k - mu_p[1])
  return(c(x_k, Eykm))
}

Vu <- function(mu_p, Sigma_p){
```

```
  return(matrix(c(Sigma_p[1,1] - Sigma_p[1,2]^2/Sigma_p[2,2], 0, 0,
   ↪ 0), nrow = 2))
}

Vv <- function(mu_p, Sigma_p){
  return(matrix(c(0, 0, 0, Sigma_p[2,2] -
   ↪ Sigma_p[2,1]^2/Sigma_p[1,1]), nrow = 2))
}

theta_update <- function(z, yj, xk, mu_p, Sigma_p){
  n <- nrow(data)

  zsum <- apply(z, MARGIN = 1, sum)
  Eus <- sapply(yj, Eu, mu_p = mu_p, Sigma_p = Sigma_p)
  Eusum <- apply(Eus, MARGIN = 1, sum)
  Evs <- sapply(xk, Ev, mu_p = mu_p, Sigma_p = Sigma_p)
  Evsum <- apply(Evs, MARGIN = 1, sum)
  mu_new <- (zsum + Eusum + Evsum)/n

  zminusmu <- z - matrix(rep(mu_new, ncol(z)), nrow = 2)
  Euminusmu <- Eus - matrix(rep(mu_new, ncol(Eus)), nrow = 2)
  Evminusmu <- Evs - matrix(rep(mu_new, ncol(Evs)), nrow = 2)

  Sigma_new <- (1/n)*(quadra_sum(zminusmu) + quadra_sum(Euminusmu) +
   ↪ quadra_sum(Evminusmu) + Vu(mu_p, Sigma_p)*length(yj) + Vv(mu_p,
   ↪ Sigma_p)*length(xk))

  return(list(mu = mu_new, Sigma = Sigma_new))
}


# initialize with result from question 2
theta_old <- list(mu = mu_2, Sigma = Sigma_2)
iter <- 0
print(noquote(paste(c("iteration", "mu_x", "mu_y", "sigma_xx",
 ↪ "sigma_xy", "sigma_yy"), collapse = '&')))
repeat{
    iter <- iter + 1
    theta_new <- theta_update(z, yj, xk, theta_old$mu,
     ↪ theta_old$Sigma)
    print(noquote(paste(c(iter, c(theta_new$mu, theta_new$Sigma)[-4]),
     ↪ collapse='&')))
    diff <- (c(theta_new$mu, theta_new$Sigma) - c(theta_old$mu,
     ↪ theta_old$Sigma))[-4]
    diffnorm <- sqrt(sum(diff^2))
    if(diffnorm < 1e-8)
      break
    theta_old <- theta_new
}

theta_new
```

```r
# 3 using conditional mean to replace missing data
theta_update_3 <- function(z, yj, xk, mu_p, Sigma_p){
  n <- nrow(data)

  zsum <- apply(z, MARGIN = 1, sum)
  Eus <- sapply(yj, Eu, mu_p = mu_p, Sigma_p = Sigma_p)
  Eusum <- apply(Eus, MARGIN = 1, sum)
  Evs <- sapply(xk, Ev, mu_p = mu_p, Sigma_p = Sigma_p)
  Evsum <- apply(Evs, MARGIN = 1, sum)
  mu_new <- (zsum + Eusum + Evsum)/n

  zminusmu <- z - matrix(rep(mu_new, ncol(z)), nrow = 2)
  Euminusmu <- Eus - matrix(rep(mu_new, ncol(Eus)), nrow = 2)
  Evminusmu <- Evs - matrix(rep(mu_new, ncol(Evs)), nrow = 2)

  Sigma_new <- (1/n)*(quadra_sum(zminusmu) + quadra_sum(Euminusmu) +
  ↪  quadra_sum(Evminusmu))

  return(list(mu = mu_new, Sigma = Sigma_new))
}

theta_old <- list(mu = mu_2, Sigma = Sigma_2)
iter <- 0
print(noquote(paste(c("iteration", "mu_x", "mu_y", "sigma_xx",
↪  "sigma_xy", "sigma_yy"), collapse = '&')))
repeat{
  iter <- iter + 1
  theta_new <- theta_update_3(z, yj, xk, theta_old$mu,
  ↪  theta_old$Sigma)
  print(noquote(paste(c(iter, c(theta_new$mu, theta_new$Sigma)[-4]),
  ↪  collapse='&')))
  diff <- (c(theta_new$mu, theta_new$Sigma) - c(theta_old$mu,
  ↪  theta_old$Sigma))[-4]
  diffnorm <- sqrt(sum(diff^2))
  if(diffnorm < 1e-8)
    break
  theta_old <- theta_new
}

theta_new
```