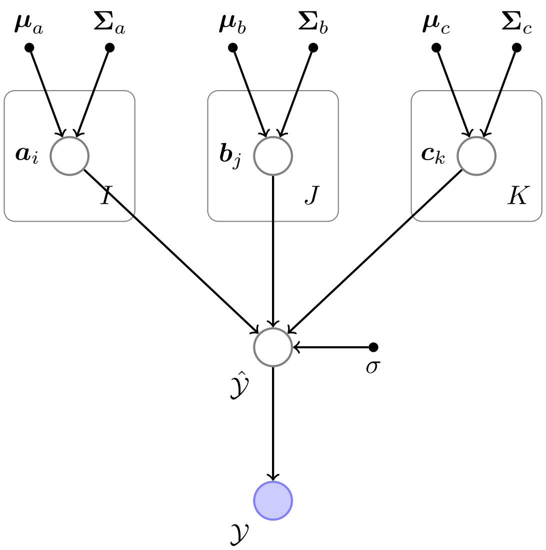

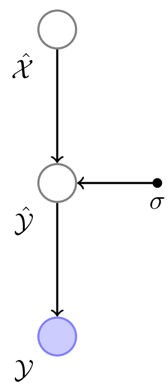

1 probability Model

According *Canonical polyadic(CP) Tensor Fatorization*, the approximation of the tensor $\mathcal{X} \in \mathbb{R}^{N \times M \times L}$ of mode 3 is given by

$$\mathcal{X} \approx \hat{\mathcal{X}} = \sum_{r=1}^{R} \boldsymbol{a}_{:,r} \circ \boldsymbol{b}_{:}, r \circ \boldsymbol{c}_{:,r}. \tag{1}$$

Assuming that each $\boldsymbol{a}_i \in \mathbb{R}^{N \times 1}$ has a Gaussian distribution with a mean $\boldsymbol{\mu}_a \in \mathbb{R}^{N \times 1}$ and a covariance matrix $\boldsymbol{\Sigma}_a \in \mathbb{R}^{N \times N}$, the conditional distribution of $\boldsymbol{a}_i$ is

$$p(\boldsymbol{a}_i | \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) = \mathcal{N}(\boldsymbol{a}_i | \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a). \tag{2}$$

And the conjugate prior probability of $\boldsymbol{\mu}_a$ and $\boldsymbol{\Sigma}_a$ is

$$p(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) = p(\boldsymbol{\Sigma}_a)p(\boldsymbol{\mu}_a|\boldsymbol{\Sigma}_A) = NIW(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a|\boldsymbol{m}_0^a, \kappa_0^a, \nu_0^a, \boldsymbol{S}_0^a), \tag{3}$$
$$p(\boldsymbol{\mu}_a|\boldsymbol{\Sigma}_a) = \mathcal{N}(\boldsymbol{\mu}_a|\boldsymbol{m}_0^a, \frac{1}{\kappa_0^a}\boldsymbol{\Sigma}_a), \tag{4}$$
$$p(\boldsymbol{\Sigma}_a) = IW(\boldsymbol{\Sigma}_a|\nu_0^a, \boldsymbol{S}_0^a). \tag{5}$$

Similarly, for $\boldsymbol{b}_j \in \mathbb{R}^{M\times 1}$, we have

$$p(\boldsymbol{b}_j|\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b) = \mathcal{N}(\boldsymbol{b}_j|\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b) \tag{6}$$
$$p(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b) = NIW(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b|\boldsymbol{m}_0^\beta, \kappa_0^\beta, \nu_0^\beta, \boldsymbol{S}_0^\beta), \tag{7}$$
$$p(\boldsymbol{\mu}_b|\boldsymbol{\Sigma}_b) = \mathcal{N}(\boldsymbol{\mu}_b|\boldsymbol{m}_0^\beta, \frac{1}{\kappa_0^\beta}\boldsymbol{\Sigma}_b), \tag{8}$$
$$p(\boldsymbol{\Sigma}_b) = IW(\boldsymbol{\Sigma}_b|\nu_0^\beta, \boldsymbol{S}_0^\beta), \tag{9}$$

