低秩矩阵完整化问题的几种高效解法

林陈冉

2017年1月1日

摘要 本文是在阅读 *Solving Low-Rank Matrix Completion Problems Efficiently*[3] 后所写的读书报告. 文章大部分内容依照原文翻译, 有些是查阅其他文献所得, 也加入了一些自己的理解和证明. 文章首先提出了低秩矩阵完整化的几个数学模型, 再对其中三个模型分别提出了一种较为高效的算法, 分别是ADAL法, RBR法和FPCA法. 第一种算法ADAL法, 是考虑原问题的对偶问题, 再用增广拉格朗日函数, 采用交替方向法求解. 第二种算法RBR法, 是基于半定规划, 采用逐行求解不断迭代的方式. 第三种算法FPCA法, 主要是对FPC法的奇异值计算的过程进行了优化, 通过求一个低秩的矩阵的奇异值分解来近似分解整个原矩阵.

1 简介

1.1 什么是低秩矩阵完整化

简单来说, 低秩矩阵完整化问题, 就是在仅仅知道矩阵的少部分元素的情况下, 恢复出这个矩阵的所有元素. 显而易见, 补全一个完全随机的矩阵几乎是不可能的, 也是意义不大的. 一般情况下, 我们认为所需要补全的矩阵是有一定规律的, 也就是说, 这个矩阵的秩比较小. 这个问题在统计, 图像处理, 计算几何, 机器学习, 信号处理, 模型控制等方面有广泛应用, 比如著名的NetFlix大奖赛问题.

1.2 数学模型

通常我们最感兴趣的是这样的一个最优化问题:

$$\begin{aligned} \min \quad & rank(X) \\ \text{s.t.} \quad & X_{ij} = M_{ij}, \forall (i,j) \in \Omega \end{aligned} \tag{1.1}$$

其中 $X, M \in \mathbb{R}^{p\times q}$, Ω 是已知元素的下标 (i,j) 构成的集合, $|\Omega| = m$.

(1.1)是下面的问题的特殊情况:

$$\begin{aligned} \min \quad & rank(X) \\ \text{s.t.} \quad & \mathcal{A}(X) = b \end{aligned} \tag{1.2}$$

其中$b=(b_1,\cdots,b_m)\in\mathbb{R}^m$, $\mathcal{A}:\mathbb{R}^{p\times q}\to\mathbb{R}^m$ 是线性算子.

但是(1.1), (1.2)都是"NP-难"的[7], 因此需要一定的转化. 这里我们用核范数来近似矩阵的秩, 把(1.1)与(1.2)转化为以下形式:

$$\begin{aligned}\min\quad &\|X\|_*\\ \text{s.t.}\quad &\mathcal{A}(X)=b\end{aligned}\tag{1.3}$$

$$\begin{aligned}\min\quad &\|X\|_*\\ \text{s.t.}\quad X_{ij}=&M_{ij},\forall(i,j)\in\Omega\end{aligned}\tag{1.4}$$

很自然的, 一个重要的问题是: (1.1)与(1.3), 或者(1.2)与(1.4)什么时候等价? 略过证明, 直接描述以下重要的结论:

对于(1.3), Candes和陶哲轩等给出了一个证明. 当满足某些正则条件, 若已知元素个数 $|\Omega|=m=O(nr\cdot polylog(n))$, 其中 $n=\max\{p,q\}$, $polylog$ 是多重对数函数, 则矩阵有很高概率可以通过(1.3)恢复.[1]

对于(1.4), Recht等给出了一个证明. 将线性映射 $\mathcal{A}$ 的矩阵形式记作 A , 即 $\mathcal{A}(X)=A\mathrm{vec}(X)$, 其中 $\mathrm{vec}(X)\in\mathbb{R}^{pq}$ 为矩阵 X 的向量化. 当 A 是一个随机高斯矩阵, 若向量 b 的维数 $m=O(r(p+q)log(pq))$, 则矩阵有很高概率可以通过(1.4)恢复.[8]

实际上, 由线性代数知识容易知道, (1.4)要求的线性约束条件并不总能成立, 因此有时需要适当松弛. 考虑(1.4)的罚函数:

$$\min\quad \|X\|_*+\frac{1}{2\mu}\|\mathcal{A}(X)-b\|_2^2\tag{1.5}$$

其中 μ 是某个给定常数.

本文主要将对(1.3), (1.4), (1.5)分别给出一种高效的解法.

1.3 一些自己的理解

1.3.1 $\mathcal{A}$ 与$\mathcal{A}^*$

我们可以考虑下(1.2)中给出的线性算子 $\mathcal{A}$ 到底是什么. $\mathcal{A}:\mathbb{R}^{p\times q}\to\mathbb{R}^m$ 可以看作 m 个线性算子的组合, 即 $\mathcal{A}(X)=(\mathcal{A}_1(X),\cdots,\mathcal{A}_m(X))$, $(\mathcal{A}_i:\mathbb{R}^{p\times q}\to\mathbb{R})\in(\mathbb{R}^{p\times q})^*$, $i=1,\cdots,m$.

