方法都是基于 L2 范数的，L2 范数与 F 范数一样对异常值敏感(Qifa Ke, Kanade, 2005)。

FPCA 是一种具有统计基础的矩阵分解方法，相较于 2DPCA，FPCA 有一个因子协方差模型，该模型假定数据服从矩阵正态分布，是一种的概率方法(Dryden et al., 2009)。然而，正是因为正态假定使得 FPCA 对非正态特征（如偏态和重尾）缺乏稳定性和稳健性(Lange et al., 1989; Lin et al., 2015)。

RFPCA 是一种新的 FPCA 的稳健扩展方法，从多元 t 分布能够有效处理异常值从而得到稳健估计的特性出发(Lange et al., 1989)，考虑在 FPCA 的基础上，将矩阵正态分布替换为矩阵变元 t 分布(Gupta et al., 2013)，有效提高了 FPCA 的稳健性。但是，当 RFPCA 应用于人脸识别中时，局限于使用人脸的灰度图像，图像的颜色信息完全消失(El-Melegy, Kamal, 2017)。若为了保留图像的颜色信息，将三维数据转换为二维数据，RFPCA 又面临着“维数灾难”问题。

实际上，在判别分析中，小样本、高维度的数据情形有着相似困境，不但计算耗时，而且计算中的估计偏差会导致判别结果有误，基于此，Friedman (1989)提出一种群协方差矩阵正则化方法(Mkhadri et al., 1997)，能够有效提高分类精度，并且能够实现快速计算。该方法通过替换协方差矩阵的极大似然估计得以实现，替换后协方差矩阵的表达式中含有正则化参数，的值是通过最小化交叉验证错分类风险确定的。当使用 Expectation Maximization (EM) 算法(Dempster et al., 1977)进行极大似然估计 (Maximum Likelihood Estimation, ML Estimation) 时，存在矩阵奇异或退化的问题，Fraley 和 Raftery (2007)提议用基于 EM 算法的最大后验估计 (Maximum A Posteriori Estimation, MAP Estimation) 替换 ML 估计，其中涉及一个分散但恰当的共轭先验，是一种基于贝叶斯模型的正则化方法，Mkhadri et al. (1997)将其称为正则化预测判别分析。

很自然地，有学者考虑将正则化方法应用到 PCA 上来，Raiko et al. (2007)尝试在目标函数上添加惩罚项，得到一种正则化 PCA (Regularized PCA)。Zou et al. (2006)提出了一种新的稀疏主成分分析法 (SPCA)，利用 lasso(弹性网) 来生成带有稀疏载荷的修正主成分。通过基于 l_0 及 l_1 惩罚的正则化方法，添加正则化稀疏惩罚项，Sun et al. (2013)得到一种稳健的 2DPCA (robust 2DPCA)，这种稳健 2DPCA 能够在保留空间信息的同时，对异常值不敏感。除此之外，通过一种结构化稀疏正则化方法，即在二维图像中加入结构化稀疏先验，也能达到前者的

效果(Sun et al., 2015)。但是，这些都是在非概率方法中考察正则化，在进行人脸图像识别时，非概率方法无法有效应对不确定性来源，如光照和姿态变化(Zhao et al., 2003)。

由Dempster et al. (1977)提出的 EM 算法是一种通过迭代方式计算参数极大似然估计的算法，可应用于有数据缺失的情形。该算法能够确保似然函数在每一步迭代时都处于上升趋势，且收敛到局部极大值。两种 EM 型算法 Expectation Conditional Maximization (ECM) 算法(Meng, Rubin, 1993) 和 ECM of Either (ECME) 算法(Liu, Rubin, 1994) 是 EM 算法的拓展，相比于 EM 算法，这两种算法的收敛速度更快。

第三节　　研究内容及创新之处

为了解决 RFPCA 在高维矩阵型数据上进行参数估计时面临的矩阵奇异或退化问题，本文提出了正则化稳健主成分分析方法 (Regularized Robust PCA, R-RPCA)。与 RFPCA 一样，R-RPCA 使用矩阵变元 t 分布Gupta et al. (2013)，且具有稳健性。本文的创新点包括：

