Robust Multi-view Subspace Learning with Non-identical and Non-independent Distributed Complex Noises

Zongsheng Yue, Hongwei Yong, Deyu Meng, *Member, IEEE*, Qian Zhao, Yee Leung, and Lei Zhang, *Fellow, IEEE*

***Abstract*—In this supplementary material, we provide more details on the computations involved in the proposed variational inference algorithm and more experiment results.**

I. NIID-MSL Model

A. Model Formulation

Basically, we decompose the observed data into:

$$\boldsymbol{X}^v = \boldsymbol{S}^v + \boldsymbol{E}^v, \tag{1}$$

where $\boldsymbol{E}^v = \{e_{ij}^v\}_{d\times n}$ denotes the residual term (i.e., noise component) and $\boldsymbol{S}^v \in \mathbb{R}^{d\times n}$ is the expected data located on the latent subspace, d and n represent the dimensionality and the number of samples in each view.

Firstly, we model the noise term $\boldsymbol{E}^v$ as follows:

$$\xi_k \sim \text{Gam}(e_0, f_0), \qquad e_{ij}^v \sim \mathcal{N}(0, (\xi_{c^v_{z^v_{ij}}})^{-1}), \tag{2a}$$

$$c_t^v \sim \text{Multi}(\boldsymbol{\beta}), \qquad z_{ij}^v \sim \text{Multi}(\boldsymbol{\pi^v}), \tag{2b}$$

$$\beta_k = \beta_k^{'}\prod_{l=1}^{k-1}(1-\beta_l^{'}), \qquad \pi_t^v = \pi_t^{v'}\prod_{s=1}^{t-1}(1-\pi_s^{v'}), \tag{2c}$$

$$\beta_k^{'} \sim \text{Beta}(1,\gamma), \qquad \pi_t^{v'} \sim \text{Beta}(1,\alpha^v), \tag{2d}$$

$$\gamma \sim \text{Gam}(m_0, n_0), \qquad \alpha^v \sim \text{Gam}(g_0, h_0). \tag{2e}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are the concentration parameters, which mainly affect the number of Gaussian components of the second-level GMM in each view and the first-level GMM for the entire dataset, respectively.

As for the expected data $\boldsymbol{S}^v$, we embedded each view into a latent space $\boldsymbol{R}$ with a dictionary $\boldsymbol{L}^v$ as conventional MSL methods, i.e.,

$$\boldsymbol{S}^v = \sum_{r=1}^{l}\boldsymbol{L}_{\cdot r}^v\boldsymbol{R}_{r\cdot}, \tag{3a}$$

$$\boldsymbol{R}_{r\cdot} \sim \mathcal{N}(\boldsymbol{0}, \frac{1}{\tau_r}I_n), \tag{3b}$$

$$\boldsymbol{L}_{\cdot r}^v \sim \mathcal{N}(\boldsymbol{0}, \frac{1}{\lambda_r^v}I_d), \tag{3c}$$

$$\lambda_r^v \sim \text{Gam}(a_0, b_0),$$

$$\tau_r \sim \text{Gam}(c_0, d_0). \tag{3d}$$

Combining Eqs. (1) - (3), the goal of our proposed NIID-MSL turns to infer the posteriors of all involved variables:

$$p(\boldsymbol{L},\boldsymbol{R},\boldsymbol{\xi},\boldsymbol{C},\boldsymbol{Z},\boldsymbol{\beta},\boldsymbol{\pi},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\tau},\gamma|\boldsymbol{\mathcal{X}}), \tag{4}$$

where $\boldsymbol{C} = \{c_t^v\}$, $\boldsymbol{Z} = \{z_{ij}^v\}$.