and for $\boldsymbol{c}_k \in \mathbb{R}^{L\times 1}$, we have

$$p(\boldsymbol{c}_k|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) = \mathcal{N}(\boldsymbol{c}_j|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \tag{10}$$
$$p(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) = NIW(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c|\boldsymbol{m}_0^c, \kappa_0^c, \nu_0^c, \boldsymbol{S}_0^c), \tag{11}$$
$$p(\boldsymbol{\mu}_c|\boldsymbol{\Sigma}_c) = \mathcal{N}(\boldsymbol{\mu}_c|\boldsymbol{m}_0^c, \frac{1}{\kappa_0^c}\boldsymbol{\Sigma}_c), \tag{12}$$
$$p(\boldsymbol{\Sigma}_c) = IW(\boldsymbol{\Sigma}_c|\nu_0^c, \boldsymbol{S}_0^c). \tag{13}$$

For each entry $\hat{y}_{nml}$ of the tensor $\hat{\mathcal{Y}}$, we have the conditional distribution

$$p(\hat{y}_{nml}|\hat{x}_{nml}, \sigma^2) = \mathcal{N}(\hat{y}_{nml}|\hat{x}_{nml}, \sigma^2), \tag{14}$$

and $\hat{x}_{nml}$ is the entry of $\hat{\mathcal{X}}$.
Further, for each entry y_{nml} of the tensor $\mathcal{Y}$, we have the conditional distribution

$$p(y_{nml}|\hat{y}_{nml}\, is\, observed) = \theta; \tag{15}$$

And

$$\begin{aligned} p(\hat{x}_{nml} = h|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a, \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b, \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) &= \frac{dF_{\hat{x}_{nml}}(h)}{dh} \\ &= \frac{\int_{(\boldsymbol{a}_n \odot \boldsymbol{b}_m \odot \boldsymbol{c}_l)\boldsymbol{1} \le h} \mathcal{N}(\boldsymbol{a}_n|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)\mathcal{N}(\boldsymbol{b}_m|\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b)\mathcal{N}(\boldsymbol{c}_l|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)d\boldsymbol{a}_n d\boldsymbol{b}_m d\boldsymbol{c}_l}{dh} \end{aligned} \tag{16}$$

$$
\begin{aligned}
&p(\hat{\mathcal{Y}}, \hat{\mathcal{X}}, \sigma^2, \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a, \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b, \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \\
&= p(\hat{\mathcal{Y}}|\hat{\mathcal{X}}) p(\hat{\mathcal{X}}|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a, \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b, \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) p(\sigma^2) p(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) p(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b) p(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \\
&= \prod_{n,m,l} \mathcal{N}(\hat{y}_{nml}|\hat{x}_{nml}, \sigma^2) \prod_{n,m,l} p(\hat{x}_{nml}|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a, \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b, \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) Gam(\frac{1}{\sigma^2}|a_0, \beta_0) \\
&\quad \times NIW(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a|\boldsymbol{m}_0^a, \kappa_0^a, \nu_0^a, \boldsymbol{S}_0^a) \\
&\quad \times NIW(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b|\boldsymbol{m}_0^\beta, \kappa_0^\beta, \nu_0^\beta, \boldsymbol{S}_0^\beta) \\
&\quad \times NIW(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c|\boldsymbol{m}_0^c, \kappa_0^c, \nu_0^c, \boldsymbol{S}_0^c)
\end{aligned} \tag{17}
$$

