Supplementary Material of
Low-rank Matrix Factorization under General Mixture Noise Distributions

Paper 1291

1. Proof of Theorem 1

Theorem 1. *Let $l_P^G(\mathbf{\Theta}) = l(\mathbf{\Theta}) - P(\boldsymbol{\pi}; \lambda)$, where $l(\Theta)$ is defined in (8). If we assume that $\{\mathbf{\Theta}^{(t)}\}$ is the sequence generated by Algorithm 1 and the sequence of likelihood values $\{l_P^G(\mathbf{\Theta}^{(t)})\}$ is bounded above, then there exits a constant $l^\star$ such that*

$$\lim_{t\to\infty} l_P^G(\mathbf{\Theta}^{(t)}) = l^\star, \tag{1}$$

where

$$\mathbf{\Theta}^{(t)} = \arg\max_{\mathbf{\Theta}} \left\{ \Omega(\mathbf{\Theta}|\mathbf{\Theta}^{(t-1)}) + P(\boldsymbol{\pi}^{(t-1)}; \lambda) - P(\boldsymbol{\pi}; \lambda) \right\}, \tag{2}$$

and

$$\Omega(\mathbf{\Theta}|\mathbf{\Theta}^{(t-1)}) = \sum_{\mathbf{Z}} \mathbb{P}(\mathbf{Z}|\mathbf{E}; \mathbf{\Theta}^{(t-1)}) \log \frac{\mathbb{P}(\mathbf{E}, \mathbf{Z}; \mathbf{\Theta})}{\mathbb{P}(\mathbf{E}, \mathbf{Z}; \mathbf{\Theta}^{(t-1)})}. \tag{3}$$

Proof. (i) First, we calculate that

$$\begin{aligned} l_P^G(\mathbf{\Theta}) - l_P^G(\mathbf{\Theta}^{(t)}) &= l(\mathbf{\Theta}) - l(\mathbf{\Theta}^{(t)}) + P(\pi^{(t)}; \lambda) - P(\pi; \lambda) \\ &= \log \sum_{\mathbf{Z}} \mathbb{P}(\mathbf{Z}|\mathbf{E}; \mathbf{\Theta}^{(t)}) \frac{\mathbb{P}(\mathbf{E}|\mathbf{Z}; \mathbf{\Theta})\mathbb{P}(\mathbf{Z}; \mathbf{\Theta})}{\mathbb{P}(\mathbf{Z}|\mathbf{E}; \mathbf{\Theta}^{(t)})} \\ &\quad - \log \mathbb{P}(\mathbf{E}; \mathbf{\Theta}^{(t)}) + P(\pi^{(t)}; \lambda) - P(\pi; \lambda) \\ &\geq \sum_{\mathbf{Z}} \mathbb{P}(\mathbf{Z}|\mathbf{E}; \mathbf{\Theta}^{(t)}) \log \frac{\mathbb{P}(\mathbf{E}|\mathbf{Z}; \mathbf{\Theta})\mathbb{P}(\mathbf{Z}; \mathbf{\Theta})}{\mathbb{P}(\mathbf{Z}|\mathbf{E}; \mathbf{\Theta}^{(t)})} \\ &\quad - \log \mathbb{P}(\mathbf{E}; \mathbf{\Theta}^{(t)}) + P(\pi^{(t)}; \lambda) - P(\pi; \lambda) \\ &= \sum_{\mathbf{Z}} \mathbb{P}(\mathbf{Z}|\mathbf{E}; \mathbf{\Theta}^{(t)}) \log \frac{\mathbb{P}(\mathbf{E}|\mathbf{Z}; \mathbf{\Theta})\mathbb{P}(\mathbf{Z}; \mathbf{\Theta})}{\mathbb{P}(\mathbf{Z}|\mathbf{E}; \mathbf{\Theta}^{(t)})\mathbb{P}(\mathbf{E}; \mathbf{\Theta}^{(t)})} \\ &\quad + P(\pi^{(t)}; \lambda) - P(\pi; \lambda). \end{aligned}$$

Let $\Omega(\mathbf{\Theta}|\mathbf{\Theta}^{(t)}) = \sum_{\mathbf{Z}} \mathbb{P}(\mathbf{Z}|\mathbf{E}; \mathbf{\Theta}^{(t)}) \log \frac{\mathbb{P}(\mathbf{E}|\mathbf{Z};\mathbf{\Theta})\mathbb{P}(\mathbf{Z};\mathbf{\Theta})}{\mathbb{P}(\mathbf{Z}|\mathbf{E};\mathbf{\Theta}^{(t)})\mathbb{P}(\mathbf{E};\mathbf{\Theta}^{(t)})}$, then

$$l_P^G(\mathbf{\Theta}) \geq l_P^G(\mathbf{\Theta}^{(t)}) + \Omega(\mathbf{\Theta}|\mathbf{\Theta}^{(t)}) + P(\pi^{(t)}; \lambda) - P(\pi; \lambda). \tag{4}$$