B. Variational Assumption

The full likelihood of the proposed NIID-MSL model is expressed as:

$$\begin{aligned}
&p(\boldsymbol{L},\boldsymbol{R},\boldsymbol{\xi},\boldsymbol{C},\boldsymbol{Z},\boldsymbol{\beta},\boldsymbol{\pi},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\tau},\gamma,\boldsymbol{X})\\
&= p(\boldsymbol{X}|\boldsymbol{L},\boldsymbol{R},\boldsymbol{\xi},\boldsymbol{C},\boldsymbol{Z})p(\boldsymbol{L}|\boldsymbol{\lambda})p(\boldsymbol{\lambda})p(\boldsymbol{R}|\boldsymbol{\tau})p(\boldsymbol{\tau})p(\boldsymbol{\xi})p(\boldsymbol{C}|\boldsymbol{\beta})\\
&\quad p(\boldsymbol{\beta}^{'}|\gamma)p(\gamma)p(\boldsymbol{Z}|\boldsymbol{\pi})p(\boldsymbol{\pi}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})\\
&= \prod_{v,i,j}\prod_{t}\left\{\prod_{k}\mathcal{N}\left(x_{ij}^v\middle|\boldsymbol{L}_{i\cdot}^{v\,T}\boldsymbol{R}_{\cdot j},\xi_k^{-1}\right)^{\boldsymbol{1}[c_t^v=k]}\right\}^{\boldsymbol{1}[z_{ij}^v=t]}\\
&\quad \prod_{v,i,j}\text{Multi}(z_{ij}^v|\boldsymbol{\pi}^v)\prod_{v,t}\text{Multi}(c_t^v|\boldsymbol{\beta})\prod_{k}\text{Gam}(\xi_k|e_0,f_0)\\
&\quad \prod_{v,r}\mathcal{N}\left(\boldsymbol{L}_{\cdot r}^v\middle|\boldsymbol{0},\lambda_r^{v-1}\boldsymbol{I}_m\right)\text{Gam}(\lambda_r^v|a_0,b_0)\\
&\quad \prod_{r}\mathcal{N}\left(\boldsymbol{R}_{r\cdot}\middle|\boldsymbol{0},{\tau_r}^{-1}\boldsymbol{I_n}\right)\text{Gam}(\tau_r|c_0,d_0)\\
&\quad \prod_{v,t}\text{Beta}(\pi_t^{v'}|1,\alpha^v)\prod_{v}\text{Gam}(\alpha^v|m_0,n_0)\\
&\quad \prod_{k}\text{Beta}(\beta_k^{'}|1,\gamma)\text{Gam}(\gamma|g_0,h_0).
\end{aligned} \tag{5}$$

In the main text, we have introduced the variational inference to calculate the posterior of this model and assumed the approximation of posterior have a factorized form as follows:

$$\begin{aligned}
&q(\boldsymbol{L},\boldsymbol{R},\boldsymbol{\xi},\boldsymbol{C},\boldsymbol{Z},\boldsymbol{\beta},\boldsymbol{\pi},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\tau},\gamma)\\
&= \prod_{i=1}^{d}q(\boldsymbol{L}_{i\cdot}^v|\boldsymbol{\mu}_i^v,\boldsymbol{\Sigma}_i^v)\prod_{j=1}^{n}q(\boldsymbol{R}_{\cdot j}|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)\prod_{k=1}^{K}q(\xi_k|e_k,f_k)\\
&\quad \prod_{v=1}^{V}\prod_{i,j}^{d,n}q(z_{ij}^v|\boldsymbol{\rho}_{ij}^v)\prod_{v=1}^{V}q(\alpha^v|m^v,n^v)q(\gamma|g,h)\\
&\quad \prod_{v=1}^{V}\prod_{t=1}^{T}q(c_t^v|\boldsymbol{\varphi}_t^v)q(\pi_t^{v'}|r_t^v,w_t^v)\prod_{k=1}^{K}q(\beta_k^{'}|s_k^1,s_k^2)\\
&\quad \prod_{v=1}^{V}\prod_{r=1}^{l}q(\lambda_r^v|a_r^v,b_r^v)\prod_{r=1}^{l}q(\tau_r|c_r,d_r).
\end{aligned} \tag{6}$$

Next, we give detailed deduction of each factorized distribution involved in posterior of Eq. (6). $E_{\boldsymbol{x}\setminus x_i}[f(\boldsymbol{x})]$ denotes the expectation of $f(\boldsymbol{x})$ on set of $\boldsymbol{x}$ with x_i removed. For notations convenience, we introduced $\boldsymbol{\Theta}$ to denote all the parameters that need to be inferenced, i.e.,

$$\boldsymbol{\Theta} = \{\boldsymbol{L},\boldsymbol{R},\boldsymbol{\xi},\boldsymbol{C},\boldsymbol{Z},\boldsymbol{\beta},\boldsymbol{\pi},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\tau},\gamma,\boldsymbol{X}\}.$$