May be wrong! (Analysis without the node $\mathcal{Y}$)
Then, the joint distribution of the complete data is

$$\begin{aligned}
& p(\hat{\mathcal{Y}}, \hat{\mathcal{X}}, \sigma^2, \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a, \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b, \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \\
&= p(\hat{\mathcal{Y}}, \{\boldsymbol{a}_i\}, \{\boldsymbol{b}_j\}, \{\boldsymbol{c}_k\}, \sigma^2, \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a, \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b, \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \\
&= p(\hat{\mathcal{Y}}|\hat{\mathcal{X}}, \sigma^2)p(\sigma^2)p(\{\boldsymbol{a}_i\}|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)p(\{\boldsymbol{b}_j\}|\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b)p(\{\boldsymbol{c}_k\}|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)p(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)p(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b)p(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \\
&= \prod_{n,m,l} \mathcal{N}(\hat{y}_{nml}|\hat{x}_{nml}, \sigma^2) Gam(\frac{1}{\sigma^2}|a_0, \beta_0) \\
&\quad \times \prod_i \mathcal{N}(\boldsymbol{a}_i|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) \prod_j \mathcal{N}(\boldsymbol{b}_j|\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b) \prod_k \mathcal{N}(\boldsymbol{c}_k|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \\
&\quad \times NIW(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a|\boldsymbol{m}_0^a, \kappa_0^a, \nu_0^a, \boldsymbol{S}_0^a) \\
&\quad \times NIW(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b|\boldsymbol{m}_0^\beta, \kappa_0^\beta, \nu_0^\beta, \boldsymbol{S}_0^\beta) \\
&\quad \times NIW(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c|\boldsymbol{m}_0^c, \kappa_0^c, \nu_0^c, \boldsymbol{S}_0^c)
\end{aligned} \tag{18}$$

For $\{\boldsymbol{a}_i\}$, the posterior probability is

$$\begin{aligned}
& p(\{\boldsymbol{a}_i\}|\hat{\mathcal{Y}}, \{\boldsymbol{b}_j\}, \{\boldsymbol{c}_k\}, \sigma^2, \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a, \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b, \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \\
&= p(\{\boldsymbol{a}_i\}|\hat{\mathcal{Y}}, \{\boldsymbol{b}_j\}, \{\boldsymbol{c}_k\}, \sigma^2, \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) \\
&\propto \prod_{n,m,l} \mathcal{N}(\hat{y}_{nml}|\hat{x}_{nml}, \sigma^2) \prod_i \mathcal{N}(\boldsymbol{a}_i|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)
\end{aligned} \tag{19}$$

Similarly,

$$\begin{aligned}
& p(\{\boldsymbol{b}_j\}|\hat{\mathcal{Y}}, \{\boldsymbol{a}_i\}, \{\boldsymbol{c}_k\}, \sigma^2, \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b) \\
&\propto \prod_{n,m,l} \mathcal{N}(\hat{y}_{nml}|\hat{x}_{nml}, \sigma^2) \prod_j \mathcal{N}(\boldsymbol{b}_j|\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b)
\end{aligned} \tag{20}$$

$$\begin{aligned}
& p(\{\boldsymbol{c}_k\}|\hat{\mathcal{Y}}, \{\boldsymbol{a}_i\}, \{\boldsymbol{b}_j\}, \sigma^2, \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \\
&\propto \prod_{n,m,l} \mathcal{N}(\hat{y}_{nml}|\hat{x}_{nml}, \sigma^2) \prod_k \mathcal{N}(\boldsymbol{c}_k|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)
\end{aligned} \tag{21}$$

The posterior probabilities of the other parameters are

$$p(\sigma^2|\hat{\mathcal{Y}},\{\boldsymbol{a}_i\},\{\boldsymbol{b}_j\},\{\boldsymbol{c}_k\}) \propto \prod_{n,m,l} \mathcal{N}(\hat{y}_{nml}|\hat{x}_{nml},\sigma^2) Gam(\frac{1}{\sigma^2}|a_0,\beta_0) \tag{22}$$

$$p(\boldsymbol{\mu}_a,\boldsymbol{\Sigma}_a|\hat{\mathcal{Y}},\{\boldsymbol{a}_i\}) \propto \prod_{n,m,l} \mathcal{N}(\hat{y}_{nml}|\hat{x}_{nml},\sigma^2) \prod_i \mathcal{N}(\boldsymbol{a}_i|\boldsymbol{\mu}_a,\boldsymbol{\Sigma}_a) \tag{23}$$

$$\times NIW(\boldsymbol{\mu}_a,\boldsymbol{\Sigma}_a|\boldsymbol{m}_0^a,\kappa_0^a,\nu_0^a,\boldsymbol{S}_0^a) \tag{24}$$

