"Robust Principal Component Analysis with Complex Noise": Supplementary Material

Qian Zhao† TIMMY.ZHAOQIAN@GMAIL.COM
Deyu Meng† DYMENG@MAIL.XJTU.EDU.CN
Zongben Xu† ZBXU@MAIL.XJTU.EDU.CN
Wangmeng Zuo§ CSWMZUO@GMAIL.COM
Lei Zhang‡ CSLZHANG@COMP.POLYU.EDU.HK
†School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China
§School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
‡Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

In this supplementary material, we give the full hierarchical Bayesian model for MoG-RPCA and present the details of the variational inference process for inferring the posterior of the model.

1. Hierarchical Model for MoG-RPCA

We adopt the RPCA model

$$\mathbf{Y} = \mathbf{L} + \mathbf{E}.$$

Denote by y_{ij} and e_{ij} the elements in the i-th row and j-th column of $\mathbf{Y}$ and $\mathbf{E}$, respectively. We formulate the matrix $\mathbf{L} \in \mathbb{R}^{m\times n}$ with rank $l \leq \min(m,n)$ as the product of two matrices $\mathbf{U} \in \mathbb{R}^{m\times R}$ and $\mathbf{V} \in \mathbb{R}^{n\times R}$ as:

$$\mathbf{L} = \mathbf{U}\mathbf{V}^T = \sum_{r=1}^{R} \mathbf{u}_{\cdot r}\mathbf{v}_{\cdot r}^T,$$

where $R \geq l$, and $\mathbf{u}_{\cdot r}$ and $\mathbf{v}_{\cdot r}$ are the r-th columns of $\mathbf{U}$ and $\mathbf{V}$, respectively. The full hierarchical form of the proposed MoG-RPCA model can then be expressed by:

$$\begin{aligned}
y_{ij} &= \mathbf{u}_{i\cdot}\mathbf{v}_{j\cdot}^T + e_{ij}\\
\mathbf{u}_{\cdot r} &\sim \mathcal{N}(\mathbf{u}_{\cdot r}|\mathbf{0}, \gamma_r^{-1}\mathbf{I}_m)\\
\mathbf{v}_{\cdot r} &\sim \mathcal{N}(\mathbf{v}_{\cdot r}|\mathbf{0}, \gamma_r^{-1}\mathbf{I}_n)\\
\gamma_r &\sim \mathrm{Gam}(\gamma_r|a_0, b_0)\\
e_{ij} &\sim \prod_{k=1}^{K} \mathcal{N}(e_{ij}|\mu_k, \tau_k^{-1})^{z_{ijk}}\\
\mathbf{z}_{ij} &\sim \mathrm{Mutinomial}(\mathbf{z}_{ij}|\boldsymbol{\pi})\\
\boldsymbol{\pi} &\sim \mathrm{Dir}(\boldsymbol{\pi}|\alpha_0)\\
\mu_k, \tau_k &\sim \mathcal{N}(\mu_k|\mu_0, (\beta_0\tau_k)^{-1})\mathrm{Gam}(\tau_k|c_0, d_0).
\end{aligned}$$

The full likelihood of this generative model can be expressed as:

$$\begin{aligned}
&p(\mathbf{U}, \mathbf{V}, \boldsymbol{\mathcal{Z}}, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\pi}, \boldsymbol{\gamma}, \mathbf{Y})\\
&= p(\mathbf{Y}|\mathbf{U}, \mathbf{V}, \boldsymbol{\mathcal{Z}}, \boldsymbol{\mu}, \boldsymbol{\tau})p(\boldsymbol{\mathcal{Z}}|\boldsymbol{\pi})p(\boldsymbol{\mu}|\boldsymbol{\tau})p(\boldsymbol{\tau})p(\mathbf{U}|\boldsymbol{\gamma})p(\mathbf{V}|\boldsymbol{\gamma})p(\boldsymbol{\gamma})\\
&= \prod_{ij}\prod_{k=1}^{K} p(y_{ij}|\mathbf{u}_{i\cdot}, \mathbf{v}_{j\cdot}, \mu_k, \tau_k^{-1})^{z_{ijk}} \prod_{ij} p(\mathbf{z}_{ij}|\boldsymbol{\pi})p(\boldsymbol{\pi})\\
&\quad \prod_{k=1}^{K} p(\mu_k, \tau_k) \prod_{r=1}^{R} \{p(\mathbf{u}_{\cdot r}|\gamma_r)p(\mathbf{v}_{\cdot r}|\gamma_r)p(\gamma_r)\}\\
&= \prod_{ij}\prod_{k=1}^{K} \mathcal{N}(y_{ij}|\mathbf{u}_{i\cdot}\mathbf{v}_{j\cdot}^T + \mu_k, \tau_k^{-1})^{z_{ijk}} \prod_{ij}\prod_{k=1}^{K} \pi_k^{z_{ijk}}\\
&\quad \mathrm{Dir}(\boldsymbol{\pi}|\alpha_0) \prod_{k=1}^{K} \left\{\mathcal{N}(\mu_k|\mu_0, (\beta_0\tau_k)^{-1})\mathrm{Gam}(\tau_k|c_0, d_0)\right\}\\
&\quad \prod_{r=1}^{R} \left\{\mathcal{N}(\mathbf{u}_{\cdot r}|\mathbf{0}, \gamma_r^{-1}\mathbf{I}_m)\mathcal{N}(\mathbf{v}_{\cdot r}|\mathbf{0}, \gamma_r^{-1}\mathbf{I}_n)\mathrm{Gam}(\gamma_r|a_0, b_0)\right\}.
\end{aligned}$$

