Low-rank Matrix Factorization under General Mixture Noise Distributions

作者Xiangyong Cao1 Yang Chen1 Qian Zhao1 Deyu Meng1;2; Yao Wang1 Dong Wang1 Zongben Xu1;2

一.工作动机

low rank matrix factorization (LRMF)是一个常用的子空间学习策略，当前的LRMF技术主要集中于L1范式和L2范式，前者对应拉普拉斯噪音，后者对应高斯噪音。实际应用里，L2范式高斯噪音会对超常值和非高斯噪音很敏感，L1范式鲁棒性好，但是仅限于拉普拉斯噪音，没法处理其他类型的噪音。最新进展，假设噪音服从mixture of Gaussians（MoG），其中有几个缺点：缺少好的方式选择混合高斯的组分的数量；有限个混合组分难以模拟拉普拉斯分布；真实的噪音有些是sub-and super-Gaussians，难以用混合高斯表示。本文的目的在于使LRMF能适应更复杂的噪音。

二.模型

LRMF问题数学形式，Y∈Rm×n with entries yijs

$$\min_{\mathbf{U},\mathbf{V}} ||\mathbf{W} \odot (\mathbf{Y} - \mathbf{U}\mathbf{V}^T)||, \tag{1}$$

where $\mathbf{W}$ is the indicator matrix with $w_{ij} = 0$ if y_{ij} is missing and 1 otherwise, and $\mathbf{U} \in \mathcal{R}^{m\times r}$ and $\mathbf{V} \in \mathcal{R}^{n\times r}$ are low-rank matrices ($r < min(m,n)$). The operator $\odot$ denotes the Hadamard product (the component-wise multiplication) and $||\cdot||$ corresponds to a certain noise measure.

1.PMoEP 模型

假设噪音服从MoEP分布

$$\mathbb{P}(e_{ij}) = \sum_{k=1}^{K} \pi_k f_{p_k}(e_{ij}; 0, \eta_k),$$

其中πk为混合比例，η和p为混合组分的参数，k代表第k个混合组分，总数为K，f是混合组分的函数

$$f_p(e; 0, \eta) = \frac{p\eta^{\frac{1}{p}}}{2\Gamma(\frac{1}{p})} \exp\{-\eta|e|^p\},$$

p为形状参数

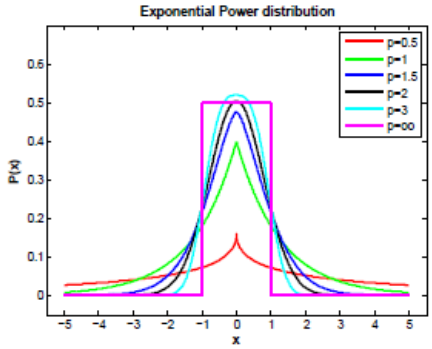


Figure 2. The probability density function of EP distributions.

设定一个指示变量zij = [zij1; zij2;...; zijK]T，当eij来自第k个分布时，相应的zijk=1，

$$\mathbb{P}(e_{ij}|\mathbf{z}_{ij}) = \prod_{k=1}^{K} f_{p_k}(e_{ij};0,\eta_k)^{z_{ijk}},$$

$$\mathbb{P}(\mathbf{z}_{ij};\boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{z_{ijk}}.$$

做出complete likelihood function, Ω是没有缺失的yijs的下标集

Denoting $\mathbf{E} = (e_{ij})_{m\times n}$, $\mathbf{Z} = (\mathbf{z}_{ij})_{m\times n}$ and $\boldsymbol{\Theta} = \{\boldsymbol{\pi}, \boldsymbol{\eta}, \mathbf{U}, \mathbf{V}\}$ with $\boldsymbol{\eta} = [\eta_1, \eta_2, \ldots, \eta_K]^T$, the *complete likelihood function* can be written as

$$\mathbb{P}(\mathbf{E},\mathbf{Z};\boldsymbol{\Theta}) = \prod_{i,j\in\Omega}\prod_{k=1}^{K}[\pi_k f_{p_k}(e_{ij};0,\eta_k)]^{z_{ijk}}, \quad (7)$$