赋予矩阵内积后, $\mathbb{R}^{p\times q}$ 可以看作Hilbert的, $(\mathbb{R}^{p\times q})^* = \mathbb{R}^{p\times q}$. 由Ries表示定理, $\exists A_i \in \mathbb{R}^{p\times q}$, s.t. $\mathcal{A}_i(X) = \langle A_i, X\rangle$, 则 $\mathcal{A}(X) = (\langle A_1, X\rangle, \cdots, \langle A_m, X\rangle)$. 这说明从 $\mathbb{R}^{p\times q}$ 到 $\mathbb{R}^m$ 的线性映射实质上就是 m 个属于 $\mathbb{R}^{p\times q}$ 的矩阵.

我们还可以考虑 $\mathcal{A}^*$, 显然 $\mathcal{A}^* : \mathbb{R}^m \to \mathbb{R}^{p\times q}$.

$$\langle \mathcal{A}(X), y\rangle = \langle ((\langle A_i, X\rangle), (y_i)\rangle = \sum_{i=1}^{m} y_i \langle A_i, X\rangle = \langle X, \sum_{i=1}^{m} y_i A_i\rangle = \langle X, \mathcal{A}^*(y)\rangle$$

故 $\mathcal{A}^*(y) = \sum_{i=1}^{m} y_i A_i$.

1.3.2 核范数与秩

大部分文献中, 都是直接使用核范数来近似秩, 并未说明为什么. 实际上, $\|A\|_* = \|\sigma(A)\|_1$, rank $A = \|\sigma(A)\|_0$, 由此可以看出, 核范数与秩的关系, 即1-范数与0-范数的关系, 故用核范数来近似秩.

2 预备知识

2.1 矩阵范数

以下默认 $A \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$, $k = \max\{m, n\}$, $\sigma(A) = (\sigma_1(A), \cdots, \sigma_k(A)) \in \mathbb{R}^k$ 是矩阵 A 的奇异值向量, $\sigma_i(A)$ 是矩阵 A 的第i个奇异值.

2.1.1 Schatten-p范数

矩阵的Schatten-p范数是矩阵奇异值向量的p-范数[9]

$$\|A\|_{S_p} = (\sum_{i}^{k} \sigma_i(A)^p)^{\frac{1}{p}} = \|\sigma(A)\|_p$$

2.1.2 诱导-p范数

矩阵的诱导-p范数[6]是将矩阵看作线性算子, 由向量的p-范数诱导而来

$$\|A\|_p = \sup_{\|x\|_p \le 1} \|Ax\|_p$$

常见的有2-范数, 又称谱范数:

$$\|A\|_2 = \max_i \sigma_i(A)$$

实际上2-范数也是 $p = \infty$ 时的Schatten-p范数.

1-范数, 又称列范数:

$$\|A\|_1 = \max_{1\le j\le n}\sum_{i=1}^{m}|a_{ij}|$$

∞-范数, 又称行范数:

$$\|A\|_\infty = \max_{1\le i\le m}\sum_{j=1}^{n}|a_{ij}|$$

2.1.3 核范数

矩阵的核范数, 又称迹范数, 是在 $p=1$ 时的Schatten-p范数[6]

$$\|A\|_* = \sum_{i}^{k}\sigma_i(X) = tr(\sqrt{AA^\top}) = \|A\|_{tr}$$

特别的, 当 A 是对称正定的方阵, 则有 $A=\sqrt{AA^\top}$, 即

$$\|A\|_* = tr(A)$$

2.1.4 F-范数

矩阵的F-范数定义类似于向量的2-范数[6]

$$\|A\|_F = (\sum_{i,j}a_{ij}^2)^{\frac{1}{2}}$$

定义矩阵内积 $\langle A,X\rangle = tr(A^\top X)$, F-范数与矩阵内积是相容的, 即 $\|A\|_F = \langle A,A\rangle^{\frac{1}{2}}$, 可以看作内积诱导的范数. F-范数也是 $p=2$ 时的Schatten-p范数.

2.2 罚函数

对于约束优化问题:

$$\begin{aligned}\min_x \quad & f(x)\\ \text{s.t.} \quad & c_i(x)=0, \quad i=1,\cdots,m_e;\\ & c_i(x)\ge 0, \quad i=m_e+1,\cdots,m\end{aligned} \tag{2.1}$$

我们可以通过把约束条件平方后, 乘上一个很大的常数 μ , 再加到目标函数上, 转化为罚函数[12]

$$P(x,\mu) = f(x) + \mu\|\bar{c}(x)\|_2^2 \tag{2.2}$$

其中 $\bar{c}(x) = (\bar{c}_1(x),\cdots,\bar{c}_m(x))\in\mathbb{R}^m$, $\bar{c}_i(x)$如下定义: $\bar{c}_i(x) = \begin{cases} c_i(x), & i\le m_e\\ \max\{0,c_i(x)\}, & i>m_e\end{cases}$.

当 $\mu \to \infty$, (2.3)和原问题(2.1)等价.

$$\min_{x} \quad P(x,\mu) \tag{2.3}$$

但在实际中, μ 无法取到无穷大, 为了克服这一缺点, 可以定义增广Lagrange函数[12](以下叙述仅考虑等值约束的情况, 即 $m_e = m$):

$$P(x,\lambda,\mu) = f(x) - \lambda^\top c(x) + \frac{1}{2}\mu\|c(x)\|_2^2 \tag{2.4}$$

其中 λ , λ , μ, $c(x) \in \mathbb{R}^m$ 给定.