(ii) In the M step of Algorithm 1, it is obvious that

$$\begin{aligned} \mathbf{\Theta}^{(t+1)} &= \arg\max_{\mathbf{\Theta}} \left\{ \sum_{\mathbf{Z}} \mathbb{P}(\mathbf{Z}|\mathbf{E}; \mathbf{\Theta}^{(t)}) \log \mathbb{P}(\mathbf{E}, \mathbf{Z}; \mathbf{\Theta}) - P(\pi; \lambda) \right\} \\ &= \arg\max_{\mathbf{\Theta}} \left\{ \sum_{\mathbf{Z}} \mathbb{P}(\mathbf{Z}|\mathbf{E}; \mathbf{\Theta}^{(t)}) \frac{\log \mathbb{P}(\mathbf{E}, \mathbf{Z}; \mathbf{\Theta})}{\log \mathbb{P}(\mathbf{E}, \mathbf{Z}; \mathbf{\Theta}^{(t)})} - P(\pi; \lambda) \right\} \\ &= \arg\max_{\mathbf{\Theta}} \left\{ \Omega(\mathbf{\Theta}|\mathbf{\Theta}^{(t)}) + P(\pi^{(t)}; \lambda) - P(\pi; \lambda) \right\}. \end{aligned}$$

Thus, we have

$$\begin{aligned} &\Omega(\mathbf{\Theta}^{(t+1)}|\mathbf{\Theta}^{(t)}) + P(\pi^{(t)}; \lambda) - P(\pi^{(t+1)}; \lambda) \\ &\geq \Omega(\mathbf{\Theta}^{(t)}|\mathbf{\Theta}^{(t)}) + P(\pi^{(t)}; \lambda) - P(\pi^{(t)}; \lambda) \\ &= 0 \end{aligned}$$

Then, we can easily derive that

$$l_P^G(\mathbf{\Theta}^{(t+1)}) \geq l_P^G(\mathbf{\Theta}^{(t)}). \tag{5}$$

Based on (5), the sequence $\{l_P^G(\mathbf{\Theta}^{(t)})\}_{t=1}^\infty$ is nondecreasing and bounded above. Therefore, there exits a constant $l^\star$ such that

$$\lim_{t\to\infty} l_P^G(\mathbf{\Theta}^{(t)}) = l^\star. \tag{6}$$

□

2. Exponential Power Distribution

2.1. Three different forms of Exponential Power Distribution

The Exponential Power Distribution ($\mu = 0$) has the following three equvalent forms:

$$f_p(x; 0, \sigma) = \frac{1}{2\sigma p^{\frac{1}{p}}\Gamma(1+\frac{1}{p})} \exp\left\{ -\frac{|x|^p}{p\sigma^p} \right\}. \tag{7}$$

Let $\tau = (p\sigma^p)^{\frac{1}{p}}$, then

$$f_p(x; 0, \tau) = \frac{1}{2\tau\Gamma(1+\frac{1}{p})} \exp\left\{ -|\frac{x}{\tau}|^p \right\}. \tag{8}$$

Let $\eta = \frac{1}{\tau^p}$, then

$$f_p(x; 0, \eta) = \frac{\eta^{\frac{1}{p}}}{2\Gamma(1+\frac{1}{p})} \exp\left\{ -\eta|x|^p \right\}. \tag{9}$$

By making use of the property of Γ function $\Gamma(1+x) = x\Gamma(x)$, we get

$$\Gamma(1+\frac{1}{p}) = \frac{1}{p}\Gamma(\frac{1}{p}). \tag{10}$$

2.2. Draw Samples from Exponential Power Distribution

The second form of exponential power distribution is

$$f_p(x; 0, \tau) = \frac{1}{2\tau\Gamma(1+\frac{1}{p})} \exp\left\{ -|\frac{x}{\tau}|^p \right\}. \tag{11}$$

Sampling from the exponential power distribution contains two cases: $p \geq 1$ and $0 < p < 1$.

2.2.1 case 1: $p \geq 1$

We adopt the method proposed in [2, 3, 4].

2.2.2 case 2: $0 < p < 1$

When $0 < p < 1$, the method proposed in [5] is used. We sample the distribution in two steps:

$$(w|p) \sim \frac{1+p}{2} Ga(2+\frac{1}{p}, 1) + \frac{1-p}{2} Ga(1+\frac{1}{p}, 1), \quad (12)$$

$$(\beta|\tau, w, p) \sim \frac{1}{\tau w^{\frac{1}{p}}} \left\{ 1 - |\frac{\beta}{\tau w^{\frac{1}{p}}}| \right\}_+, \quad (13)$$

where w is a intermediate variable. (12) can be sampled directly but (13) is difficult. Therefore, we adopt the slice sampling strategy in [1].

References

[1] C. M. Bishop et al. *Pattern recognition and machine learning*, volume 4. springer New York, 2006. 2

[2] M. Chiodi. Generation of pseudo random variates from a normal distribution of order p. *Statistica Applicata (Italian Journal of Applied Statistics)*, 7(4):401–416, 1995. 2

[3] G. Marsaglia and T. A. Bray. A convenient method for generating normal variables. *Siam Review*, 6(3):260–264, 1964. 2

[4] A. M. Mineo, M. Ruggieri, et al. A software tool for the exponential power distribution: The normalp package. *Journal of Statistical Software*, 12(4):1–24, 2005. 2

[5] N. G. Polson, J. G. Scott, and J. Windle. The bayesian bridge. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(4):713–733, 2014. 2