Infer C and Z:

$$\begin{aligned}
&\ln q(z_{ij}^v)\\
&= E_{\boldsymbol{\Theta}\setminus \boldsymbol{Z}}\big[p(\boldsymbol{L},\boldsymbol{R},\boldsymbol{\xi},\boldsymbol{C},\boldsymbol{Z},\boldsymbol{\beta},\boldsymbol{\pi},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\tau},\boldsymbol{\gamma},\boldsymbol{X})\big] + const\\
&= \sum_t \mathbf{1}[z_{ij}^v = t]\Big\{\sum_k \varphi_{tk}^v E_{\boldsymbol{\Theta}\setminus \boldsymbol{Z}}\Big[\mathcal{N}\Big(x_{ij}^v - \boldsymbol{L}_{i\cdot}^{v\,T}\boldsymbol{R}_{\cdot j}\Big|0,\xi_k^{-1}\Big)\Big]\\
&\qquad + E\big[\ln \pi_t^v\big]\Big\} + const\\
&= \sum_t \mathbf{1}[z_{ij}^v = t]\Big\{\sum_k \varphi_{tk}^v\Big(-\frac{1}{2}\ln 2\pi + \frac{1}{2}E\big[\ln \xi_k\big]\\
&\quad -\frac{1}{2}E\big[\xi_k\big]E\Big[\Big(x_{ij}^v - \boldsymbol{L}_{i\cdot}^{v\,T}\boldsymbol{R}_{j\cdot}\Big)^2\Big]\Big) + E\big[\ln \pi_t^v\big]\Big\} + const,
\end{aligned} \tag{7}$$

$$\begin{aligned}
&\ln q(c_t^v)\\
&= E_{\boldsymbol{\Theta}\setminus \boldsymbol{C}}\big[p(\boldsymbol{L},\boldsymbol{R},\boldsymbol{\xi},\boldsymbol{C},\boldsymbol{Z},\boldsymbol{\beta},\boldsymbol{\pi},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\tau},\boldsymbol{\gamma},\boldsymbol{X})\big] + const\\
&= \sum_k \mathbf{1}[c_t^v = k]\Big\{\sum_{i,j} \rho_{ijt}^v E_{\boldsymbol{\Theta}\setminus \boldsymbol{C}}\Big[\mathcal{N}\Big(x_{ij}^v - \boldsymbol{L}_{i\cdot}^{v\,T}\boldsymbol{R}_{\cdot j}\Big|0,\xi_k^{-1}\Big)\Big]\\
&\qquad + E\big[\ln \beta_k\big]\Big\} + const\\
&= \sum_k \mathbf{1}[c_t^v = k]\Big\{\sum_{i,j} \rho_{ijt}^v\Big(-\frac{1}{2}\ln 2\pi + \frac{1}{2}E\big[\ln \xi_k\big]\\
&\quad -\frac{1}{2}E\big[\xi_k\big]E\Big[\Big(x_{ij}^v - \boldsymbol{L}_{i\cdot}^{v\,T}\boldsymbol{R}_{j\cdot}\Big)^2\Big]\Big) + E\big[\ln \beta_k\big]\Big\} + const,
\end{aligned} \tag{8}$$

Taking the exponential of both sides of Eq. (7), Eq. (8) and normalizing the right side, we obtain

$$q(z_{ij}^v|\boldsymbol{\rho}_{ij}^v) = \text{Multi}(\boldsymbol{\rho_{ij}^v}), \quad q(c_t^v|\boldsymbol{\varphi_t^v}) = \text{Multi}(\boldsymbol{\varphi}_t^v), \tag{9}$$

where

$$\rho_{ijt}^v = \frac{\rho_{ijt}^{v\ \prime}}{\sum_s \rho_{ijs}^{v\ \prime}}, \quad \varphi_{tk}^v = \frac{\varphi_{tk}^{v\ \prime}}{\sum_s \varphi_{ts}^{v\ \prime}}, \tag{10}$$