(2.4)等价于以下的Powell罚函数[12]

$$P(x,\theta,\mu) = f(x) + \frac{1}{2}\mu\|c(x)-\theta\|_2^2 \tag{2.5}$$

其中 $c(x), \theta \in \mathbb{R}^m$.

(2.3)与(2.5)只相差一个与 x 无关的常数 $\frac{1}{2}\sigma\|\theta\|_2^2$.[12]

2.3 对偶问题

对于约束优化问题(2.1),它的Langrange函数为

$$L(x,\mu,\lambda) = f(x) + \sum_{i=1}^{m_e}\mu_i c_i(x) + \sum_{i=m_e+1}^{m}\lambda_i c_i(x) \tag{2.6}$$

定义 $g(\mu,\lambda) = \min_x L(x,\mu,\lambda)$, 通过这个函数我们可以把(2.1)转化为它的对偶问题[12]

$$\begin{aligned} \max_{\mu,\lambda} \quad & g(\mu,\lambda) \\ \text{s.t.} \quad & \mu \succeq 0 \end{aligned} \tag{2.7}$$

当满足一定条件(KKT条件)时, (2.7)和(2.1)是等价的.

2.4 次梯度

$f: E \to \mathbb{R}$ 是一个定义在 $\mathbb{R}^n$ 的凸子集 E 上的实值函数,则所有满足

$$f(y) - f(x) \geq u(y-x), \quad \forall y \in E \tag{2.8}$$

的向量 u 称为 f 在 x 点的次梯度, 所有这样的 u 构成的集合称为 f 在 x 点的次梯度集.[12]

次梯度集通常用 $\partial f(x)$ 表示

$$\partial f(x) = \{u \in \mathbb{R}^m | f(y) - f(x) \geq u(y-x), \quad \forall y \in E\} \tag{2.9}$$

当函数在 x 点可微, $\partial f(x) = \{f'(x)\}$; 当函数在 x 点不可微, $\partial f(x)$ 是一个非空的紧凸集.

对于不可微函数或者不可微点, 我们可以采用次梯度替代梯度进行分析. 众所周知, 一个点是最优解的必要条件是该点梯度为0. 相应的,有如下结论:

定理 2.1. *x^* 是(2.1)问题的最优解的一个必要条件是: $0 \in \partial f(x^*)$. 特别的, 若 f 是凸函数, 则是充要条件.*

特别指出, 核范数是非光滑函数的, 次梯度集如下[10]

$$\partial \|X\|_* = \{UV^\top + W | U^\top W = 0, WV = 0, \|W\|_2 \leq 1\} \tag{2.10}$$

其中 $X = UDV^\top$.

3 交替方向增广Lagrange法

3.1 问题转化

命题 3.1. *问题(1.3)的对偶问题是*

$$\begin{aligned} \max_{y \in \mathbb{R}^m} \quad & b^\top y \\ s.t. \quad & \|\mathcal{A}^*(y)\|_2 \leq 1 \end{aligned} \tag{3.1}$$

这个结论被大多数文章认为是显然的, 这里给出一个我自己的证明.

证明. 问题(1.3)的拉格朗日函数为

$$L(x,y) = \|X\|_* - (\mathcal{A}(X) - b)^\top y \tag{3.2}$$

其中 $y \in \mathbb{R}^m$. 由于 $\|X\|_*$ 不光滑, 我们考虑(3.2)的次梯度集. 又(3.2)是凸函数, 函数在某点取到极值, 等价于 0 属于该点的次梯度集, 即

$$0 \in \partial \|X\|_* - \mathcal{A}^*(y)$$

这说明 $\exists W$, $\|W\|_2 \leq 1$, $\mathcal{A}^*(y) = UV^\top + W$, 其中 $X = UDV^\top$. 由 $\|W\|_2 \leq 1$ 易知 $\|\mathcal{A}^*(y)\|_2 \leq 1$.

由核范数次梯度的定义, $U^\top \mathcal{A}^*(y) = V^\top$, $\mathcal{A}^*(y)V = U$, 即 $\mathcal{A}^*(y) = UV^\top$, 那么可以知道

$$\begin{aligned} g(y) =& \min_x L(x,y) \\ =& \|X\|_* - \mathcal{A}(X)^\top y + b^\top y \\ =& \|UDV^\top\|_* - \langle UDV^\top, \mathcal{A}^*(y) \rangle + b^\top y \\ =& tr(D) - tr(VD^\top U^\top UV^\top) + b^\top y \\ =& b^\top y \end{aligned} \tag{3.3}$$

由此我们可以得到结论, 即(1.3)的对偶是(3.1). □

引入一个形式上的变量 S , 将(3.1)变为如下的等价形式

$$\begin{aligned} \min_{y \in \mathbb{R}^m} \quad & -b^\top y \\ \text{s.t.} \quad & \mathcal{A}^*(y) - S = 0 \\ & \|S\|_2 \le 1 \end{aligned} \tag{3.4}$$

可以考虑(3.4)的增广Lagrange函数

$$L(y,S,X,\mu) = -b^\top y + \langle X, \mathcal{A}^*(y) - S \rangle + \frac{1}{2\mu} \|\mathcal{A}^*(y) - S\|_F^2 \tag{3.5}$$

其中 $X \in \mathbb{R}^{p \times q}$, $\|\cdot\|_F$ 是矩阵的 $F-$范数.