(1) 基于矩阵变元 t 分布，提出了能够解决 RFPCA 在高维矩阵型数据上所面临的问题的方法 R-RPCA；

(2) R-RPCA 利用最大后验估计 (MAP Estimation) 来进行参数估计，有效解决了 RFPCA 利用极大似然估计 (ML Estimation) 来进行参数估计时参数在迭代过程中发生矩阵奇异或退化的问题；

(3) 使用 k 折交叉验证选取最优的正则化参数 λ，以此训练出最优模型，再通过 EM 型算法，得到 R-RPCA 的 MAP；

(4) 与基于矩阵正态分布的正则化主成分分析 (Regularized PCA, R-PCA) 相比较，R-RPCA 是稳健的；

(5) 矩阵变元 t 分布中的协方差矩阵分为可分协方差矩阵和不可分协方差矩阵。实验结果表明，在进行最大后验估计时，针对可分协方差矩阵，引入一个正则化参数 λ 表现更好；针对不可分协方差矩阵，引入两个正则化参数 λ_1 和 λ_2 表现更优。

第四节　　余下章节结构安排

本文余下章节安排如下：

第二章：相关理论与方法。本章简要回顾了文中涉及到的 EM 型算法、矩阵正态分布及其基础上的 FPCA、矩阵变元 t 分布及其基础上的 RFPCA。

第三章：正则化稳健主成分分析。本章引出了正则化稳健主成分分析 (R-RPCA) 方法并对其做了介绍，另外，在引入两种先验信息的情况下分别给出了 R-RPCA 第一种正则化 (MAP1) 和第二种正则化 (MAP2) 的最大后验估计算法。

第四章：实验。本章进行了一系列在模拟数据上的模拟实验以及 R-RPCA 的 MAP1 和 MAP2 在真实数据上的应用。模拟实验包括 ECM 和 ECME 算法的收敛性、两种算法对初始化的敏感性、参数估计的准确性、模型稳健性、异常值检测。真实实验包括人脸识别和多元时间序列的分类。

第五章：总结与展望。本章对本文的主要内容进行了简要的总结，重申了创新之处，并概括了实验结论，同时阐述了后续可能的研究方向。

第二章　相关理论与方法

第一节　　EM 及其拓展算法

一、EM 算法

EM 算法(Dempster et al., 1977) 是一种通过迭代方式计算参数极大似然估计的算法，常用于含有潜变量的概率模型中，可应用于有数据缺失以及存在潜在(即无法观测) 变量的情形。

记完全数据 $\mathbf{X}=\{\mathbf{X}_o;\mathbf{Z}\}$，其中 $\mathbf{X}_o$ 为观测数据，$\mathbf{Z}$ 为缺失数据，则 EM 算法主要分为两步：

E-step : 计算完全数据对数似然函数的条件期望

$$Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)})=\mathbb{E}[\mathcal{L}\,(\boldsymbol{\theta} \mid \mathbf{X}) \mid \mathbf{X}_o,\boldsymbol{\theta}^{(t)}] \tag{2.1}$$

M-step : 关于 $\boldsymbol{\theta}$ 极大化 **E-step** 的条件期望，得到 $\boldsymbol{\theta}$ 的估计值

$$\boldsymbol{\theta}^{(t+1)}=\underset{\boldsymbol{\theta}}{\mathrm{Argmax}}Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}) \tag{2.2}$$

循环 **E-step** 和 **M-step** 两个步骤，当对数似然 $\mathcal{L}$ 的值小于某个给定的取值时，退出迭代过程。

二、EM 型算法

本节介绍两种 EM 型算法，即 ECM 算法和 ECME 算法。

• ECM 算法由Meng 和 Rubin (1993)提出，是 EM 算法的拓展。该算法将 EM 算法的最大化步骤 (**M-step**) 替换成条件最大化步骤 (**CM-step**)。

将参数 $\boldsymbol{\theta}$ 划分成 $\boldsymbol{\theta}_1$ 和 $\boldsymbol{\theta}_2$，$Q(\boldsymbol{\theta})$ 为完全对数似然函数的期望，则 **CM-step** 分为两小步：

CM-step1 : 给定 $\boldsymbol{\theta}_2$ 的初始值，关于 $\boldsymbol{\theta}_1$ 极大化 $Q(\boldsymbol{\theta})$，得到 $\boldsymbol{\theta}_1$ 的估计值

$$\boldsymbol{\theta}_1^{(t+1)}=\underset{\boldsymbol{\theta}_1}{\mathrm{Argmax}}Q(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2^{(t)} \mid \boldsymbol{\theta}^{(t)}) \tag{2.3}$$

