1 Appendix A

In this section, the derivation of conditional expectations in the E-step is presented, and the details are shown as following. There are five conditional expectations: $E_\Theta(Z_{ik}|\mathbf{X})$, $E_\Theta(Z_{ik}u_i|\mathbf{X})$, $E_\Theta(Z_{ik}u_i\tau_i|\mathbf{X})$, $E_\Theta(Z_{ik}u_i\tau_i^2|\mathbf{X})$, $E_\Theta(Z_{ik}\log u_i|\mathbf{X})$, will be estimated in proposed method. Before introducing them, we should calculate some probability density functions(pdf) by the truncated distribution represent in Eq. (23). If $\mathbf{X}$ can be formulated in a three-level hierarchical representation in Eq. (23), the density function of $f(x_i,u_i,\tau_i)$ is given by:

$$\begin{aligned}
f(x_i,u_i,\tau_i) &= p(x_i|u_i,\tau_i)p(t_i|u_i)p(u_i)\\
&= \frac{1}{(2\pi)^{D/2}|u_i^{-1}\Upsilon_k|^{1/2}}\exp(-\frac{u_i(x_i-\mu_k-\Delta_k\tau_i)^T\Upsilon_k^{-1}(x_i-\mu_k-\Delta_k\tau_i)}{2})\\
&\quad\times\frac{2}{(2\pi)^{1/2}|u_i^{-1}|^{1/2}}\exp(-\frac{u_i\tau_i^2}{2})\times\frac{(\nu_k/2)^{\nu_k/2}}{\Gamma(\nu_k/2)}u_i^{\frac{\nu_k}{2}}\exp(-\frac{\nu_k}{2}u_i)\\
&= \frac{2}{(2\pi)^{(D+1)/2}|\Upsilon_k|^{1/2}}u_i^{\frac{\nu_k+D+1}{2}-1}\times\frac{(\nu_k/2)^{\nu_k/2}}{\Gamma(\nu_k/2)}\times\exp(-\frac{u_i[(x_i-\mu_k)^T\Sigma_k^{-1}(x_i-\mu_k)+\nu_k]}{2}\\
&\quad\times\exp(-\frac{u_i[\tau_i-\Delta_k^T\Sigma_k^{-1}(x_i-\mu_k)]^2}{2(1-\delta_k^T\delta_k)})
\end{aligned}\tag{1}$$

Then the density function of $f(x_i,u_i)$ is written as:

$$\begin{aligned}
f(x_i,u_i) &= \int_0^\infty f(x_i,u_i,\tau_i)d\tau_i\\
&= \frac{2}{(2\pi)^{(D+1)/2}|\Upsilon_k|^{1/2}}u_i^{\frac{\nu_k+D+1}{2}-1}\times\frac{(\nu_k/2)^{\nu_k/2}}{\Gamma(\nu_k/2)}\times\int_0^\infty\exp(-\frac{u_i[(x_i-\mu_k)^T\Sigma_k^{-1}(x_i-\mu_k)+\nu_k]}{2})\\
&\quad\times\exp(-\frac{u_i[\tau_i-\Delta_k^T\Sigma_k^{-1}(x_i-\mu_k)]^2}{2(1-\delta_k^T\delta_k)})d\tau_i\\
&= \frac{2}{(2\pi)^{(D+1)/2}|\Sigma_k|^{1/2}}u_i^{\frac{\nu_k+D}{2}-1}\times\frac{(\nu_k/2)^{\nu_k/2}}{\Gamma(\nu_k/2)}\times\exp(-\frac{u_i[(x_i-\mu_k)^T\Sigma_k^{-1}(x_i-\mu_k)+\nu_k]}{2})\\
&\quad\times\Phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i})
\end{aligned}\tag{2}$$