$$\begin{aligned}
\rho_{ijt}^{v\ \prime} \propto \exp\Big\{&\sum_k \varphi_{tk}^{v\ \prime}\Big(\frac{1}{2}\ln 2\pi + \frac{1}{2}E\big[\ln \xi_k\big]\\
&-\frac{1}{2}E\big[\xi_k\big]E\Big[\Big(x_{ij}^v - \boldsymbol{L}_{i\cdot}^{v\,T}\boldsymbol{R}_{\cdot j}\Big)^2\Big]\Big) + E\big[\ln \pi_t^v\big]\Big\},
\end{aligned} \tag{11}$$

$$\begin{aligned}
\varphi_{tk}^{v\ \prime} \propto \exp\Big\{&\sum_{i,j} \rho_{ijt}^v\Big(\frac{1}{2}\ln 2\pi + \frac{1}{2}E\big[\ln \xi_k\big]\\
&-\frac{1}{2}E\big[\xi_k\big]E\Big[\Big(x_{ij}^v - \boldsymbol{L}_{i\cdot}^{v\,T}\boldsymbol{R}_{\cdot j}\Big)^2\Big]\Big) + E[\ln \beta_k]\Big\}.
\end{aligned} \tag{12}$$

Infer ξ:

$$\begin{aligned}
&\ln q(\xi_k)\\
&= E_{\boldsymbol{\Theta}\setminus \boldsymbol{\xi}}\big[p(\boldsymbol{L},\boldsymbol{R},\boldsymbol{\xi},\boldsymbol{C},\boldsymbol{Z},\boldsymbol{\beta},\boldsymbol{\pi},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\tau},\boldsymbol{\gamma},\boldsymbol{X})\big] + const\\
&= \sum_{v,i,j,t} \rho_{ijt}^v\varphi_{tk}^v E_{\boldsymbol{\Theta}\setminus \boldsymbol{\xi_k}}\Big[\mathcal{N}\Big(x_{ij}^v - \boldsymbol{L}_{i\cdot}^{v\,T}\boldsymbol{R}_{j\cdot}\Big|0,\xi_k^{-1}\Big)\Big]\\
&\qquad + (e_0 - 1)\ln \xi_k - f_0\xi_k\\
&= \Big(\frac{1}{2}\sum_{v,i,j,t} \rho_{ijt}^v\varphi_{tk}^v + e_0 - 1\Big)\ln \xi_k\\
&\quad - \Big\{\frac{1}{2}\sum_{v,i,j,t} \rho_{ijt}^v\varphi_{tk}^v E\Big[\big(x_{ij}^v - \boldsymbol{L}_{i\cdot}^v\boldsymbol{R}_{\cdot j}\big)^2\Big] + f_0\Big\}\xi_k + const,
\end{aligned} \tag{13}$$

Aftering taking exponential of both side of Eq. (13), we have:

$$q(\xi_k|e_k, f_k) = \text{Gam}(\xi_k|e_k, f_k), \tag{14}$$

where

$$e_k = \frac{1}{2}\sum_{v,i,j,t} \rho_{ijt}^v\varphi_{tk}^v + e_0, \tag{15}$$

$$f_k = \frac{1}{2}\sum_{v,i,j,t} \rho_{ijt}^v\varphi_{tk}^v E\Big[\Big(x_{ij}^v - \boldsymbol{L}_{i\cdot}^{v\,T}\boldsymbol{R}_{\cdot j}\Big)^2\Big] + f_0. \tag{16}$$

Infer π' and β':

$$\begin{aligned}
&\ln q(\pi_t^{v\prime})\\
&= E_{\boldsymbol{\Theta}\setminus \boldsymbol{\pi'}}\big[p(\boldsymbol{L},\boldsymbol{R},\boldsymbol{\xi},\boldsymbol{C},\boldsymbol{Z},\boldsymbol{\beta},\boldsymbol{\pi},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\tau},\boldsymbol{\gamma},\boldsymbol{X})\big] + const\\
&= \sum_{i,j} \rho_{ijt}^v \ln \pi_t^v + (E[\alpha^v] - 1)\ln(1 - \pi_t^{v\prime}) + const\\
&= \Big(\sum_{i,j,s=t+1} \rho_{ijs}^v + E[\alpha^v] - 1\Big)\ln(1 - \pi_t^{v\prime})\\
&\qquad + \Big(\sum_{i,j} \rho_{ijt}^v\Big)\ln \pi_t^{v\prime} + const,
\end{aligned} \tag{17}$$