CM-step2 : 关于 $\boldsymbol{\theta}_2$ 极大化 $Q(\boldsymbol{\theta})$，得到得到 $\boldsymbol{\theta}_2$ 的估计值

$$\boldsymbol{\theta}_2^{(t+1)} = \underset{\boldsymbol{\theta}_2}{\mathrm{Argmax}} Q(\boldsymbol{\theta}_1^{(t+1)}, \boldsymbol{\theta}_2 \mid \boldsymbol{\theta}^{(t)}) \tag{2.4}$$

• ECME 算法由Liu 和 Rubin (1994)提出，是 ECM 算法的拓展。与 ECM 算法只有 CMQ 步骤不同，ECME 算法还有 CML 步骤，即存在极大化 $\mathcal{L}(\boldsymbol{\theta})$ 的步骤。

若参数 $\boldsymbol{\theta}$ 划分成 $\boldsymbol{\theta}_1$ 和 $\boldsymbol{\theta}_2$，则 ECM 算法的 **CM-step2** 应替换为 CML 步骤，即

$$\boldsymbol{\theta}_2^{(t+1)} = \underset{\boldsymbol{\theta}_2}{\mathrm{Argmax}} \mathcal{L}(\boldsymbol{\theta}_1^{(t+1)}, \boldsymbol{\theta}_2) \tag{2.5}$$

第二节　　矩阵正态分布和因子化主成分分析

一、矩阵正态分布

定义 2.1　若矩阵 $\mathbf{X} \in \mathbb{R}^{d_c \times d_r}$ 服从矩阵正态分布 $\mathbf{X} \sim \mathcal{N}_{d_c,d_r}(\mathbf{M}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r)$，其中，$\mathbf{M} \in \mathbb{R}^{d_c \times d_r}$ 为均值矩阵，$\boldsymbol{\Sigma}_c \in \mathbb{R}^{d_c \times d_c}$ 为行协方差矩阵，$\boldsymbol{\Sigma}_r \in \mathbb{R}^{d_r \times d_r}$ 为列协方差矩阵，向量化后 $\mathbf{vec}(\mathbf{X}) \sim \mathcal{N}_{d_c d_r}(\mathbf{vec}(\mathbf{M}), \boldsymbol{\Sigma}_r \otimes \boldsymbol{\Sigma}_c)$。$\mathbf{X}$ 的概率密度函数为：

$$p(\mathbf{X}) = (2\pi)^{-\frac{d_c d_r}{2}} |\boldsymbol{\Sigma}_c|^{-\frac{d_r}{2}} |\boldsymbol{\Sigma}_r|^{-\frac{d_c}{2}} \exp\{-\frac{1}{2}\delta_{\mathbf{X}}(\boldsymbol{\theta})\} \tag{2.6}$$

其中

$$\delta_{\mathbf{X}}(\boldsymbol{\theta}) = \mathrm{tr}[\boldsymbol{\Sigma}_c^{-1}(\mathbf{X} - \mathbf{M})\boldsymbol{\Sigma}_r^{-1}(\mathbf{X} - \mathbf{M})'] \tag{2.7}$$

由定义 2.1可知，所谓矩阵正态分布，就是在多元正态分布的基础上具有可分离的协方差结构 $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_r \otimes \boldsymbol{\Sigma}_c$。

二、因子化主成分分析

在主成分分析 (PCA) 中，假定变量 $\mathbf{x}$ 服从多元 t 分布，其协差阵为 $\boldsymbol{\Sigma}$，对 $\mathbf{x}$ 做正交线性变换 $\mathbf{y} = \mathbf{T}'\mathbf{x}$，其中 $\mathbf{T} \in \mathbb{R}^{d \times q}$ 且 $q < d$，那么 $\mathbf{y}$ 的协差阵为 $\mathbf{T}'\boldsymbol{\Sigma}\mathbf{T}$。PCA 的目的是在目标函数中寻找最优的矩阵 $\mathbf{T}$(Jolliffe, 2002)，即

$$\max_{\mathbf{T}} \mathrm{tr}\{\mathbf{T}'\boldsymbol{\Sigma}\mathbf{T}\}, s.t. \mathbf{T}'\mathbf{T} = \mathbf{I}_q \tag{2.8}$$