Similarly, the density function of $f(x_i,\tau_i)$ is written as:

$$\begin{aligned}
f(x_i,\tau_i) &= \int_0^\infty f(x_i,u_i,\tau_i)du_i\\
&= \frac{2}{(2\pi)^{(D+1)/2}|\Upsilon_k|^{1/2}}\times\frac{(\nu_k/2)^{\nu_k/2}}{\Gamma(\nu_k/2)}\times\int_0^\infty u_i^{\frac{\nu_k+D+1}{2}-1}\times\exp(-\frac{u_i(x_i-\mu_k)^T\Sigma_k^{-1}(x_i-\mu_k)+\nu_k}{2})\\
&\quad\times\exp(-\frac{u_i[\tau_i-\Delta_k^T\Sigma_k^{-1}(x_i-\mu_k)]^2}{2(1-\delta_k^T\delta_k)})du_i\\
&= \frac{2}{(2\pi)^{(D+1)/2}|\Upsilon_k|^{1/2}}\times\frac{(\nu_k/2)^{\nu_k/2}}{\Gamma(\nu_k/2)}\times\Gamma(\frac{\nu_k+D+1}{2})\times\left[\begin{array}{c}\frac{\nu_k+(x_i-\mu_k)^T\Sigma_k^{-1}(x_i-\mu_k)}{2}\\ +\frac{((1-\delta^T\delta)^{-1/2}\tau_i-\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k))^2}{2}\end{array}\right]^{-(\frac{\nu_k+D+1}{2}-1)}
\end{aligned}\tag{3}$$

Then the density of $f(u_i|x_i)$ is calculated by:

$$\begin{aligned}
f(u_i|x_i) &= f(x_i,u_i)/f(x_i)\\
&= \frac{f(x_i,u_i)}{2^D t_D(x_i|\mu_k,\Sigma_k,\nu_k)T(\sqrt{(\nu_k+D)/(\nu_k+d)}A,\nu_k+D)}\\
&= \frac{u_i^{\frac{\nu_k+D}{2}-1}\exp(-u_i[(x_i-\mu_k)^T\Sigma_k^{-1}(x_i-\mu_k)+\nu_k]/2)}{\Gamma((\nu_k+D)/2)\times T(\sqrt{(\nu_k+D)/(\nu_k+d)}A,\nu_k+D)}\times\Phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i})((\nu_k+D)/2)^{\frac{\nu_k+D}{2}}
\end{aligned}\tag{4}$$

For $f(\tau_i|x_i,u_i)$, it can be calculated by:

$$\begin{aligned} f(\tau_i|x_i,u_i) &= f(x_i,u_i,\tau_i)/f(x_i,u_i) \\ &= \frac{\sqrt{u_i}}{\sqrt{2\pi}(1-\delta^T\delta)^{1/2}\times\Phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i})}\times\exp(-\frac{u_i[\tau_i-\Delta_k^T\Sigma_k^{-1}(x_i-\mu_k)]^2}{2(1-\delta_k^T\delta_k)}) \end{aligned} \tag{5}$$

thus, $\tau_i|x_i,u_i \sim TN(\Delta_k^T\Sigma_k^{-1}(x_i-\mu_k), u_i^{-1}(1-\delta^T\delta))$.

In the next step, we will employ the pdf above to estimated the conditional expectations. Firstly, employing the proposition in [1]:
If $x\sim\Gamma(\alpha,\beta)$, then for any $a\in\mathbb{R}$, $E(\Phi(a\sqrt{x}))=T_{2\alpha}(a\sqrt{\frac{\alpha}{\beta}})$,
then we get:

$$\begin{aligned} E(u_i|x_i) &= \int_0^\infty u_i f(u_i|x_i)du_i \\ &= \frac{\Gamma(\frac{\nu_k+D+2}{2})}{\Gamma(\frac{\nu_k+D}{2})\times T(\sqrt{\frac{D+\nu_k}{d+\nu_k}}A,\nu_k+D)}\times\frac{(\frac{\nu_k+d}{2})^{\frac{\nu_k+D}{2}}}{(\frac{[(x_i-\mu_k)^T\Sigma_k^{-1}(x_i-\mu_k)+\nu_k]}{2})^{\frac{\nu_k+D+2}{2}}} \\ &\quad\times\int_0^\infty\gamma(u_i|\frac{\nu_k+D+2}{2},\frac{[(x_i-\mu_k)^T\Sigma_k^{-1}(x_i-\mu_k)+\nu_k]}{2})\times\Phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i})du_i \\ &= \frac{\Gamma(\frac{\nu_k+D+2}{2})}{\Gamma(\frac{\nu_k+D}{2})\times T(\sqrt{\frac{D+\nu_k}{d+\nu_k}}A,\nu_k+D)}\times\frac{(\frac{\nu_k+d}{2})^{\frac{\nu_k+D}{2}}}{(\frac{[(x_i-\mu_k)^T\Sigma_k^{-1}(x_i-\mu_k)+\nu_k]}{2})^{\frac{\nu_k+D+2}{2}}}\times E(\Phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i})) \\ &= (\frac{\nu_k+D}{d+\nu_k})\times\frac{T(\sqrt{\frac{D+\nu_k+2}{d+\nu_k}}A,\nu_k+D+2)}{T(\sqrt{\frac{D+\nu_k}{d+\nu_k}}A,\nu_k+D)} \end{aligned} \tag{6}$$

As for $E(\tau_i|u_i,x_i), E(\tau_i^2|u_i,x_i)$, we use the Proposition in [1]: If $\tau_i|x_i,u_i\sim TN(\Delta_k^T\Sigma_k^{-1}(x_i-\mu_k),u_i^{-1}(1-\delta^T\delta))$, then we have:

$$E(\tau_i|u_i,x_i)=\Delta_k^T\Sigma_k^{-1}(x_i-\mu_k)+\frac{\phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i})}{\Phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i})}\times\sqrt{u_i^{-1}(1-\delta^T\delta)} \tag{7}$$

$$E(\tau_i^2|u_i,x_i)=(\Delta_k^T\Sigma_k^{-1}(x_i-\mu_k))^2+u_i^{-1}(1-\delta^T\delta)+\frac{\phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i})}{\Phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i})}\Delta_k^T\Sigma_k^{-1}(x_i-\mu_k)\sqrt{u_i^{-1}(1-\delta^T\delta)} \tag{8}$$

Then $E(u_i\tau_i|x_i)$ is:

$$\begin{aligned} E(u_i\tau_i|x_i) &= E(u_iE(\tau_i|u_i,x_i)|x_i) \\ &= E(u_i|x_i)\times\Delta_k^T\Sigma_k^{-1}(x_i-\mu_k)+\sqrt{(1-\delta^T\delta)}\times E(\sqrt{u_i}\frac{\phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i})}{\Phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i})}|x_i) \end{aligned} \tag{9}$$

As for $E(\sqrt{u_i}\frac{\phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i})}{\Phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i})}|x_i)$, it can be written as:

$$\begin{aligned} E(\sqrt{u_i}\frac{\phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i})}{\Phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i})}|x_i) &= \int_0^\infty\sqrt{u_i}\frac{\phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i})}{\Phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i})}\frac{f(u_i,x_i)}{f(x_i)}du_i \\ &= \frac{t_D(x_i|\mu_k,\Sigma_k,\nu_k)}{f(x_i)}\times\frac{2\Gamma(\frac{\nu_k+D+1}{2})\times(\nu_k+d)^{(D+\nu_k)/2}}{\Gamma(\frac{\nu_k+D}{2})\pi^{1/2}(A^2+\nu_k+d)^{\frac{\nu_k+D+d}{2}}} \end{aligned} \tag{10}$$

Then, $E(u_i\tau_i|x_i)$ is:

$$E(u_i\tau_i|x_i)=E(u_i|x_i)S_{1,i,j}+M_iS_{2,i,j} \tag{11}$$

And $E(u_i\tau_i^2|x_i)$ is:

$$
\begin{aligned}
E(u_i\tau_i^2|x_i) &= E(u_iE(\tau_i^2|u_i,x_i)|x_i) \\
&= E(u_i|x_i) \times (\Delta_k^T\Sigma_k^{-1}(x_i-\mu_k))^2 + (1-\delta^T\delta) + (\Delta_k^T\Sigma_k^{-1}(x_i-\mu_k))\sqrt{u_i^{-1}(1-\delta^T\delta)} \\
&\quad \times E(\sqrt{u_i}\frac{\phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i})}{\Phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i})}|x_i) \\
&= S_{1,i,j}^2E(u_i|x_i) + M_i^2 + M_iS_{2,i,j}S_{1,i,j}
\end{aligned}
\tag{12}
$$