取对数，加入惩罚项。ε是一个非常小正数，**λ是调整参数**，Dk是第k个组分的自由参数，下式为惩罚项

$$P(\boldsymbol{\pi};\lambda) = n\lambda\sum_{k=1}^{K} D_k \log\frac{\epsilon+\pi_k}{\epsilon},$$

得到PMoEP模型

$$\max_{\boldsymbol{\Theta}}\left\{l_P^C(\boldsymbol{\Theta}) = l^C(\boldsymbol{\Theta}) - P(\boldsymbol{\pi};\lambda)\right\},$$

2. EM算法

Algorithm 1 EM Algorithm for PMoEP LRMF

Input:

Data $\mathbf{Y} = \{y_{ij}\}_{m\times n}$. Initialize $\mathbf{\Theta}^{(t)} = \{\boldsymbol{\pi}^{(t)}, \boldsymbol{\eta}^{(t)}, \mathbf{U}^{(t)}, \mathbf{V}^{(t)}\}$, the number of components K_{start}, preset candidates $\mathbf{p} = [p_1, \ldots, p_{K_{start}}]$, tolerance ϵ and $t = 0$.

Output:

Parameter $\hat{\mathbf{\Theta}}$, the number of final components $\hat{K}$ and posterior probability $\boldsymbol{\gamma} = \{\gamma_{ijk}\}_{m\times n\times \hat{K}}$.

1: **repeat**
2: (*E step*): Update the posterior probability $\boldsymbol{\gamma}^{(t)}$ by (12).
3: (*M step for* $\boldsymbol{\pi}$): Update parameter $\boldsymbol{\pi}^{(t)}$ by (15), and remove the component with $\pi_k^{(t)} = 0$.
4: (*M step for* $\boldsymbol{\eta}$): Update parameters $\boldsymbol{\eta}^{(t)}$ by (16).
5: (*M step for* $\mathbf{U}, \mathbf{V}$): Update $\mathbf{U}^{(t)}, \mathbf{V}^{(t)}$ by maximizing (18) using ALM strategy (19)-(24).
6: **until** *converge*;

$$y_{ij} = \mathbf{u}_i \mathbf{v}_j^T + e_{ij},$$

Expectation：计算zijk在eij下的条件期望

$$\gamma_{ijk}^{(t+1)} = \frac{\pi_k^{(t)} f_{p_k}(y_{ij} - \mathbf{u}_i^{(t)}(\mathbf{v}_j^{(t)})^T)|0, \eta_k^{(t)})}{\sum_{l=1}^{K} \pi_l^{(t)} f_{p_l}(y_{ij} - \mathbf{u}_i^{(t)}(\mathbf{v}_j^{(t)})^T)|0, \eta_l^{(t)}))}.$$

代入惩罚似然的式子，得到Q函数

$$Q(\mathbf{\Theta}, \mathbf{\Theta}^{(t)}) = \sum_{i,j\in\Omega} \sum_{k=1}^{K} \gamma_{ijk}^{(t+1)} [\log f_{p_k}(e_{ij}; 0, \eta_k) + \log \pi_k] - n\lambda \sum_{k=1}^{K} D_k \log \frac{\epsilon + \pi_k}{\epsilon}.$$

Maximation：

update for π：Q函数加入拉格朗日乘数τ实施限制条件，对其求一阶导数 $\sum_{k=1}^{K} \pi_k = 1$

$$\pi_k^{(t+1)} = \max\left\{0, \frac{1}{1-\lambda\hat{D}}\left[\frac{\sum_{i,j\in\Omega} \gamma_{ijk}^{(t+1)}}{|\Omega|} - \lambda D_k\right]\right\},$$

update for η：对Q函数求一阶导数

$$\eta_k^{(t+1)} = \frac{N_k}{p_k \sum_{i,j\in\Omega} \gamma_{ijk}^{(t+1)} |y_{ij} - \mathbf{u}_i^{(t)}(\mathbf{v}_j^{(t)})^T|^{p_k}}, \qquad N_k = \sum_{i,j\in\Omega} \gamma_{ijk}^{(t+1)}.$$

update for U、V：最大化如下函数

$$\sum_{i,j\in\Omega} \sum_{k=1}^{K} \gamma_{ijk}^{(t+1)} \log f_k(y_{ij} - \mathbf{u}_i^{(t)}(\mathbf{v}_j^{(t)})^T; \eta_k^{(t+1)}).$$