由此我们得到(3.1)的一个等价问题

$$\begin{aligned} \min_{y,S} \quad & L(y,S,X,\mu) \\ \text{s.t.} \quad & \|S\|_2 \le 1 \end{aligned} \tag{3.6}$$

通过选取最佳的 μ 和 X 可以求出(3.6)的最优解.

采用如下的迭代过程以求解问题(3.6)

$$\begin{aligned} \mu^{k+1} =& \alpha \mu^k, \quad \alpha \in (0,1) \\ X^{k+1} =& X^k + \frac{\mathcal{A}^*(y^{k+1}) - S^{k+1}}{\mu^k} \\ (y^{k+1}, S^{k+1}) =& \arg\min_{y,S} L(y,S,X^k,\mu^k) \end{aligned} \tag{3.7}$$

3.2 交替方向求解

但实际上, (3.7)第3式并不容易解, 因此我们考虑使用交替方向法来求解这个方程, 即

$$y^{k+1} = \arg\min_{y,S} L(y, S^k, X^k, \mu^k) \tag{3.8}$$

$$S^{k+1} = \arg\min_{y,S} L(y^{k+1}, S, X^k, \mu^k) \tag{3.9}$$

定理 3.1. *当固定* S^k *, (3.8)的最优解为:*

$$y^{k+1} = \mu^k(b - \mathcal{A}(X^k)) + \mathcal{A}(S^k) \tag{3.10}$$

当固定 y^{k+1} *, (3.9)的最优解为:*

$$S^{k+1} = UDiag(\min\{\sigma, 1\})V^\top \tag{3.11}$$

其中 U, V *来自* $Y = \mathcal{A}^*(y^{k+1}) + \mu^k X^k$ *的奇异值分解, 即* $Y = UDiag(\sigma)V^\top$ *.*

该结果是比较容易得到的, 文章中直接忽略了, 这里给出一个我自己的证明.

证明. 首先来求解(3.8), 当固定 S^k

$$L(y, S^k, X^k, \mu^k) = -b^\top y + \mathcal{A}(X^k)^\top y - \langle X, S^k\rangle + \frac{1}{2\mu^k}\|\mathcal{A}^*(y) - S^k\|_F^2 \tag{3.12}$$

记 $T = \mathcal{A}^*(y) - S^k$, $\mathcal{A}_{ij} = (A_{1_{ij}}, \cdots, A_{m_{ij}}) \in \mathbb{R}^m$,考虑 $\|T\|_F^2$

$$\|T\|_F^2 = \sum_{i,j}((\sum_{k=1}^m y_k A_{k_{ij}}) - S_{ij}^k)^2 = \sum_{i,j}(\mathcal{A}_{ij}^\top y - S_{ij}^k)^2$$

则它对y的梯度

$$\begin{aligned}
\frac{\partial\|T\|_F^2}{\partial y} =& \frac{\partial\sum_{i,j}(\mathcal{A}_{ij}^\top y - S_{ij}^k)^2}{\partial y} \\
=& \sum_{i,j}\frac{\partial(\mathcal{A}_{ij}^\top y - S_{ij}^k)^2}{\partial y} \\
=& 2\sum_{i,j}(\mathcal{A}_{ij}^\top y - S_{ij}^k)\mathcal{A}_{ij} \\
=& 2(\sum_{i,j} T_{ij}A_{k_{ij}}) = 2(\langle A_i, T\rangle) \\
=& 2\mathcal{A}(\mathcal{A}^*(y) - S^k)
\end{aligned} \tag{3.13}$$

由此可以给出(3.12)的梯度

$$\frac{\partial L(y, S^k, X^k, \mu^k)}{\partial y} = -b + \mathcal{A}(X^k) + \frac{1}{\mu^k}\mathcal{A}(\mathcal{A}^*(y) - S^k) \tag{3.14}$$

令(3.14)式等于0, 可得

$$y^{k+1} = \mu^k(b - \mathcal{A}(X^k)) + \mathcal{A}(S^k) \tag{3.15}$$

再考虑(3.9). 当固定 y^{k+1}

$$\begin{aligned}
&L(y^{k+1}, S, X^k, \mu^k) \\
=& -b^\top y^{k+1} + \mathcal{A}(X^k)^\top y^{k+1} - \langle X, S\rangle + \frac{1}{2\mu^k}\|\mathcal{A}^*(y^{k+1}\|) - S_F^2 \\
=& \frac{1}{2\mu^k}(\|S\|_F^2 - 2\langle Y, S\rangle + \|\mathcal{A}^*(y^{k+1}\|)_F^2) - b^\top y^{k+1} + \mathcal{A}(X^k)^\top y^{k+1}
\end{aligned} \tag{3.16}$$

其中 $Y=\mathcal{A}^*(y^{k+1})+\mu^k X^k$.

$$\frac{\partial L(y^{k+1},S,X^k,\mu^k)}{\partial S}=\frac{1}{2\mu^k}(\frac{\partial\|S\|_F^2}{\partial S}-2\frac{\partial\langle Y,S\rangle}{\partial S})=\frac{S-Y}{\mu^k} \tag{3.17}$$

令(3.17)式等于0, 则可得 $S^{k+1}=Y$, 但由约束条件 $\|S\|_2\leq 1$, 需对 Y 进行修正, 即

$$S^{k+1}=U\mathrm{Diag}(\min\{\sigma,1\})V^\top \tag{3.18}$$

其中 $Y=U\mathrm{Diag}(\sigma)V^\top$. □

根据 Y 的定义, 可以简化 X^{k+1} 的计算方法

$$X^{k+1}=\frac{\mu^k X^k+\mathcal{A}^*(y^{k+1})-S^{k+1}}{\mu^k}=\frac{Y-S^{k+1}}{\mu^k} \tag{3.19}$$

重复进行上述迭代,到目标精度停止,下面给出相应算法

算法 1.