2. Update Equations

The variational update equations for inferring the posterior of the variables involved in the MoG-RPCA model are given as follows.

Infer $\mathbf{U}$:

$$q(\mathbf{u}_{i\cdot}) = \mathcal{N}(\mathbf{u}_{i\cdot}|\boldsymbol{\mu}_{\mathbf{u}_{i\cdot}}, \boldsymbol{\Sigma}_{\mathbf{u}_{i\cdot}}),$$

Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 2014. JMLR: W&CP volume 32.

Copyright 2014 by the author(s).

where $\langle\cdot\rangle$ denotes the expectation, and

$$\boldsymbol{\Sigma}_{\mathbf{u}_{i\cdot}} = \left(\sum_{k=1}^{K}\langle\tau_k\rangle\sum_{j=1}^{n}\langle z_{ijk}\rangle\langle\mathbf{v}_{j\cdot}^T\mathbf{v}_{j\cdot}\rangle + \boldsymbol{\Gamma}\right)^{-1},$$
$$\boldsymbol{\mu}_{\mathbf{u}_{i\cdot}}^T = \boldsymbol{\Sigma}_{\mathbf{u}_{i\cdot}}\left\{\sum_{k=1}^{K}\langle\tau_k\rangle\sum_{j=1}^{n}\langle z_{ijk}\rangle(y_{ij}-\langle\mu_k\rangle)\langle\mathbf{v}_{j\cdot}\rangle\right\}^T.$$

Infer $\mathbf{V}$:

$$q(\mathbf{v}_{j\cdot}) = \mathcal{N}(\mathbf{v}_{j\cdot}|\boldsymbol{\mu}_{\mathbf{v}_{j\cdot}}, \boldsymbol{\Sigma}_{\mathbf{v}_{j\cdot}}),$$

where

$$\boldsymbol{\Sigma}_{\mathbf{v}_{j\cdot}} = \left(\sum_{k=1}^{K}\langle\tau_k\rangle\sum_{i=1}^{m}\langle z_{ijk}\rangle\langle\mathbf{u}_{i\cdot}^T\mathbf{u}_{i\cdot}\rangle + \boldsymbol{\Gamma}\right)^{-1},$$
$$\boldsymbol{\mu}_{\mathbf{v}_{j\cdot}}^T = \boldsymbol{\Sigma}_{\mathbf{v}_{j\cdot}}\left\{\sum_{k=1}^{K}\langle\tau_k\rangle\sum_{i=1}^{m}\langle z_{ijk}\rangle(y_{ij}-\langle\mu_k\rangle)\langle\mathbf{u}_{i\cdot}\rangle\right\}^T.$$

Infer $\boldsymbol{\gamma}$:

$$q(\gamma_r) = \mathrm{Gam}(\gamma_r|a_r, b_r),$$

where

$$a_r = a_0 + \frac{m+n}{2},$$
$$b_r = b_0 + \frac{1}{2}\left(\langle\mathbf{u}_{\cdot r}^T\mathbf{u}_{\cdot r}\rangle + \langle\mathbf{v}_{\cdot r}^T\mathbf{v}_{\cdot r}\rangle\right).$$

Infer $\boldsymbol{\mathcal{Z}}$:

$$q(\mathbf{z}_{ij}) = \prod_{k=1}^{K} {r_{ijk}}^{z_{ijk}},$$

where

$$r_{ijk} = \frac{\rho_{ijk}}{\sum_k \rho_{ijk}},$$
$$\rho_{ijk} = \frac{1}{2}\langle\ln\tau_k\rangle - \frac{1}{2}\ln 2\pi\langle(y_{ij}-\mathbf{u}_{i\cdot}\mathbf{v}_{j\cdot}^T-\mu_k)^2\rangle - \frac{1}{2}\langle\tau_k\rangle + \langle\ln\pi_k\rangle.$$