then we take exponential of both side of Eq. (17) and can get:

$$q(\pi_r^{v\prime}|r_t^v, w_t^v) = \text{Beta}(\pi_r^{v\prime}|r_t^v, w_t^v), \tag{18}$$

where

$$r_t^v = \sum_{i,j} \rho_{ijt}^v + 1, \tag{19}$$

$$w_t^v = \sum_{i,j,s=t+1} \rho_{ijs}^v + E[\alpha^v]. \tag{20}$$

Similarly, we have:

$$q(\beta_k^{'}|s_k^1, s_k^2) = \text{Beta}(\beta_k^{'}|s_k^1, s_k^2), \tag{21}$$

where

$$s_k^1 = \sum_{v,t} \varphi_{tk}^v + 1, \tag{22}$$

$$s_k^2 = \sum_{v,t,l=k+1} \varphi_{tl}^v + E[\gamma]. \tag{23}$$

Infer α and γ:

$$\begin{aligned}
&\ln q(\alpha^v)\\
&= E_{\boldsymbol{\Theta}\setminus \boldsymbol{\alpha}}\big[p(\boldsymbol{L},\boldsymbol{R},\boldsymbol{\xi},\boldsymbol{C},\boldsymbol{Z},\boldsymbol{\beta},\boldsymbol{\pi},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\tau},\boldsymbol{\gamma},\boldsymbol{X})\big] + const\\
&= \sum_t \Big((\alpha^v - 1)E[\ln(1 - \pi_t^{v\prime})] + \ln \alpha^v\Big) + (m_0 - 1)\ln \alpha^v\\
&\qquad - n_0\alpha^v + const\\
&= (T + m_0 - 1)\ln \alpha^v - \Big(n_0 - \sum_t E\big[\ln(1 - \pi_t^{v\prime})\big]\Big) + const,
\end{aligned} \tag{24}$$

From Eq. (24), We can easily get the following equations of $\boldsymbol{\alpha}$:

$$q(\alpha^v|m^v, n^v) = \text{Gam}(\alpha^v|m^v, n^v), \tag{25}$$

where

$$m^v = T + m_0, \tag{26}$$

$$n^v = n_0 - \sum_t E\big[\ln(1 - \pi_t^{v\prime})\big]. \tag{27}$$

Similarly, we can update variable γ as follows:

$$q(\gamma|g,h) = \mathrm{Gam}(\gamma|g,h), \tag{28}$$

where

$$g = K + g_0, \tag{29}$$
$$h = h_0 - \sum_k E\left[\ln(1-\beta_k^{'})\right]. \tag{30}$$

Infer $\boldsymbol{L}$ and $\boldsymbol{R}$:

$$\begin{aligned}&\ln q(\boldsymbol{L}_{i\cdot}^v)\\ &= E_{\boldsymbol{\Theta}\backslash \boldsymbol{L}}\left[p(\boldsymbol{L},\boldsymbol{R},\boldsymbol{\xi},\boldsymbol{C},\boldsymbol{Z},\boldsymbol{\beta},\boldsymbol{\pi},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\tau},\boldsymbol{\gamma},\boldsymbol{X})\right] + const\\ &= \sum_{j,t,k}\rho_{ijt}^v\varphi_{tk}^v\left[-\frac{1}{2}E[\xi_k]E\left[\left(x_{ij}^v - \boldsymbol{L}_{i\cdot}^{v\,T}\boldsymbol{R}_{\cdot j}\right)^2\right]\right]\\ &\qquad -\frac{1}{2}\boldsymbol{L}_{i\cdot}^{v\,T}\boldsymbol{\Lambda}_{\boldsymbol{v}}^{\boldsymbol{L}}\boldsymbol{L}_{i\cdot}^v + const,\end{aligned} \tag{31}$$