步1 给出 μ^0 , X^0 , y^0 , S^0 , ϵ , α , 设置计数器 $k=0$;
步2 若 $\frac{\|X^{k+1}-X^k\|_F}{\max\{1,\|X^k\|_F\}}\leq\epsilon$, 则停止;
步3 计算 $y^{k+1}=\mu^k(b-\mathcal{A}(X^k))+\mathcal{A}(S^k)$;
步4 计算 $Y=\mathcal{A}^*(y^{k+1})+\mu^k X^k$, 并计算其SVD: $Y=UDiag(\sigma)V^\top$;
步5 计算 $S^{k+1}=UDiag(min\{\sigma,1\})V^\top$;
步6 计算 $X^{k+1}=\frac{Y-S^{k+1}}{\mu^k}$;
步7 计算 $\mu^{k+1}=\alpha\mu^k$, $k=k+1$,转步2.

注 1. 本章除了两个自己证明的定理, 均为原文[3]的内容.

4 RBR法

4.1 SDP问题与Schur补

对于问题(1.4), 我们可以考虑把它转化为一个半定规划(SDP)问题.

一个标准的半定规划问题是

$$\begin{aligned}\min_{X\in S^n}\quad & \langle C,X\rangle\\ \text{s.t.}\quad & \mathcal{A}(X)=b, X\succ 0\end{aligned} \tag{4.1}$$

其中 $b\in\mathbb{R}^m$, $\mathcal{A}(X)=(\langle A^{(1)}X\rangle,\cdots,\langle A^{(m)}X\rangle)$, $C,A^{(i)}\in S^n$, S^n 是全体对称矩阵.

对于一个对称正定矩阵 $X \in S^n$, 我们可以把它写成分块矩阵的形式

$$X = \begin{pmatrix} \xi & y^\top \\ y & B \end{pmatrix} \tag{4.2}$$

其中$\xi \in \mathbb{R}$, $y \in \mathbb{R}^{n-1}$, $B \in S^{n-1}$.

容易验证的, X 可以表示为以下形式

$$X = \begin{pmatrix} 1 & y^\top B^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} \xi - y^\top B^{-1} y & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} 1 & 0 \\ B^{-1} y & I \end{pmatrix} \tag{4.3}$$

记 $(X/B) = \xi - y^\top B^{-1} y$, 称为X对于B的Schur补.

显然的

$$X \succeq 0 \Leftrightarrow B \succeq 0, (X/B) \geq 0 \tag{4.4}$$

4.2 SOCP问题

约定以下记号

$$X_{\alpha,\beta} = \begin{cases} x_{\alpha\beta} & \alpha, \beta \in \mathbb{R} \\ (x_{\alpha\beta_1}, \cdots, x_{\alpha\beta_n}) & \alpha \in \mathbb{R}, \beta = \{\beta_1, \cdots, \beta_n\} \\ (x_{\alpha_1\beta}, \cdots, x_{\alpha_m\beta})^\top & \alpha = \{\alpha_1, \cdots, \alpha_m\}, \beta \in \mathbb{R} \\ \begin{pmatrix} x_{\alpha_1\beta_1} & \cdots & x_{\alpha_1\beta_n} \\ \cdots & \cdots & \cdots \\ x_{\alpha_m\beta_1} & \cdots & x_{\alpha_m\beta_n} \end{pmatrix} & \alpha = \{\alpha_1, \cdots, \alpha_m\}, \beta = \{\beta_1, \cdots, \beta_n\} \end{cases} \tag{4.5}$$

$$i^c = \{1, \cdots, n\} \backslash \{i\} = \{1, \cdots, i-1, i+1, \cdots, n\} \tag{4.6}$$

令 $X = \begin{pmatrix} \xi & y^\top \\ y & B \end{pmatrix} = \begin{pmatrix} X_{i,i} & X_{i,i^c} \\ X_{i^c,i} & X_{i^c,i^c} \end{pmatrix}$, 等号在相差一个初等变化下成立. 基于(4.4), 令 i 取遍 $\{1, \cdots, n\}$, 逐行解如下的SOCP问题来解决SDP问题(4.1)

$$\begin{aligned} \min_{[\xi;y] \in \mathbb{R}^n} \quad & \bar{c}^\top \begin{pmatrix} \xi \\ y \end{pmatrix} \\ \text{s.t.} \quad & \bar{A} \begin{pmatrix} \xi \\ y \end{pmatrix} = \bar{b} \\ & (X/B) \geq \delta \end{aligned} \tag{4.7}$$

其中

$$\bar{c} = \begin{pmatrix} C_{i,i} \\ 2Ci^c, i \end{pmatrix}, \quad \bar{A} = \begin{pmatrix} A_{i,i}^{(1)} & 2A_{i,i^c}^{(1)} \\ \cdots & \cdots \\ A_{i,i}^{(m)} & 2A_{i,i^c}^{(m)} \end{pmatrix}, \quad \bar{b} = \begin{pmatrix} b_1 - \langle A_{i^c,i^c}^{(1)}, B \rangle \\ \cdots \\ b_m - \langle A_{i^c,i^c}^{(m)}, B \rangle \end{pmatrix} \tag{4.8}$$

若 X 是半正定的, 即 $X \succeq 0$, 取 $\delta = 0$;

若 X 是正定的, 即 $X \succ 0$, 用大于零的数来限制Schur补, 取$\delta > 0$.