$$p(\boldsymbol{\mu}_b,\boldsymbol{\Sigma}_b|\hat{\mathcal{Y}},\{\boldsymbol{b}_j\}) \propto \prod_{n,m,l} \mathcal{N}(\hat{y}_{nml}|\hat{x}_{nml},\sigma^2) \prod_j \mathcal{N}(\boldsymbol{b}_j|\boldsymbol{\mu}_b,\boldsymbol{\Sigma}_b) \tag{25}$$

$$\times NIW(\boldsymbol{\mu}_b,\boldsymbol{\Sigma}_b|\boldsymbol{m}_0^\beta,\kappa_0^\beta,\nu_0^\beta,\boldsymbol{S}_0^\beta) \tag{26}$$

$$p(\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c|\hat{\mathcal{Y}},\{\boldsymbol{c}_k\}) \propto \prod_{n,m,l} \mathcal{N}(\hat{y}_{nml}|\hat{x}_{nml},\sigma^2) \prod_k \mathcal{N}(\boldsymbol{c}_k|\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c) \tag{27}$$

$$\times NIW(\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c|\boldsymbol{m}_0^c,\kappa_0^c,\nu_0^c,\boldsymbol{S}_0^c) \tag{28}$$

2 Inference using Ep method

$$\begin{aligned} q(\boldsymbol{\theta}) = &\prod_n \mathcal{N}(\boldsymbol{a}_n|\boldsymbol{\mu}_n^a,\boldsymbol{\Sigma}_n^a) \prod_m \mathcal{N}(\boldsymbol{b}_m|\boldsymbol{\mu}_m^b,\boldsymbol{\Sigma}_m^b) \prod_l \mathcal{N}(\boldsymbol{c}_l|\boldsymbol{\mu}_l^c,\boldsymbol{\Sigma}_l^c) \\ &\times Gam(\frac{1}{\sigma^2}|a_0,\beta_0) \\ &\times NIW(\boldsymbol{\mu}_a,\boldsymbol{\Sigma}_a|\boldsymbol{m}_0^a,\kappa_0^a,\nu_0^a,\boldsymbol{S}_0^a) \\ &\times NIW(\boldsymbol{\mu}_b,\boldsymbol{\Sigma}_b|\boldsymbol{m}_0^\beta,\kappa_0^\beta,\nu_0^\beta,\boldsymbol{S}_0^\beta) \\ &\times NIW(\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c|\boldsymbol{m}_0^c,\kappa_0^c,\nu_0^c,\boldsymbol{S}_0^c) \end{aligned} \tag{29}$$

Take $\boldsymbol{a}_n$ for example,

$$\begin{aligned} q^{\backslash nml}(\boldsymbol{\theta}) = &\prod_{n'\neq n} \mathcal{N}(\boldsymbol{a}_{n'}|\boldsymbol{\mu}_{n'}^a,\boldsymbol{\Sigma}_{n'}^a) \prod_{m'\neq m} \mathcal{N}(\boldsymbol{b}_{m'}|\boldsymbol{\mu}_{m'}^b,\boldsymbol{\Sigma}_{m'}^b) \prod_{l'\neq l} \mathcal{N}(\boldsymbol{c}_{l'}|\boldsymbol{\mu}_{l'}^c,\boldsymbol{\Sigma}_{l'}^c) \\ &\times Gam(\frac{1}{\sigma^2}|a_0,\beta_0) \\ &\times NIW(\boldsymbol{\mu}_a,\boldsymbol{\Sigma}_a|\boldsymbol{m}_0^a,\kappa_0^a,\nu_0^a,\boldsymbol{S}_0^a) \\ &\times NIW(\boldsymbol{\mu}_b,\boldsymbol{\Sigma}_b|\boldsymbol{m}_0^\beta,\kappa_0^\beta,\nu_0^\beta,\boldsymbol{S}_0^\beta) \\ &\times NIW(\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c|\boldsymbol{m}_0^c,\kappa_0^c,\nu_0^c,\boldsymbol{S}_0^c) \end{aligned} \tag{30}$$