若变量服从矩阵正态分布，协差阵被分解为 $\mathbf{\Sigma}=\mathbf{\Sigma}_r\otimes\mathbf{\Sigma}_c$，相应地，矩阵 $\mathbf{T}$ 被分解为 $\mathbf{T}=\mathbf{T}_r\otimes\mathbf{T}_c$，PCA 的目标函数则拆解成两个相互独立的子目标函数，即

$$\max_{\mathbf{T}_c}\operatorname{tr}\{\mathbf{T}_c'\mathbf{\Sigma}_c\mathbf{T}_c\}, s.t. \mathbf{T}_c'\mathbf{T}_c=\mathbf{I}_{q_c} \quad \text{和} \quad \max_{\mathbf{T}_r}\operatorname{tr}\{\mathbf{T}_r'\mathbf{\Sigma}_r\mathbf{T}_r\}, s.t. \mathbf{T}_r'\mathbf{T}_r=\mathbf{I}_{q_r} \tag{2.9}$$

$\mathbf{\Sigma}_c$ 和 $\mathbf{\Sigma}_r$ 的特征值与特征向量序列分别为 $\{(\mathbf{u}_i,\lambda_i)\}_{i=1}^c$ 和 $\{(\mathbf{v}_i,\eta_i)\}_{i=1}^r$，那么式 (2.9) 中优化问题的解分别为

$$\mathbf{T}_c=[\mathbf{u}_1,\ldots,\mathbf{u}_{q_c}], \mathbf{T}_r=[\mathbf{u}_1,\ldots,\mathbf{r}_{q_r}], \mathbf{\Sigma}_c\text{和}\mathbf{\Sigma}_r\text{的前}q_c,q_r\text{个特征向量} \tag{2.10}$$

由特征值所构成的对角矩阵为 $\mathbf{\Lambda}_c=\operatorname{diag}\{\lambda_1,\lambda_2,\ldots,\lambda_{q_c}\}$ 和 $\mathbf{\Lambda}_r=\operatorname{diag}\{\eta_1,\eta_2,\ldots,\eta_{q_r}\}$。在矩阵正态分布的假定条件下，分别由式 (2.12) 中的 $\widetilde{\mathbf{\Sigma}}_c$ 和式 (2.13) 中的 $\widetilde{\mathbf{\Sigma}}_r$ 得到 $\mathbf{\Sigma}_c$ 和 $\mathbf{\Sigma}_r$ 的估计。

三、矩阵正态分布的极大似然估计

$\mathcal{X}=\{\mathbf{X}_n\}_{n=1}^N$ 是一组事先给定的观测，参数为 $\boldsymbol{\theta}=(\mathbf{M},\mathbf{\Sigma}_c,\mathbf{\Sigma}_r)$，观测数据的对数似然函数 $\mathcal{L}$(省略常数项) 为

$$\mathcal{L}(\boldsymbol{\theta}|\mathcal{X})=-\frac{1}{2}\sum_{n=1}^N\{d_r\ln|\mathbf{\Sigma}_c|+d_c\ln|\mathbf{\Sigma}_r|+\delta_{\mathbf{X}_n}(\boldsymbol{\theta})\} \tag{2.11}$$

其中 $\delta_{\mathbf{X}_n}(\boldsymbol{\theta})$ 由式 (2.7) 得到。

参数 $\mathbf{M}$ 的极大似然估计为 $\overline{\mathbf{X}}=-\frac{1}{N}\sum_{n=1}^N\mathbf{X}_n$，令 $\mathbf{M}=\overline{\mathbf{X}}$。利用 CM 算法(Liu, Rubin, 1994) 迭代极大化似然函数 $\mathcal{L}$ 可得到 $\mathbf{\Sigma}_c$ 和 $\mathbf{\Sigma}_r$ 的极大似然估计，分为两步：

CM-step1 : 给定 $\mathbf{\Sigma}_r$，关于 $\mathbf{\Sigma}_c$ 极大化似然函数 $\mathcal{L}$ 得

$$\widetilde{\mathbf{\Sigma}}_c=\frac{1}{Nd_r}\sum_{n=1}^N(\mathbf{X}_n-\overline{\mathbf{X}})\mathbf{\Sigma}_r^{-1}(\mathbf{X}_n-\overline{\mathbf{X}})' \tag{2.12}$$