考虑(4.7)的罚函数

$$F(X, \mu) = \bar{c}^\top \begin{pmatrix} \xi \\ y \end{pmatrix} + \frac{1}{2\mu} \|\bar{A}[\xi; y] - \bar{b}\|_2^2 \tag{4.9}$$

其中 $\mu > 0$ 是给定的.

(4.7)等价于以下问题

$$\begin{aligned} \min_X \quad & F(X, \mu) \\ \text{s.t.} \quad & X \succeq 0 \end{aligned} \tag{4.10}$$

4.3 用SDP和RBR法补全矩阵

回头考虑(1.4). 当 $X \in S^n$ 对称正定, 有 $\|X\|_* = tr(X)$, (1.4)等价于下面的SDP问题

$$\begin{aligned} \min_X \quad & tr(X) = \langle E, X \rangle \\ \text{s.t.} \quad & X_{ij} = M_{ij}, \quad \forall (i,j) \in \Omega \end{aligned} \tag{4.11}$$

当 $X \in \mathbb{R}^{p \times q}$ 不是对称正定的, 可以考虑一个更大的对称正定矩阵 W . 当补全了 W , 则 X 自然被补全了(当然, 当 X 是对称正定矩阵时, 我们也可以这么做)

$$\begin{aligned} \min_X \quad & tr(X) \\ \text{s.t.} \quad & X = \begin{pmatrix} X_1 & W \\ W^\top & X_2 \end{pmatrix} \succ 0 \\ & W_{ij} = M_{ij}, \quad \forall (i,j) \in \Omega \end{aligned} \tag{4.12}$$

其中 $X \in S^n$, $n = p + q$,$W_1 \in S^p$, $W_2 \in S^q$, $X, W_1, W_2 \succ \delta$.

我们主要讨论一般的情况, 即问题(4.12). 采用RBR法, 对于某个 i , 我们把向量 y 分为两个部分, 即

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad y_1 = X_{\alpha_i, i}, \quad y_2 = X_{\beta_i, i} \tag{4.13}$$

其中 $\alpha_i = \begin{cases} \{j+p|(i,j\in\Omega)\}, i\le p \\ \{j|(j,i-p)\in\Omega\}, p<i\le n \end{cases}$, $\beta_i = \{1,\cdots,p\}\backslash(\alpha_i\cup\{i\})$, y_1 是 X 第 i 列除去第 i 行后所有已知元素构成的列向量,y_2 是 X 第 i 列除去第 i 行后所有未知元素构成的列向量.

相应的, $B = \begin{pmatrix} X_{\alpha_i,\alpha_i} & X_{\alpha_i,\beta_i} \\ X_{\beta_i,\alpha_i} & X_{\beta_i,\beta_i} \end{pmatrix}$, $\xi = X_{i,i}$. 同时,可以给出 $\bar{A}\begin{pmatrix}\xi \\ y\end{pmatrix}$,$\bar{b}$ 和 $\bar{c}$ 的显式表达

$$\bar{b} = \begin{cases} (M_{i,\alpha_i-p})^\top, & i\le p \\ M_{\alpha_i,i-p}\ , & p<i\le n \end{cases}, \quad \bar{A}\begin{pmatrix}\xi \\ y\end{pmatrix} = y_1, \quad \bar{c} = (1,\overbrace{0,\cdots,0}^{n-1}) \tag{4.14}$$

故(4.10)化为以下形式

$$\begin{aligned} \min \quad & \xi + \frac{1}{2\mu}\|y_1-\bar{b}\|_2^2 \\ \text{s.t.} \quad & \xi - y^\top B^{-1}y \ge \delta \end{aligned} \tag{4.15}$$

定理 4.1. 问题*(4.15)*的最优解为

$$\begin{aligned} y_1 =& (2\mu I + X_{\alpha,\alpha})^{-1}X_{\alpha,\alpha} \\ y_2 =& \frac{1}{2\mu}X_{\beta,\alpha}(\bar{b}-y_1) \\ \xi =& \frac{1}{2\mu}y_1^\top(\bar{b}-y_1)+\delta \end{aligned} \tag{4.16}$$

证明. [11] 容易发现, 当 $\xi = y^\top B^{-1}y+\delta$ 时, 才可能取到最小值, 则(4.15)等价于

$$\min_y y^\top B^{-1}y + \frac{1}{2\mu}\|y_1-\bar{b}\|_2^2 \tag{4.17}$$

令其梯度为0

$$\begin{pmatrix} X_{\alpha,\alpha} & X_{\alpha,\beta} \\ X_{\beta,\alpha} & X_{\beta,\beta} \end{pmatrix}^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \frac{1}{2\mu}\begin{pmatrix} y_1-\bar{b} \\ 0 \end{pmatrix} = 0 \tag{4.18}$$