Infer $\boldsymbol{\mu}, \boldsymbol{\tau}$:

$$q(\mu_k, \tau_k) = \mathcal{N}(\mu_k|m_k, (\beta_k\tau_k)^{-1})\mathrm{Gam}(\tau_k|c_k, d_k),$$

where

$$\beta_k = \beta_0 + \sum_{ij}\langle z_{ijk}\rangle,$$
$$m_k = \frac{1}{\beta_k}(\beta_0\mu_0 + \sum_{ij}\langle z_{ijk}\rangle(y_{ij} - \langle\mathbf{u}_{i\cdot}\rangle\langle\mathbf{v}_{j\cdot}\rangle^T)),$$
$$c_k = c_0 + \frac{1}{2}\sum_{ij}\langle z_{ijk}\rangle,$$
$$d_k = d_0 + \frac{1}{2}\{\sum_{ij}\langle z_{ijk}\rangle\langle(y_{ij}-\mathbf{u}_{i\cdot}\mathbf{v}_{j\cdot}^T)^2\rangle + \beta_0\mu_0^2 - \frac{1}{\beta_k}(\sum_{ij}\langle z_{ijk}\rangle(y_{ij}-\langle\mathbf{u}_{i\cdot}\rangle\langle\mathbf{v}_{j\cdot}\rangle^T) + \beta_0\mu_0)^2\}.$$

Infer $\boldsymbol{\pi}$:

$$q(\boldsymbol{\pi}) = \mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\alpha}),$$

where

$$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K),$$
$$\alpha_k = \alpha_{0k} + \sum_{ij}\langle z_{ijk}\rangle.$$

3. Calculation of Expectations

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

$$\langle\tau_k\rangle = \frac{c_k}{d_k}$$
$$\langle z_{ijk}\rangle = r_{ijk}$$
$$\langle\ln\tau_k\rangle = \psi(c_k) - \ln d_k$$
$$\langle\ln\pi_k\rangle = \psi(\alpha_k) - \psi(\hat{\alpha}),\ \ \hat{\alpha} = \sum_{k=1}^{K}\alpha_k$$
$$\langle(y_{ij}-\mathbf{u}_{i\cdot}\mathbf{v}_{j\cdot}^T-\mu_k)^2\rangle = \langle(y_{ij}-\mathbf{u}_{i\cdot}\mathbf{v}_{j\cdot}^T)^2\rangle - 2\langle\mu_k\rangle(y_{ij}-\langle\mathbf{u}_{i\cdot}\rangle\langle\mathbf{v}_{j\cdot}\rangle^T) + \langle\mu_k^2\rangle$$
$$\langle\mu_k\rangle = m_k$$
$$\langle\mu_k^2\rangle = (\beta_k\tau_k)^{-1} + m_k^2$$
$$\langle(y_{ij}-\mathbf{u}_{i\cdot}\mathbf{v}_{j\cdot}^T)^2\rangle = y_{ij}^2 + \mathrm{tr}\left(\langle\mathbf{u}_{i\cdot}^T\mathbf{u}_{i\cdot}\rangle\langle\mathbf{v}_{j\cdot}^T\mathbf{v}_{j\cdot}\rangle\right) - 2y_{ij}\langle\mathbf{u}_{i\cdot}\rangle\langle\mathbf{v}_{j\cdot}\rangle^T$$
$$\langle\mathbf{u}_{i\cdot}^T\mathbf{u}_{i\cdot}\rangle = \boldsymbol{\Sigma}_{\mathbf{u}_{i\cdot}} + \langle\mathbf{u}_{i\cdot}\rangle\langle\mathbf{u}_{i\cdot}\rangle^T$$
$$\langle\mathbf{v}_{j\cdot}^T\mathbf{v}_{j\cdot}\rangle = \boldsymbol{\Sigma}_{\mathbf{v}_{j\cdot}} + \langle\mathbf{v}_{j\cdot}\rangle\langle\mathbf{v}_{j\cdot}\rangle^T$$
$$\boldsymbol{\Gamma} = \mathrm{diag}\left(\langle\boldsymbol{\gamma}\rangle\right),\ \ \langle\gamma_r\rangle = \frac{a_r}{b_r}$$
$$\langle\mathbf{u}_{\cdot r}^T\mathbf{u}_{\cdot r}\rangle = \langle\mathbf{u}_{\cdot r}\rangle^T\langle\mathbf{u}_{\cdot r}\rangle + \sum_{i=1}^{m}\left(\boldsymbol{\Sigma}_{\mathbf{u}_{i\cdot}}\right)_{rr}$$
$$\langle\mathbf{v}_{\cdot r}^T\mathbf{v}_{\cdot r}\rangle = \langle\mathbf{v}_{\cdot r}\rangle^T\langle\mathbf{v}_{\cdot r}\rangle + \sum_{j=1}^{n}\left(\boldsymbol{\Sigma}_{\mathbf{v}_{j\cdot}}\right)_{rr},$$

where $\psi(\cdot)$ is the digamma function defined by $\psi(x) = \frac{d}{dx}\ln\Gamma(x)$.