CM-step2 : 给定由 **CM-step1** 得到的 $\widetilde{\mathbf{\Sigma}}_c$，关于 $\mathbf{\Sigma}_r$ 极大化似然函数 $\mathcal{L}$ 得

$$\widetilde{\mathbf{\Sigma}}_r = \frac{1}{Nd_c}\sum_{n=1}^{N}(\mathbf{X}_n - \overline{\mathbf{X}})'\widetilde{\mathbf{\Sigma}}_c^{-1}(\mathbf{X}_n - \overline{\mathbf{X}}) \tag{2.13}$$

循环 **CM-step1** 和 **CM-step2** 两个步骤，当对数似然 $\mathcal{L}$ 的值小于某个给定的取值时，退出迭代过程。

第三节　矩阵变元 t 分布和基于矩阵变元 t 分布的稳健因子主成分分析

为方便后续推导使用,先对伽玛分布的相关知识进行简要介绍(Bishop, Nasrabadi, 2006)。

若变量 τ 服从伽玛分布 $\tau \sim \mathrm{Ga}(\alpha, \beta)$，$\tau$ 和 $\ln\tau$ 的期望分别为

$$\mathbb{E}[\tau] = \frac{\alpha}{\beta} \tag{2.14}$$

$$\mathbb{E}[\ln\tau] = \psi(\alpha) - \ln\beta \tag{2.15}$$

其中 $\psi(x) = d\ln(\Gamma(x))/dx$ 为 digamma 函数。

一、矩阵变元 t 分布

定义 2.2　称矩阵 $\mathbf{X} \in \mathbb{R}^{d_c\times d_r}$ 服从矩阵变元 t 分布 $\mathbf{X} \sim Mt_{d_c,d_r}(\mathbf{M}, \mathbf{\Sigma}_c, \mathbf{\Sigma}_r, \nu)$，其中 $\mathbf{M} \in \mathbb{R}^{d_c\times d_r}$ 为均值矩阵，$\mathbf{\Sigma}_c \in \mathbb{R}^{d_c\times d_c}$ 为行协方差矩阵，$\mathbf{\Sigma}_r \in \mathbb{R}^{d_r\times d_r}$ 为列协方差矩阵且 ν 为自由度，向量化后 $\mathbf{vec}(\mathbf{X}) \sim t_{d_cd_r}(\mathbf{vec}(\mathbf{M}), \mathbf{\Sigma}_r\otimes\mathbf{\Sigma}_c, \nu)$。$\mathbf{X}$ 的概率密度函数为:

$$p(\mathbf{X}) = \frac{|\mathbf{\Sigma}_c|^{-\frac{d_r}{2}}|\mathbf{\Sigma}_r|^{-\frac{d_c}{2}}\Gamma(\frac{\nu+d_cd_r}{2})}{(\pi\nu)^{\frac{d_cd_r}{2}}\Gamma(\frac{\nu}{2})}\left(1+\frac{\delta_{\mathbf{X}}(\boldsymbol{\theta})}{\nu}\right)^{-\frac{\nu+d_cd_r}{2}} \tag{2.16}$$

其中 $\delta_{\mathbf{X}}(\boldsymbol{\theta})$ 见于式 (2.7)。

矩阵 $\mathbf{X}$ 可以在引入一个权重变量 τ 的条件下表示为一个潜变量模型。在 τ 服从伽玛分布 $\mathrm{Ga}(\nu/2, \nu/2)$ 的条件下，矩阵 $\mathbf{X} \in \mathbb{R}^{d_c\times d_r}$ 服从矩阵变元正态分布

$\mathcal{N}_{d_c,d_r}(\mathbf{W},\mathbf{\Sigma}_c/\tau,\mathbf{\Sigma}_r)$，即

$$\begin{cases} \mathbf{X}|\tau \sim \mathcal{N}_{d_c,d_r}(\mathbf{W},\mathbf{\Sigma}_c/\tau,\mathbf{\Sigma}_r) \\ \tau \sim \mathrm{Ga}(\nu/2,\nu/2) \end{cases} \tag{2.17}$$