where $\boldsymbol{\Lambda}_{\boldsymbol{v}}^{\boldsymbol{L}} = diag(E[\boldsymbol{\lambda^v}])$. Taking exponential of both sides of Eq. (31), and normalizing the result, we obtain the posterior distribution of $\boldsymbol{L}_{i\cdot}^v$:

$$q(\boldsymbol{L}_{i\cdot}^v|\boldsymbol{\mu}_i^v,\boldsymbol{\Sigma}_i^v) = \mathcal{N}\left(\boldsymbol{L}_{i\cdot}^v|\boldsymbol{\mu}_i^v,\boldsymbol{\Sigma}_i^v\right), \tag{32}$$

where

$$\boldsymbol{\Sigma_i^v} = \left(\sum_{j,t,k}\rho_{ijt}^v\varphi_{tk}^v E\left[\xi_k\right]E[\boldsymbol{R}_{\cdot j}\boldsymbol{R}_{\cdot j}^T] + \boldsymbol{\Lambda}_v^L\right)^{-1}, \tag{33}$$
$$\boldsymbol{\mu}_i^v = \boldsymbol{\Sigma_i^v}\sum_{j,t,k}\rho_{ijt}^v\varphi_{tk}^v E[\xi_k]x_{ij}^v E[\boldsymbol{R}_{\cdot j}]. \tag{34}$$

Similarly, each column of $\boldsymbol{R}$ is also a Gaussian distribution, i.e.,

$$q(\boldsymbol{R}_{\cdot j}|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j) = \mathcal{N}\left(\boldsymbol{R}_{\cdot j}|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j\right), \tag{35}$$

where

$$\boldsymbol{\Sigma_j} = \left(\sum_{v,i,t,k}\rho_{ijt}^v\varphi_{tk}^v E\left[\xi_k\right]E[\boldsymbol{L}_{i\cdot}^v\boldsymbol{L}_{i\cdot}^{v\,T}] + \boldsymbol{\Lambda}^R\right)^{-1}, \tag{36}$$
$$\boldsymbol{\mu}_j = \boldsymbol{\Sigma_j}\sum_{v,i,t,k}\rho_{ijt}^v\varphi_{tk}^v E[\xi_k]x_{ij}^v E[\boldsymbol{L}_{i\cdot}^v]. \tag{37}$$

and $\boldsymbol{\Lambda}^R = diag(E[\boldsymbol{\tau}])$.

Infer $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$:

$$\begin{aligned}&\ln q(\lambda_r^v)\\ &= E_{\boldsymbol{\Theta}\backslash\boldsymbol{\lambda}}\left[p(\boldsymbol{L},\boldsymbol{R},\boldsymbol{\xi},\boldsymbol{C},\boldsymbol{Z},\boldsymbol{\beta},\boldsymbol{\pi},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\tau},\boldsymbol{\gamma},\boldsymbol{X})\right] + const\\ &= \left(\frac{d}{2}+a_0-1\right)\ln\lambda_r^v - \left(\frac{1}{2}E\left[\boldsymbol{L}_{\cdot r}^{v\,T}\boldsymbol{L}_{\cdot r}^v\right]+b_0\right)\lambda_r^v + const,\end{aligned} \tag{38}$$

Thus, we can get the following Updated equations:

$$q(\lambda_r^v|a_r^v,b_r^v) = \mathrm{Gam}(\lambda_r^v|a_r^v,b_r^v), \tag{39}$$

where

$$a_r^v = \frac{d}{2} + a_0, \tag{40}$$
$$b_r^v = \frac{1}{2}E\left[\boldsymbol{L}_{\cdot r}^{v\,T}\boldsymbol{L}_{\cdot r}^v\right] + b_0. \tag{41}$$

Similarly, we can update $\boldsymbol{\tau}$ as following:

$$q(\tau_r|c_r,d_r) = \mathrm{Gam}(\tau_r|c_r,d_r), \tag{42}$$

where

$$c_r = \frac{n}{2} + c_0, \tag{43}$$
$$d_r = \frac{1}{2}E\left[\boldsymbol{R}_{r\cdot}^T\boldsymbol{R}_{r\cdot}\right] + d_0. \tag{44}$$