相当于最小化

$$\min_{\mathbf{U},\mathbf{V}} \sum_{k=1}^{K} ||\mathbf{W}_{(k)} \odot (\mathbf{Y} - \mathbf{U}\mathbf{V}^T)||_{p_k}^{p_k}, \tag{18}$$

where the element $w_{(k)ij}$ of $\mathbf{W}_{(k)} \in \mathcal{R}^{m\times n}$ is

$$w_{(k)ij} = \begin{cases} (\eta_k^{(t+1)} \gamma_{ijk}^{(t+1)})^{\frac{1}{p_k}}, & i,j \in \Omega \\ 0, & i,j \notin \Omega \end{cases}.$$

使用augmented Lagrange multipliers(ALM)得到结果

3.调整参数λ选择

λ在1.中是一个固定的数字，实际应用需要选择。使用Bayesian information criterion(BIC)

$$\mathrm{BIC}(\lambda) = \sum_{i,j\in\Omega} \log\{\sum_{k=1}^{\hat{K}} \hat{\pi}_k f_k(e_{ij};\hat{\eta}_k)\} - \frac{1}{2}(\sum_{k=1}^{\hat{K}} D_k)\log|\Omega|. \tag{29}$$

and then select $\hat{\lambda}$ by

$$\hat{\lambda} = \arg\max_{\lambda} \mathrm{BIC}(\lambda), \tag{30}$$

where $|\Omega|$ is the number of the observations, $\hat{K}$ is the estimate of the number of components, $\hat{\pi}_k$ is the estimate of parameter π_k and $\hat{\eta}_k$ is the estimate of parameter η_k for maximizing (10) with a given λ.

三、实验

1.生成数据实验

$\mathbf{U}_{gt} \in \mathcal{R}^{40\times 4}$ and $\mathbf{V}_{gt} \in \mathcal{R}^{20\times 4}$, i.e., $\mathbf{Y}_{gt} = \mathbf{U}_{gt}\mathbf{V}_{gt}^T$.

两个低秩矩阵相乘，随机生成三十个秩为4的40×20矩阵，随机去掉20%，添加4种噪音在80%上

non-missing entries as follows: (1) *Gaussian noise*: 80% of the entries were corrupted with $\mathcal{N}(0, 0.04)$. (2) *Sparse noise*: 10% of the entries were corrupted with uniformly distributed noise between [-20,20]. (3) *EP noise*: 80% of the entries were corrupted with $EP(0, 0.2^p p), p = 0.2$. (4) *Mixture noise*: 20% of the entries were corrupted with uniformly distributed noise between [-5,5], 20% are contaminated with Gaussian noise $\mathcal{N}(0, 0.04)$ and the remaining 40% are corrupted with Gaussian noise $\mathcal{N}(0, 0.01)$. We de-

6种衡量方式

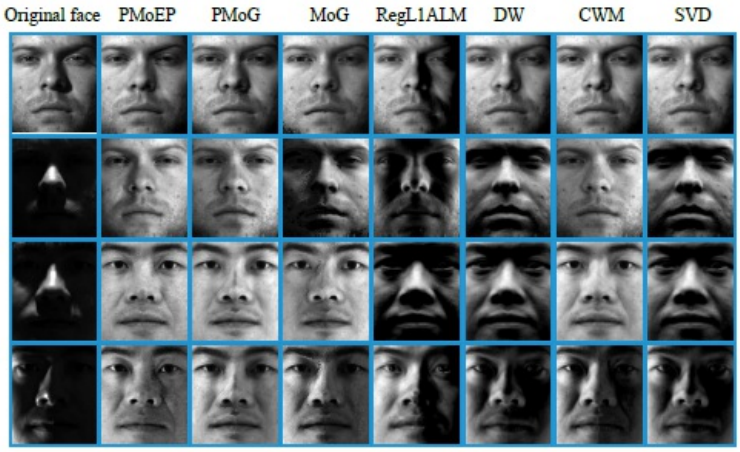
$$C1 = ||\mathbf{W} \odot (\mathbf{Y}_{no} - \tilde{\mathbf{U}}\tilde{\mathbf{V}}^T)||_1, \quad C2 = ||\mathbf{W} \odot (\mathbf{Y}_{no} - \tilde{\mathbf{U}}\tilde{\mathbf{V}}^T)||_2,$$
$$C3 = ||\mathbf{Y}_{gt} - \tilde{\mathbf{U}}\tilde{\mathbf{V}}^T||_1, \quad C4 = ||\mathbf{Y}_{gt} - \tilde{\mathbf{U}}\tilde{\mathbf{V}}^T||_2,$$
$$C5 = subspace(\mathbf{U}_{gt}, \tilde{\mathbf{U}}), \quad C6 = subspace(\mathbf{V}_{gt}, \tilde{\mathbf{V}}),$$