根据式 (2.17),可得边际分布 $\mathbf{X} \sim Mt_{d_c,d_r}(\mathbf{M},\mathbf{\Sigma}_c,\mathbf{\Sigma}_r,\nu)$ 或 $\mathbf{vec}(\mathbf{X}) \sim t_{d_cd_r}(\mathbf{vec}(\mathbf{M}),\mathbf{\Sigma}_r\otimes\mathbf{\Sigma}_c,\nu)$。$\mathbf{X}$，其协方差矩阵为

$$\mathrm{cov}(\mathrm{vec}(\mathbf{X})) = \frac{\nu}{\nu-2}\mathbf{\Sigma}_r\otimes\mathbf{\Sigma}_c \tag{2.18}$$

其后验分布为

$$\tau|\mathbf{X} \sim \mathrm{Ga}(\frac{\nu+d_cd_r}{2},\frac{\nu+\delta_{\mathbf{X}}(\boldsymbol{\theta})}{2}) \tag{2.19}$$

二、稳健因子主成分分析

在矩阵变元 t 分布的假定条件下，存在式 (2.18) 中给出的协差阵的分解。在省略常数的情形下，将分解 $\mathbf{\Sigma}=\mathbf{\Sigma}_r\otimes\mathbf{\Sigma}_c$ 代入式 (2.9) 和式 (2.10)，得到 RFPCA 中优化问题的解 $\mathbf{T}_c$ 和 $\mathbf{T}_r$，$\mathbf{\Sigma}_c$ 和 $\mathbf{\Sigma}_r$ 的估计分别由式 (2.25) 中的 $\widetilde{\mathbf{\Sigma}}_c$ 和式 (2.26) 中的 $\widetilde{\mathbf{\Sigma}}_r$ 得到。

三、矩阵变元 t 分布的极大似然估计

$\mathcal{X}=\{\mathbf{X}_n\}_{n=1}^N$ 是一组事先给定的观测，参数 $\boldsymbol{\theta}=(\mathbf{M},\mathbf{\Sigma}_c,\mathbf{\Sigma}_r,\nu)$，根据式 (2.16)，观测数据的对数似然函数 $\mathcal{L}$ 为

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}|\mathcal{X}) = &-\frac{N}{2}[d_cd_r\ln\pi + d_r\ln|\mathbf{\Sigma}_c| + d_c\ln|\mathbf{\Sigma}_r| + 2\ln\Gamma(\frac{\nu}{2}) - 2\ln\Gamma(\frac{\nu+d_cd_r}{2}) \\ &-\nu\ln\nu] - \frac{\nu+d_cd_r}{2}\sum_{n=1}^N\ln(\nu+\delta_{\mathbf{X}_n}(\boldsymbol{\theta}))\end{aligned} \tag{2.20}$$

其中 $\delta_{\mathbf{X}_n}(\boldsymbol{\theta})$ 由式 (2.7) 得到。直接极大化式 (2.20) 中的 $\mathcal{L}$ 来估计参数是很困难的，所以利用 EM 型算法来进行参数估计。

• ECM 算法含有一个求条件期望的步骤 (E-step) 以及四个在给定参数下极大化条件期望的步骤 (CM-steps)：

E-step：给定一组缺失数据 $\boldsymbol{\tau}=\{\tau_n\}_{n=1}^N$，此时完全数据的对数似然函数可表示为 $\sum_{n=1}^N\ln\{p(\mathbf{X}_n|\tau_n)p(\tau_n)\}$，将其关于条件分布 $p(\tau|\mathcal{X},\boldsymbol{\theta})$ 求期望，可得

$$
\begin{aligned}
\mathcal{Q}(\boldsymbol{\theta}|\mathcal{X}) = & -\frac{N}{2}[d_c d_r \ln(2\pi) + d_r \ln|\boldsymbol{\Sigma}_c| + d_c \ln|\boldsymbol{\Sigma}_r|] + \frac{d_c d_r}{2}\sum_{n=1}^{N}\mathbb{E}(\ln\tau_n|\mathbf{X}_n) \\
& - \frac{1}{2}\sum_{n=1}^{N}\mathbb{E}(\tau_n|\mathbf{X}_n)\delta_{\mathbf{X}_n}(\boldsymbol{\theta}) + N[\frac{\nu}{2}\ln(\frac{\nu}{2}) - \ln\Gamma(\frac{\nu}{2})] \\
& + (\frac{\nu}{2}-1)\sum_{n=1}^{N}\mathbb{E}(\ln\tau_n|\mathbf{X}_n) - \frac{\nu}{2}\sum_{n=1}^{N}\mathbb{E}(\tau_n|\mathbf{X}_n)
\end{aligned}
\tag{2.21}
$$