变形可得

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \frac{1}{2\mu}\begin{pmatrix} X_{\alpha,\alpha} \\ X_{\beta,\alpha} \end{pmatrix}(y_1-\bar{b}) = 0 \tag{4.19}$$

则

$$\begin{aligned} y_1 =& (2\mu I + X_{\alpha,\alpha})^{-1}X_{\alpha,\alpha} \\ y_2 =& \frac{1}{2\mu}X_{\beta,\alpha}(y_1-\bar{b}) \end{aligned} \tag{4.20}$$

那么由(4.18)和(4.20), 有

$$\xi = y^\top B^{-1}y+\delta = -\frac{1}{2\mu}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}^\top \begin{pmatrix} y_1-\bar{b} \\ 0 \end{pmatrix} + \delta = \frac{1}{2\mu}y_1^\top(\bar{b}-y_1)+\delta \tag{4.21}$$

即 (4.15)的最优解为(4.16) □

我们可以让 i 循环取遍 $\{1, \cdots, n\}$, 逐行求解(4.15), 最终解决问题(1.4). 给出算法

算法 2.

步1 给出 $\delta \geq 0$, $X^1 \succ 0$, $F^0 = tr(X^1)$, $F^1 = +\infty$, $\epsilon > 0$, 设置计数器 $k = 1$, $i = 1$;

步2 若 $\frac{|F^{k-1} - F^k|}{\max\{1, |F^{k-1}|\}} \leq \epsilon$, 则停止;

步3 若 $i > n$, 则令 $i = 1$, $X^{k+1} = X^k$, $k = k + 1$, $F^k = tr(X^k)$, 转步2;

步4 求出 i 对应的 α_i , β_i;

步5 若 $|\alpha_i| = 0$, 令 $X^k_{\alpha,\alpha} = X^k_{\alpha,\beta} = X^k_{\beta,\alpha} = 0$, $X^k_{\beta,\beta} = X^k$, 否则按通常定义求出以上几个量.

步6 按(4.16), 计算当前最优解 ξ , y_1 , y_2 , 以及 $y = [y_1, y_2]$;

步7 令 $X^k_{i,i} = \xi$, $X^k_{i,i^c} = y$, $X^k_{i^c,i} = y^\top$;

步8 $i = i + 1$, 转步3.

注 2. 本章除了已经标注引用的定理(4.1), 均为原文[3]的内容.

5 FPCA法

5.1 FPC法

FPC法是一种基于不动点定理的算法, 可以解决问题(1.5).

$\|X\|_* + \frac{1}{2\mu}\|\mathcal{A}(X) - b\|_2^2$ 是一个凸函数, 求最优解只要求梯度为0的点. 但 $\|X\|_*$ 是不可微的, 故考虑次梯度, 即

$$0 \in \mu\partial\|X\|_* + g(X) \tag{5.1}$$

其中 $g(X) = \mathcal{A}^*(\mathcal{A}(X) - b)$

设 $Y = X - \psi g(X)$, $\psi > 0$ 是给定常数, 则(5.1)等价于

$$0 \in \psi\mu\partial\|X\|_* + X - (X - \psi g(X)) = \psi\mu\partial\|X\|_* + X - Y \tag{5.2}$$

(5.2)恰好是 $\psi\mu\|X\|_* + \frac{1}{2}\|X - Y\|_F^2$ 的次梯度, 即(1.4)等价于

$$\min \psi\mu\|X\|_* + \frac{1}{2}\|X - Y\|_F^2 \tag{5.3}$$

引入矩阵的Shrinkage算子[5], 定义如下

定义 5.1 (Shrinkage算子)**.** *对 $X \in \mathbb{R}^{m\times n}$ 做奇异值分解, $X = UDiag(\sigma)V^\top$, $r = rank(X)$, $U \in \mathbb{R}^{m\times r}$, $V \in \mathbb{R}^{n\times r}$, 对 $\forall\nu > 0$, 定义向量 $\bar{\sigma} = (\bar{\sigma_1}, \cdots, \bar{\sigma_r})$, 其中$\bar{\sigma}_i = \max\{\sigma_i - \nu, 0\}$. Shrinkage算子 $S_\nu : \mathcal{R}^{m\times n} \to \mathcal{R}^{m\times n}$, $S_\nu(X) = UDiag(\bar{\sigma})V^\top$*

当认为(5.3)中的 Y 是给定的情况下, 有如下定理

定理 5.1. *问题(5.3)中, 当给定 $Y \in \mathcal{R}^{m\times n}$, 最优解为 $S_{\psi\mu}(Y)$, 其中 $S_{\psi\mu}(\cdot)$ 是Shrinkage算子*

证明. [5] 不失一般性, 设 $m \le n$, 记 $\nu = \psi\mu$, $X = U\mathrm{Diag}(\sigma)V^\top$, $Y = U_Y\mathrm{Diag}(\gamma)V_Y^\top$ 是 X , Y 的SVD, $\mathrm{rank}X = r$, $\mathrm{rank}Y = t$.