结果，~U; ~V是输出

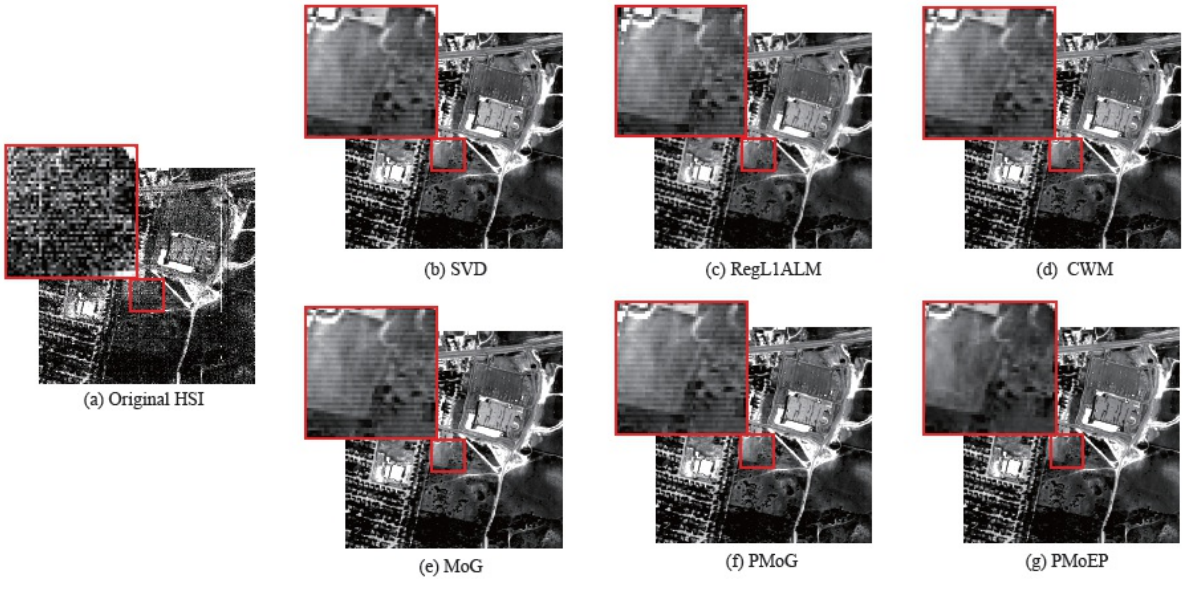
	PMoEP	PMoG	MoG [19]	CWM [20]	DW [24]	RegL1ALM [37]
			Sparse Noise			
C1	8.32e+2	7.98e+2	8.03e+2	**7.97e+2**	1.05e+3	8.63e+2
C2	1.07e+4	1.06e+4	1.08e+4	9.97e+3	**4.79e+3**	5.96e+3
C3	3.23e+1	**2.13e-12**	1.57e+1	4.10e+2	5.62e+14	1.19e+6
C4	6.73e+1	**2.42e-5**	3.45e+1	9.69e+1	1.30e+7	4.26e+3
C5	1.40e-1	**3.85e-8**	3.30e-2	3.93e-1	1.47e+1	1.49e+1
C6	6.16e-2	**2.33e-8**	1.96e-6	1.13e-1	1.44e+1	1.45e+1
			Gaussian Noise			
C1	8.08e+1	7.83e+1	8.20e+1	7.71e+1	8.16e+1	**7.35e+1**
C2	**1.65e+1**	2.41e+1	1.67e+1	2.17e+1	1.67e+1	2.14e+1
C3	**1.31e+1**	2.48e+1	1.33e+1	2.24e+1	1.32e+1	2.03e+1
C4	7.89e+1	1.06e+2	7.90e+1	9.95e+1	**7.86e+1**	9.74e+1
C5	**8.54e-2**	1.17e-1	9.24e-2	1.22e-1	8.67e-2	1.10e-1
C6	5.82e-2	8.43e-2	6.30e-2	9.96e-2	**5.77e-2**	7.06e-2
			EP Noise			
C1	3.64e+2	3.19e+2	**3.18e+2**	3.30e+2	4.33e+2	3.53e+2
C2	1.29e+3	1.25e+3	1.02e+3	1.35e+3	**6.49e+2**	8.75e+2
C3	**1.58e+2**	3.28e+3	6.98e+4	1.99e+2	1.04e+7	6.66e+4
C4	**2.40e+2**	2.65e+2	3.52e+2	2.47e+2	2.57e+3	8.22e+2
C5	**3.02e-1**	4.21e-1	4.19e-1	3.62e-1	1.08e+1	1.14e+1
C6	**2.02e-1**	2.87e-1	3.28e-1	2.53e-1	9.05e-1	1.01e+1
			Mixture Noise			
C1	4.40e+2	4.38e+2	4.27e+2	4.31e+2	5.18e+2	**4.26e+2**
C2	1.27e+3	1.33e+3	1.28e+3	1.10e+3	**8.26e+2**	1.12e+3
C3	**1.42e+2**	1.90e+2	1.88e+2	3.77e+2	2.16e+8	1.45e+4
C4	**1.68e+2**	1.91e+2	1.84e+2	3.15e+2	7.08e+3	5.08e+2
C5	**3.47e-1**	3.90e-1	3.93e-1	5.41e-1	8.44e-1	7.96e-1
C6	1.79e-1	1.78e-1	**1.62e-1**	3.91e-1	7.46e-1	6.17e-1
Time(s)	2.34	0.28	0.46	**0.062**	0.352	0.11