C. Calculation of Expectations

The expectation in the variational update equations can be calculated with respect to the current variational distributions, as listed in the followings:

$$E[\xi_k] = \frac{e_k}{f_k}, \tag{45}$$
$$E\left[\ln\xi_k\right] = \psi(e_k) - \ln f_k, \tag{46}$$
$$E\left[\ln\pi_t^{v'}\right] = \psi\left(r_t^v\right) - \psi\left(r_t^v + w_t^v\right), \tag{47}$$
$$E\left[\ln(1-\pi_t^{v'})\right] = \psi\left(w_t^v\right) - \psi\left(r_t^v + w_t^v\right), \tag{48}$$
$$E\left[\ln\pi_t^v\right] = E\left[\ln\pi_t^{v'}\right] + \sum_{s=1}^{t-1}E\left[\ln(1-\pi_s^{v'})\right], \tag{49}$$
$$E\left[\ln\beta_k^{'}\right] = \psi\left(s_k^1\right) - \psi\left(s_k^1 + s_k^2\right), \tag{50}$$
$$E\left[\ln\left(1-\beta_k^{'}\right)\right] = \psi\left(s_k^2\right) - \psi\left(s_k^1 + s_k^2\right), \tag{51}$$
$$E\left[\ln\beta_k\right] = E\left[\ln\beta_k^{'}\right] + \sum_{l=1}^{k-1}E\left[\ln(1-\beta_l^{'})\right], \tag{52}$$
$$E\left[\boldsymbol{L}_{i\cdot}^v\boldsymbol{L}_{i\cdot}^{v\,T}\right] = \boldsymbol{\mu}_i^v\boldsymbol{\mu}_i^{vT} + \boldsymbol{\Sigma}_i^v, \tag{53}$$
$$E\left[\boldsymbol{R}_{\cdot j}\boldsymbol{R}_{\cdot j}^T\right] = \boldsymbol{\mu}_j\boldsymbol{\mu}_j^{\,T} + \boldsymbol{\Sigma}_j, \tag{54}$$
$$E\left[\boldsymbol{L}_{\cdot r}^{v\,T}\boldsymbol{L}_{\cdot r}^v\right] = \sum_i\left(\boldsymbol{\mu}_{ir}^v\right)^2 + \left(\boldsymbol{\Sigma}_i^v\right)_{rr}, \tag{55}$$
$$E\left[\boldsymbol{R}_{r\cdot}^{\,T}\boldsymbol{R}_{r\cdot}\right] = \sum_j\left(\boldsymbol{\mu}_{jr}\right)^2 + \left(\boldsymbol{\Sigma}_j\right)_{rr}, \tag{56}$$
$$\begin{aligned}E\left[\left(x_{ij}^v - \boldsymbol{L}_{i\cdot}^{v\,T}\boldsymbol{R}_{\cdot j}\right)^2\right] &= x_{ij}^{v\,2} - 2x_{ij}^v\boldsymbol{\mu}_i^{vT}\boldsymbol{\mu}_j\\ &+tr\left(E\left[\boldsymbol{L}_{i\cdot}^v\boldsymbol{L}_{i\cdot}^{v\,T}\right]E\left[\boldsymbol{R}_{\cdot j}\boldsymbol{R}_{\cdot j}^T\right]\right).\end{aligned} \tag{57}$$

where $\boldsymbol{\mu}_{ir}^v$ and $\boldsymbol{\mu}_{jr}$ represent the rth element of vector $\boldsymbol{\mu}_i^v$ and $\boldsymbol{\mu}_j$ respectively, $\psi(\cdot)$ is the digamma function defined by $\psi(x) = \frac{d}{dx}\ln\Gamma(x)$

II. Supplementary Expeiments

A. Baseline Methods

In the main text, we assume that the noise of pratical multi-view data is with three characteristics, i.e., complex, non-identical and non-independent. Some previous noise modeling literatures [1]–[3] had proved the effectiveness of complex noise assumption in different real applications, thus we consider the non-identical and non-independent assumptions in these experiments. In order to demonstrate the marginal benefit of improving on these two assumptions, we design two different noise models as baselines compared with the NIID-MSL model, in which the first one fits the noise using one single

TABLE I
RRSE COMPARISON OF NIID-MSL AND TWO BASELINE METHODS ON CMU MULTI-PIE FACE DATASETS WITHOUT ANY SYNTHETIC NOISE. THE BEST RESULTS IN EACH EXPERIMENT ARE HIGHLIGHTED IN RED.