Then,

$$
\begin{aligned}
Z_{nml} &= \int f(\boldsymbol{a}_n, \boldsymbol{b}_m, \boldsymbol{c}_l) q^{\setminus nml}(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \int \mathcal{N}(y_{nml}|x_{nml}, \sigma^2)\mathcal{N}(\boldsymbol{a}_n|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)\mathcal{N}(\boldsymbol{b}_m|\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b)\mathcal{N}(\boldsymbol{c}_l|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \\
&\quad \times \prod_{n'\neq n} \mathcal{N}(\boldsymbol{a}_{n'}|\boldsymbol{\mu}^a_{n'}, \boldsymbol{\Sigma}^a_{n'}) \prod_{m'\neq m} \mathcal{N}(\boldsymbol{b}_{m'}|\boldsymbol{\mu}^b_{m'}, \boldsymbol{\Sigma}^b_{m'}) \prod_{l'\neq l} \mathcal{N}(\boldsymbol{c}_{l'}|\boldsymbol{\mu}^c_{l'}, \boldsymbol{\Sigma}^c_{l'}) \\
&\quad \times Gam(\frac{1}{\sigma^2}|a_0, \beta_0) \\
&\quad \times NIW(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a|\boldsymbol{m}^a_0, \kappa^a_0, \nu^a_0, \boldsymbol{S}^a_0) \\
&\quad \times NIW(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b|\boldsymbol{m}^\beta_0, \kappa^\beta_0, \nu^\beta_0, \boldsymbol{S}^\beta_0) \\
&\quad \times NIW(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c|\boldsymbol{m}^c_0, \kappa^c_0, \nu^c_0, \boldsymbol{S}^c_0) \\
&\quad d\{\boldsymbol{a}_n\}d\{\boldsymbol{b}_m\}d\{\boldsymbol{c}_l\}d\frac{1}{\sigma^2}d\boldsymbol{\mu}_a d\boldsymbol{\Sigma}_a \boldsymbol{\mu}_b d\boldsymbol{\Sigma}_b d\boldsymbol{\mu}_c d\boldsymbol{\Sigma}_c
\end{aligned} \tag{31}
$$

Split the integral into different terms.

$$
\begin{aligned}
&\mathcal{N}(y_{nml}|x_{nml}, \sigma^2)\mathcal{N}(\boldsymbol{a}_n|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) \\
&= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_{nml} - \boldsymbol{a}_n(\boldsymbol{b}_m \odot \boldsymbol{c}_l)^T)^2\right) (2\pi)^{-\frac{R}{2}}|\boldsymbol{\Sigma}_a|^{-1} \exp\left(-\frac{1}{2}(\boldsymbol{a}_n - \boldsymbol{\mu}_n)\boldsymbol{\Sigma}_a^{-1}(\boldsymbol{a}_n - \boldsymbol{\mu}_n)^T\right) \\
&= (2\pi)^{-\frac{R+1}{2}}(\sigma)^{-1}|\boldsymbol{\Sigma}_a|^{-1} \\
&\quad \times \exp[-\frac{1}{2}(\boldsymbol{a}_n(\frac{1}{\sigma^2}(\boldsymbol{b}_m \odot \boldsymbol{c}_l)^T(\boldsymbol{b}_m \odot \boldsymbol{c}_l) + \boldsymbol{\Sigma}_a^{-1})\boldsymbol{a}_n^T \\
&\quad - \boldsymbol{a}_n(\frac{1}{\sigma^2}(\boldsymbol{b}_m \odot \boldsymbol{c}_l)^T y_{nml} + \boldsymbol{\Sigma}_a^{-1}\boldsymbol{\mu}^T) - (\frac{1}{\sigma^2}(\boldsymbol{b}_m \odot \boldsymbol{c}_l)y_{nml} + \boldsymbol{\mu}\boldsymbol{\Sigma}_a^{-1})\boldsymbol{a}_n^T \\
&\quad + \frac{1}{\sigma^2}y^2_{nml} + \boldsymbol{\mu}_a\boldsymbol{\Sigma}_a\boldsymbol{\mu}_a^T)] \\
&= (2\pi)^{-\frac{1}{2}}(\sigma^2)^{-\frac{1}{2}}|\boldsymbol{\Sigma}_a|^{-1}|\boldsymbol{\Sigma}'|^{\frac{1}{2}}\mathcal{N}(\boldsymbol{a}_n|\boldsymbol{\mu}', \boldsymbol{\Sigma}') \exp(-\frac{1}{2}(\frac{1}{\sigma^2}y^2_{nml} + \boldsymbol{\mu}_a\boldsymbol{\Sigma}_a^{-1}\boldsymbol{\mu}_a^T - \boldsymbol{\mu}'\boldsymbol{\Sigma}'^{-1}\boldsymbol{\mu}'^T))
\end{aligned} \tag{32}
$$