2.真实数据实验

人脸识别：



高光谱图像重构：



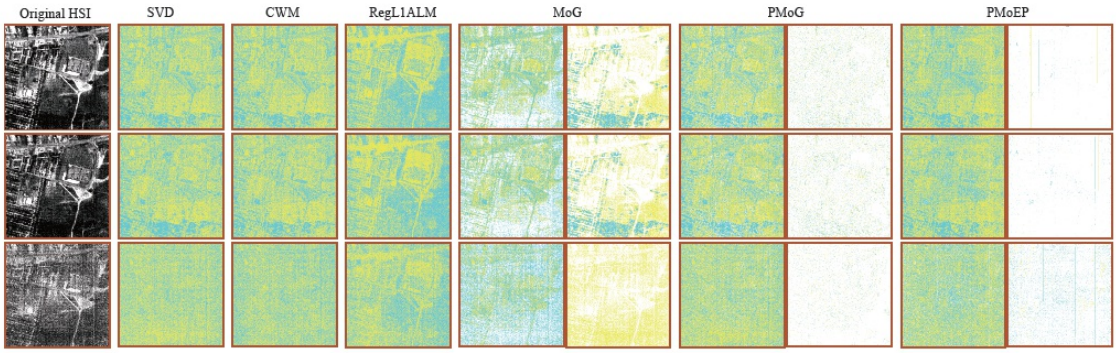


Figure 8. From left to right: original bands, and extracted noise by SVD, CWM, RegL1ALM, MoG, PMoG and PMoEP. The noises with positive and negative values are depicted in yellow and blue, respectively. This figure should be viewed in color and the details are better seen by zooming on a computer screen.

四、总结

假设噪音服从Mixture of Exponential Power (MoEP) distributions，通过惩罚似然方法和MoEP分布，构建penalized MoEP （PMoEP）model，设计一个expectation maximization (EM)算法估计参数，求解PMoEP。

优点：拟合指数族噪音，从数据中自动学习组分的数量K，更好从有干扰数据获得真实低秩矩阵。