Index	Methods		
	Baseline 1	Baseline 2	NIID-MSL
RRSE	0.0121	0.0106	0.0092
RRAE	0.0680	0.0680	0.0671

TABLE II
F-MEASURE VALUE OF NIID-MSL AND TWO BASELINE METHODS ON WALLFLOWER DATASET. THE BEST RESULTS IN EACH EXPERIMENT ARE HIGHLIGHTED IN RED.

Video	Methods		
	Baseline 1	Baseline 2	NIID-MSL
Bootstrapping	0.7326	0.7325	0.7326
Camouflage	0.7205	0.7239	0.7413
Apertu	0.9593	0.9592	0.9593
SwitchLight	0.6826	0.6804	0.6852
TimeOfDay	0.7641	0.7681	0.7594
WavingTrees	0.7180	0.6647	0.9119
Mean	0.7628	0.7548	0.7982

DPGMM for all the views (complex but i.i.d.) while the second one different DPGMM for each view of data (complex, non-identical, but independent). Since the latent subspace modeling part of these two baselines are the same with the NIID-MSL as shown in Eq. (3), we only list the noise modeling part of them as follows.

Baseline 1:

$$\xi_k \sim \text{Gam}(e_0, f_0), \quad e_{ij}^v \sim \mathcal{N}(0, (\xi_{z_{ij}^v})^{-1}), \tag{58a}$$

$$z_{ij}^v \sim \text{Multi}(\boldsymbol{\pi}), \qquad \pi_k = \pi_k^{'} \prod_{l}^{k-1}(1-\pi_k^{'}), \tag{58b}$$

$$\boldsymbol{\pi}_k^{'} \sim \text{Beta}(1, \gamma), \qquad \gamma \sim \text{Gam}(m_0, n_0). \tag{58c}$$

Baseline 2:

$$\xi_k^v \sim \text{Gam}(e_0, f_0), \quad e_{ij}^v \sim \mathcal{N}(0, (\xi_{z_{ij}^v}^v)^{-1}), \tag{59a}$$

$$z_{ij}^v \sim \text{Multi}(\boldsymbol{\pi}^v), \qquad \pi_k^v = \pi_k^{v'} \prod_{l}^{k-1}(1-\pi_k^{v'}), \tag{59b}$$

$$\boldsymbol{\pi}_k^{v'} \sim \text{Beta}(1, \gamma^v), \qquad \gamma^v \sim \text{Gam}(m_0, n_0). \tag{59c}$$

B. Experimental Results

We compare our proposed NIID-MSL methods with two baselines in Eq. (58) and Eq. (59) to validate the effectiveness of our non-identical and non-independent assumptions on the noise of multi-view data. And some experiments were carried on the real face image recovery ('No noise' case of Table III of main text) and real application of foreground detection on RGB data (part C of Section VI), becasue they can be more representative of the characteristics of noise in practical multi-view data.

Theoretically, Baseline 1 and Baseline 2 are both special cases of our proposed NIID-MSL. As shown in Fig. 2 of the main text, the NIID-MSL degenarated into Baseline 1 when the MoGs in each view of the second-level all share the same Gaussian components from the first-level. On the contrary, if they do not share any same Gaussian component, the NIID-MSL is equivalent to Baseline 2. The Tabel I and Table II list the quantitative comparison of RRSE and RRAE in face image recovery and F-Measure in foreground detection experiments, respectively. It is easy to see that NIID-MSL obtains the best or the second best performance in most of the cases, which validates the above theoretical analysis experimently.

REFERENCES

[1] D. Meng and F. De La Torre, "Robust matrix factorization with unknown noise," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1337–1344.

[2] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and L. Zhang, "Robust principal component analysis with complex noise," in *International conference on machine learning*, 2014, pp. 55–63.

[3] X. Cao, Y. Chen, Q. Zhao, D. Meng, Y. Wang, D. Wang, and Z. Xu, "Low-rank matrix factorization under general mixture noise distributions," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1493–1501.