where (assuming $\boldsymbol{\Sigma}'$ invertible)

$$
\boldsymbol{\Sigma}' = (\frac{1}{\sigma^2}(\boldsymbol{b}_m \odot \boldsymbol{c}_l)^T(\boldsymbol{b}_m \odot \boldsymbol{c}_l) + \boldsymbol{\Sigma}_a^{-1})^{-1} \tag{33}
$$

$$
\boldsymbol{\mu}' = (\frac{1}{\sigma^2}(\boldsymbol{b}_m \odot \boldsymbol{c}_l)y_{nml} + \boldsymbol{\mu}_a\boldsymbol{\Sigma}_a^{-1})\boldsymbol{\Sigma}' \tag{34}
$$

For the exponential part, we have

$$
\begin{aligned}
&\frac{1}{\sigma^2} y_{nml}^2 + \boldsymbol{\mu}_a \boldsymbol{\Sigma}_a^{-1} \boldsymbol{\mu}_a^T - \boldsymbol{\mu}' \boldsymbol{\Sigma}'^{-1} \boldsymbol{\mu}'^T \\
&= \frac{1}{\sigma^2} y_{nml}^2 + \boldsymbol{\mu}_a \boldsymbol{\Sigma}_a^{-1} \boldsymbol{\mu}_a^T \\
&- (\frac{1}{\sigma^2} (\boldsymbol{b}_m \odot \boldsymbol{c}_l) y_{nml} + \boldsymbol{\mu}_a \boldsymbol{\Sigma}_a^{-1})(\frac{1}{\sigma^2} (\boldsymbol{b}_m \odot \boldsymbol{c}_l)^T (\boldsymbol{b}_m \odot \boldsymbol{c}_l) + \boldsymbol{\Sigma}_a^{-1})^{-1} (\frac{1}{\sigma^2} (\boldsymbol{b}_m \odot \boldsymbol{c}_l) y_{nml} + \boldsymbol{\mu}_a \boldsymbol{\Sigma}_a^{-1})^T \\
&= \frac{1}{\sigma^2} y_{nml}^2 + \boldsymbol{\mu}_a \boldsymbol{\Sigma}_a^{-1} \boldsymbol{\mu}_a^T - (\frac{1}{\sigma^2} \boldsymbol{v} y_{nml} + \boldsymbol{\mu}_a \boldsymbol{\Sigma}_a^{-1})(\frac{1}{\sigma^2} \boldsymbol{v}^T \boldsymbol{v} + \boldsymbol{\Sigma}_a^{-1})^{-1} (\frac{1}{\sigma^2} \boldsymbol{v} y_{nml} + \boldsymbol{\mu}_a \boldsymbol{\Sigma}_a^{-1})^T
\end{aligned}
\tag{35}
$$