As for $E(\log u_i|x_i)$, it can be gotten by the formula $\frac{d}{d\nu_k}\int_0^\infty f(u_i|x_i)du_i = 0$, where $\frac{d}{d\nu_k}\int_0^\infty f(u_i|x_i)du_i$ is given by:

$$
\begin{aligned}
\frac{d}{d\nu_k}\int_0^\infty f(u_i|x_i)du_i &= \int_0^\infty \exp(-\frac{u_i[(x_i-\mu_k)^T\Sigma_k^{-1}(x_i-\mu_k)]}{2}) \\
&\quad \times \Phi(\lambda_k^T\Sigma_k^{-1/2}(x_i-\mu_k)\sqrt{u_i}) \times \frac{d}{d\nu_k}\{\frac{u_i^{\frac{\nu_k+D}{2}-1}\exp(-\frac{u_i\nu_k}{2})(\frac{\nu_k+d}{2})^{\frac{\nu_k+D}{2}}}{\Gamma(\frac{\nu_k+D}{2})T(\sqrt{\frac{D+\nu_k}{d+\nu_k}}A,\nu_k+D)}\}du_i \\
&= \frac{1}{2}E(\log u_i|x_i) - \frac{1}{2}E(u_i|x_i) + \frac{1}{2}[\log(\frac{\nu_k+d}{2}) + \frac{\nu_k+D}{\nu_k+d}] - \frac{1}{2}\frac{\Gamma'(\frac{\nu_k+d}{2})}{\Gamma(\frac{\nu_k+d}{2})} \\
&\quad - \frac{1}{2}\frac{t_{\nu_k+D}(\sqrt{\frac{D+\nu_k}{d+\nu_k}}A)}{T(\sqrt{\frac{D+\nu_k}{d+\nu_k}}A,\nu_k+D)}A(\frac{d+\nu_k}{D+\nu_k})^{1/2}(\frac{d-D}{(d+\nu_k)^2}) \\
&\quad - \frac{1}{2}T^{-1}(\sqrt{\frac{D+\nu_k}{d+\nu_k}}A,\nu_k+D) \times \int_{-\infty}^{\sqrt{\frac{D+\nu_k}{d+\nu_k}}A} g(p,\nu_k+D)dp
\end{aligned}
\tag{13}
$$

Then, we have

$$
\begin{aligned}
E(\log u_i) &= E(u_i|x_i) - (\log\frac{\nu_k+d}{2} + \frac{\nu_k+D}{\nu_k+d} - \psi(\frac{\nu_k+D}{2})) \\
&\quad + (\frac{t_D(\sqrt{\frac{D+\nu_k}{d+\nu_k}}A,\nu_k+D)}{T(\sqrt{\frac{D+\nu_k}{d+\nu_k}}A,\nu_k+D)} \times \frac{A(d-D)}{\sqrt{(D+\nu_k)(d+\nu_k)^3}}) \\
&\quad + T^{-1}(\sqrt{\frac{D+\nu_k}{d+\nu_k}}A,\nu_k+D) \times \int_{-\infty}^{\sqrt{\frac{D+\nu_k}{d+\nu_k}}A} g(p,\nu_k+D)dp
\end{aligned}
\tag{14}
$$

where

$$
\begin{aligned}
g(p,\nu_k+D) &= (\psi(\frac{\nu_k+2D}{2}) - \psi(\frac{\nu_k+D}{2}) - \log(\frac{p^2+\nu_k+D}{\nu_k+D}) \\
&\quad + \frac{p^2(\nu_k+D)-D(\nu_k+D)}{(p^2+\nu_k+D)(\nu_k+D)}) \cdot t_{\nu_k+D}(p)
\end{aligned}
\tag{15}
$$

References

[1] T. I. Lin, J. C. Lee, and W. J. Hsieh, "Robust mixture models using the skew-t distribution," *Statistics and Computing*, vol. 17, no. 2, pp. 81–92, 2007.