由式 (2.14)、(2.15) 和 (2.19) 可得上式 (2.21) 中的期望为

$$\mathbb{E}[\tau_n|\mathbf{X}_n] = \frac{\nu + d_c d_r}{\nu + \delta_{\mathbf{X}_n}(\boldsymbol{\theta})} \tag{2.22}$$

$$\mathbb{E}[\ln\tau_n|\mathbf{X}_n] = \psi(\frac{\nu + d_c d_r}{2}) - \ln(\frac{\nu + \delta_{\mathbf{X}_n}(\boldsymbol{\theta})}{2}) \tag{2.23}$$

其中 $\delta_{\mathbf{X}_n}(\boldsymbol{\theta})$ 由式 (2.7) 得到，$\mathbb{E}[\tau_n|\mathbf{X}_n]$ 是权重的期望。

CM-step1 : 在给定 $(\boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r, \nu)$ 的条件下，关于 **M** 极大化 **E-step** 中的 $\mathcal{Q}$，得

$$\widetilde{\mathbf{M}} = \frac{\sum_{n=1}^{N}\mathbb{E}[\tau_n|\mathbf{X}_n]\mathbf{X}_n}{\sum_{n=1}^{N}\mathbb{E}[\tau_n|\mathbf{X}_n]} \tag{2.24}$$

CM-step2 : 得到 **CM-step1** 中的 $\widetilde{\mathbf{M}}$ 后，给定 $(\widetilde{\mathbf{M}}, \boldsymbol{\Sigma}_r, \nu)$，关于 $\boldsymbol{\Sigma}_c$ 极大化 $\mathcal{Q}$，得

$$\widetilde{\boldsymbol{\Sigma}}_c = \frac{1}{Nd_r}\sum_{n=1}^{N}\mathbb{E}[\tau_n|\mathbf{X}_n](\mathbf{X}_n - \widetilde{\mathbf{M}})\boldsymbol{\Sigma}_r^{-1}(\mathbf{X}_n - \widetilde{\mathbf{M}})' \tag{2.25}$$

CM-step3 : 得到 **CM-step2** 中的 $\widetilde{\boldsymbol{\Sigma}}_c$ 后，给定 $(\widetilde{\mathbf{M}}, \widetilde{\boldsymbol{\Sigma}}_c, \nu)$，关于 $\boldsymbol{\Sigma}_r$ 极大化 $\mathcal{Q}$，得

$$\widetilde{\boldsymbol{\Sigma}}_r = \frac{1}{Nd_c}\sum_{n=1}^{N}\mathbb{E}[\tau_n|\mathbf{X}_n](\mathbf{X}_n - \widetilde{\mathbf{M}})'\widetilde{\boldsymbol{\Sigma}}_c^{-1}(\mathbf{X}_n - \widetilde{\mathbf{M}}) \tag{2.26}$$

CM-step4 : 得到 **CM-step3** 中的 $\widetilde{\boldsymbol{\Sigma}}_r$ 后，给定 $(\widetilde{\mathbf{M}}, \widetilde{\boldsymbol{\Sigma}}_c, \widetilde{\boldsymbol{\Sigma}}_r)$，关于 ν 极大化 $\mathcal{Q}$，得

$$\widetilde{\nu} = \frac{2}{y + \ln y - 1} + 0.0416 \times [1 + erf(0.6594 \times \ln\frac{2.1971}{y + \ln y - 1})] \tag{2.27}$$

其中 $y=-\frac{1}{N}\sum_{n=1}^{N}\{\mathbb{E}[\ln\tau_n|\mathbf{X}_n]-\mathbb{E}[\tau_n|\mathbf{X}_n]\}$，$erf(x)=\frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)\,\mathrm{d}t$

• ECME 算法与 ECM 算法一样，也分为 E-step 和 CM-steps，但 ECME 对于参数 ν 的更新是通过极大化对数似然函数 $\mathcal{L}$ 实现的，即