where

$$
\boldsymbol{v} = \boldsymbol{b}_m \odot \boldsymbol{c}_l \tag{36}
$$

So far, the integral is intractable. Adopt some strategies to make approximations. Simplify the model, make $\boldsymbol{a}_n$, $\boldsymbol{b}_m$ and $\boldsymbol{c}_l$ have standard normal distributions, namely $\boldsymbol{\mu}_a = \boldsymbol{\mu}_b = \boldsymbol{\mu}_c = \boldsymbol{0}$ and $\boldsymbol{\Sigma}_a = \boldsymbol{\Sigma}_b = \boldsymbol{\Sigma}_c = \boldsymbol{I}$. More importantly,(improper description) when integrate out the concerned variable, assume the other related variables as observed, i.e. when integrate out the $\boldsymbol{a}_n$, assume $\boldsymbol{b}_m$, $\boldsymbol{c}_l$ and σ are observed. Then when calculate the moments of $\boldsymbol{a}_n$,

$$
\begin{aligned}
Z_{nml} &= (2\pi)^{-\frac{1}{2}} (\sigma^2)^{-\frac{1}{2}} |I + \frac{1}{\sigma^2} \boldsymbol{v}^T \boldsymbol{v}|^{-\frac{1}{2}} \\
&\quad \times \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2} y_{nml}^2 - \frac{1}{\sigma^2} \boldsymbol{v} y_{nml} (I + \frac{1}{\sigma^2} \boldsymbol{v}^T \boldsymbol{v})^{-1} \frac{1}{\sigma^2} \boldsymbol{v}^T y_{nml}\right)\right] \\
&\quad \times \mathcal{N}(\boldsymbol{b}_m | \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b) \mathcal{N}(\boldsymbol{c}_l | \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) Gam(\frac{1}{\sigma^2} | a_0, \beta_0) \\
&= (2\pi)^{-\frac{1}{2}} (\sigma^2)^{-\frac{1}{2}} (1 + \frac{1}{\sigma^2} \boldsymbol{v} \boldsymbol{v}^T)^{-\frac{1}{2}} \exp(-\frac{1}{2} y_{nml}^2 (\sigma^2 + \boldsymbol{v} \boldsymbol{v}^T)^{-1}) \\
&\quad \times \mathcal{N}(\boldsymbol{b}_m | \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b) \mathcal{N}(\boldsymbol{c}_l | \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) Gam(\frac{1}{\sigma^2} | a_0, \beta_0) \\
&= \frac{1}{\sqrt{2\pi}} (\sigma^2 + \boldsymbol{v} \boldsymbol{v}^T)^{-\frac{1}{2}} \exp(-\frac{1}{2} y_{nml}^2 (\sigma^2 + \boldsymbol{v} \boldsymbol{v}^T)^{-1}) \\
&\quad \times \mathcal{N}(\boldsymbol{b}_m | \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b) \mathcal{N}(\boldsymbol{c}_l | \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) Gam(\frac{1}{\sigma^2} | a_0, \beta_0) \\
&= \mathcal{N}(y_{nml} | 0, \sigma^2 + \boldsymbol{v} \boldsymbol{v}^T) \mathcal{N}(\boldsymbol{b}_m | \boldsymbol{0}, \boldsymbol{I}) \mathcal{N}(\boldsymbol{c}_l | \boldsymbol{0}, \boldsymbol{I}) Gam(\frac{1}{\sigma^2} | \alpha_0, \beta_0)
\end{aligned}
\tag{37}
$$

During the derivation, we use the *Woodbury Matrix Identity*(from right to left) which is

$$
(\boldsymbol{A} + \boldsymbol{U}\boldsymbol{C}\boldsymbol{V})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{C}^{-1} + \boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U})\boldsymbol{V}\boldsymbol{A}^{-1} \tag{38}
$$

And the approximated variance and expectation of $\boldsymbol{a}_n$ is

$$
\boldsymbol{\Sigma}' = (\frac{1}{\sigma^2} \boldsymbol{v}^T \boldsymbol{v} + \boldsymbol{I})^{-1} \tag{39}
$$

$$
\boldsymbol{\mu}' = \frac{y_{nml}}{\sigma^2} \boldsymbol{v} (\frac{1}{\sigma^2} \boldsymbol{v}^T \boldsymbol{v} + \boldsymbol{I})^{-1} \tag{40}
$$

Similarly, when integrating out $\boldsymbol{b}_m$ and $\boldsymbol{c}_l$,

$$\boldsymbol{v}_{\boldsymbol{b}_m} = \boldsymbol{a}_n \odot \boldsymbol{c}_l \tag{41}$$

$$Z_{nml}^{\boldsymbol{b}_m} = \mathcal{N}(y_{nml}|0, \sigma^2 + \boldsymbol{v}_{\boldsymbol{b}_m}^T \boldsymbol{v}_{\boldsymbol{b}_m})\mathcal{N}(\boldsymbol{a}_n|\boldsymbol{0}, \boldsymbol{I})\mathcal{N}(\boldsymbol{c}_l|\boldsymbol{0}, \boldsymbol{I})Gam(\frac{1}{\sigma^2}|\alpha_0, \beta_0) \tag{42}$$

$$\boldsymbol{\Sigma}'_{\boldsymbol{b}_m} = (\frac{1}{\sigma^2}\boldsymbol{v}_{\boldsymbol{b}_m}^T \boldsymbol{v}_{\boldsymbol{b}_m} + \boldsymbol{I})^{-1} \tag{43}$$

$$\boldsymbol{\mu}'_{\boldsymbol{b}_m} = \frac{y_{nml}}{\sigma^2}\boldsymbol{v}_{\boldsymbol{b}_m}(\frac{1}{\sigma^2}\boldsymbol{v}_{\boldsymbol{b}_m}^T \boldsymbol{v}_{\boldsymbol{b}_m} + \boldsymbol{I})^{-1} \tag{44}$$

$$\boldsymbol{v}_{\boldsymbol{c}_l} = \boldsymbol{a}_n \odot \boldsymbol{b}_m \tag{45}$$

$$Z_{nml}^{\boldsymbol{c}_l} = \mathcal{N}(y_{nml}|0, \sigma^2 + \boldsymbol{v}_{\boldsymbol{c}_l}^T \boldsymbol{v}_{\boldsymbol{c}_l})\mathcal{N}(\boldsymbol{a}_n|\boldsymbol{0}, \boldsymbol{I})\mathcal{N}(\boldsymbol{b}_m|\boldsymbol{0}, \boldsymbol{I})Gam(\frac{1}{\sigma^2}|\alpha_0, \beta_0) \tag{46}$$

$$\boldsymbol{\Sigma}'_{\boldsymbol{c}_l} = (\frac{1}{\sigma^2}\boldsymbol{v}_{\boldsymbol{c}_l}^T \boldsymbol{v}_{\boldsymbol{c}_l} + \boldsymbol{I})^{-1} \tag{47}$$

$$\boldsymbol{\mu}'_{\boldsymbol{c}_l} = \frac{y_{nml}}{\sigma^2}\boldsymbol{v}_{\boldsymbol{c}_l}(\frac{1}{\sigma^2}\boldsymbol{v}_{\boldsymbol{c}_l}^T \boldsymbol{v}_{\boldsymbol{c}_l} + \boldsymbol{I})^{-1} \tag{48}$$

When integrating out $\frac{1}{\sigma^2}$,

$$Z_{nml}^{\frac{1}{\sigma^2}} = \frac{\Gamma(\alpha_0 + \frac{1}{2})\beta_0^{\alpha_0}}{\sqrt{2\pi}\Gamma(\alpha_0)[\beta_0 + \frac{1}{2}(y_{nml} - \boldsymbol{a}_n(\boldsymbol{b}_m \odot \boldsymbol{c}_l)^T)^2]^{\alpha_0 + \frac{1}{2}}} \tag{49}$$

$$\mu_{\frac{1}{\sigma^2}} = \frac{\alpha_0 + \frac{1}{2}}{\beta_0 + \frac{1}{2}(y_{nml} - \boldsymbol{a}_n(\boldsymbol{b}_m \odot \boldsymbol{c}_l)^T)^2} \tag{50}$$

$$var_{\frac{1}{\sigma^2}} = \frac{\alpha_0 + \frac{1}{2}}{[\beta_0 + \frac{1}{2}(y_{nml} - \boldsymbol{a}_n(\boldsymbol{b}_m \odot \boldsymbol{c}_l)^T)^2]^2} \tag{51}$$