CM-step4：给定 $(\widetilde{\mathbf{M}},\widetilde{\mathbf{\Sigma}}_c,\widetilde{\mathbf{\Sigma}}_r)$，关于 ν 极大化式 (2.20) 中的对数似然 $\mathcal{L}$，让 $\mathcal{L}$ 对于 ν 的导数为 0，可得

$$\begin{aligned}1+\psi(\frac{\nu+d_cd_r}{2})-\ln(\frac{\nu+d_cd_r}{2})-\psi(\frac{\nu}{2})+\ln\frac{\nu}{2}\\+\frac{1}{N}\sum_{n=1}^{N}[\ln\frac{\nu+d_cd_r}{\nu+\delta_{\mathbf{X}_n}(\theta)}-\frac{\nu+d_cd_r}{\nu+\delta_{\mathbf{X}_n}(\theta)}]=0\end{aligned}\tag{2.28}$$

ν 的更新 $\widetilde{\nu}$ 通过求解上述式子得到。

算法 1 展示了 RFPCA 的完整的 ECM 算法和 ECME 算法。

算法 1: RFPCA 的 ECM 算法 (ECME 算法)

输入: 观测数据 $\mathcal{X}$ 和参数 $\boldsymbol{\theta}=(\mathbf{M},\mathbf{\Sigma}_c,\mathbf{\Sigma}_r,\nu)$ 的初始值

1 循环下述五个步骤:

2 E-step: 计算条件期望 $\mathbb{E}[\tau_n|\mathbf{X}_n]=(\nu+d_cd_r)/(\nu+\delta_{\mathbf{X}_n}(\boldsymbol{\theta}))$

3 CM-step1: $\widetilde{\mathbf{M}}=\sum_{n=1}^{N}\mathbb{E}[\tau_n|\mathbf{X}_n]\mathbf{X}_n/\sum_{n=1}^{N}\mathbb{E}[\tau_n|\mathbf{X}_n]$

4 CM-step2: $\widetilde{\mathbf{\Sigma}}_c=\sum_{n=1}^{N}\mathbb{E}[\tau_n|\mathbf{X}_n](\mathbf{X}_n-\widetilde{\mathbf{M}})\mathbf{\Sigma}_r^{-1}(\mathbf{X}_n-\widetilde{\mathbf{M}})'/Nd_r$

5 CM-step3: $\widetilde{\mathbf{\Sigma}}_r=\sum_{n=1}^{N}\mathbb{E}[\tau_n|\mathbf{X}_n](\mathbf{X}_n-\widetilde{\mathbf{M}})'\widetilde{\mathbf{\Sigma}}_c^{-1}(\mathbf{X}_n-\widetilde{\mathbf{M}})/Nd_c$

6 CM-step4:

$\widetilde{\nu}=2/(y+\ln y-1)+0.0416\times[1+erf(0.6594\times\ln(2.1971/(y+\ln y-1)))]$

(通过式 (2.28) 得到 $\widetilde{\nu}$ 的解)

7 直到对数似然 $\mathcal{L}$ 的值小于某个给定的取值

输出: $\widetilde{\boldsymbol{\theta}}=(\widetilde{\mathbf{M}},\widetilde{\mathbf{\Sigma}}_c,\widetilde{\mathbf{\Sigma}}_r,\widetilde{\nu})$

四、极大似然估计不存在的情形

在高维矩阵型数据上，RFPCA 利用极大似然估计 (ML Estimation, MLE) 来对参数列协方差矩阵 $\mathbf{\Sigma}_c$ 和行协方差矩阵 $\mathbf{\Sigma}_r$ 进行估计时，容易存在矩阵奇异或退化问题，此时，无法得到 $\mathbf{\Sigma}_c$ 和 $\mathbf{\Sigma}_r$ 的 ML 估计。

MLE 的存在性与矩阵维数和样本量 $(m1,m2,n)$ 相关，Drton et al. (2021)提出了三种 ML 阈值，即存在性阈值 $N_e(m1,m2)$、唯一性阈值 $N_u(m1,m2)$ 和有界阈值 $N_b(m1,m2)$，分别表示：若 $n\geq N_e(m1,m2)$，MLE 存在，若 $n\geq N_u(m1,m2)$，有唯一的 MLE 值，若 $n\geq N_b(m1,m2)$，ML 有界。三种 ML 阈值满足