因为 $\partial\|X\|_* = \{UV^\top + W|U^\top W = 0, WV = 0, \|W\|_2 \le 1\}$, 找到矩阵 $\bar{U} \in \mathbb{R}^{m\times(m-r)}$, $\bar{V} \in \mathbb{R}^{n\times(n-r)}$, 使得 $\tilde{U} = [U, \bar{U}]$, $\tilde{V} = [V, \bar{V}]$ 分别是 m , n 阶正交矩阵. 那么易验证, 当 $\bar{\sigma} \in \mathbb{R}_+^{m-r}$, $\|\sigma\|_\infty \le 1$, 形如 $W = \bar{U}[\mathrm{Diag}(\bar{\sigma}), 0]\bar{V}^\top$ 的矩阵 $W \in \partial\|X\|_*$.

若 $0 \in \nu\partial\|X\|_* + X - Y$, 则说明 $\exists W \in \partial\|X\|_*$ 使 $\nu(UV^\top + W) + X - Y = 0$. 那么可以看到当 $W = \bar{U}[\mathrm{Diag}(\bar{\sigma}), 0]\bar{V}^\top$, 则要求

$$
\begin{aligned}
&\nu(UV^\top + W) + X - Y \\
=&(\nu UIV^\top + \bar{U}[\mathrm{Diag}(\bar{\sigma}), 0]\bar{V}^\top + U\mathrm{Diag}(\sigma)V^\top) - Y \\
=&\begin{pmatrix} U \\ \bar{U} \end{pmatrix} \left[\begin{pmatrix} \nu I & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \nu\mathrm{Diag}(\bar{\sigma}) & 0 \end{pmatrix} + \begin{pmatrix} \mathrm{Diag}(\sigma) & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right] \begin{pmatrix} V \\ \bar{V} \end{pmatrix}^\top - U_Y\mathrm{Diag}(\gamma)V_Y^\top \\
=&\begin{pmatrix} U \\ \bar{U} \end{pmatrix} \begin{pmatrix} \nu I + \mathrm{Diag}(\sigma) & 0 & 0 \\ 0 & \nu\mathrm{Diag}(\bar{\sigma}) & 0 \end{pmatrix} \begin{pmatrix} V \\ \bar{V} \end{pmatrix}^\top - U_Y\mathrm{Diag}(\gamma)V_Y^\top = 0
\end{aligned}
\tag{5.4}
$$

当 $\gamma_1 \ge \cdots \ge \gamma_t \ge \nu$, 那我们令 $X = S_\nu(Y)$, 则 $r = t$, $U = U_Y$, $V = V_Y$, $\sigma = (\gamma_1 - \nu, \cdots, \gamma_t - \nu)$, 对于这样的 X , 取 $\bar{\sigma} = 0$ 即 $W = 0$, (5.4)成立.

当 $\gamma_1 \ge \cdots \ge \gamma_k \ge \nu \ge \gamma_{k+1} \ge \cdots \ge \gamma_t$, 那我们令 $X = S_\nu(Y)$, 则 $r = k$, $\sigma = (\gamma_1 - \nu, \cdots, \gamma_k - \nu, 0, \cdots, 0)$, 对这样的 X , 取 $\bar{\sigma} = (\gamma_{k+1}/\nu, \cdots, \gamma_t/\nu, 0, \cdots, 0)$, 同时令 $\bar{U}$ 的前 $t-k$ 行依次为 U_Y 的 $k+1$ 至 t 行, $\bar{V}$ 的前 $t-k$ 行依次为 V_Y 的 $k+1$ 至 t 行, 即 $\begin{pmatrix} U \\ \bar{U} \end{pmatrix} = \begin{pmatrix} U_Y \\ \bar{U}' \end{pmatrix}$, $\begin{pmatrix} V \\ \bar{V} \end{pmatrix} = \begin{pmatrix} V_Y \\ \bar{V}' \end{pmatrix}$, 其中 $\bar{U}'$ 是 $\bar{U}$ 的后 $m-t$ 行, $\bar{V}'$ 是 $\bar{V}$ 的后 $n-t$ 行, 此时(5.4)也成立.

故 $S_{\psi\mu}(Y)$ 是(5.3)的解. □

5.2 FPC法的收敛性

也就是说, 我们只需要求出合适的 Y , $S_{\psi\mu}(Y)$ 就是(1.5)的解. 由 X 与 Y 的关系可以看出, 若 X^* 是(1.5)的解, 则

$$X^* = S_{\psi\mu}(X^* - \mu g(X^*)) = S_{\psi\mu}(X^* - \mu\mathcal{A}^*(\mathcal{A}(X^*) - b)) \tag{5.5}$$

即 X^* 是映射 $S_{\psi\mu} \circ h$ 的不动点, 其中 $h(X) = X - \mu g(X)$.

我们有如下的结论

定理 5.2. *Shrinkage*算子 $S_\nu(\cdot)$ 是非扩张的, 即

$$\|S_\nu(Y_1) - S_\nu(Y_2)\|_F \le \|Y_1 - Y_2\|_F \tag{5.6}$$

证明. [5] 设 $Y_1 = U_1 D_1 V_1^\top$, $Y_2 = U_2 D_2 V_2^\top$, 则 $\bar{Y}_1 = S_\nu(Y_1) = U_1 \bar{D}_1 V_1^\top$, $\bar{Y}_2 = S_\nu(Y_2) = U_2 \bar{D}_2 V_2^\top$. 其中 $D_i = \mathrm{Diag}(\sigma_i)$, $\bar{D}_i = \mathrm{Diag}(\max\{\sigma_i - \nu, 0\})$, $\sigma_i \in \mathbb{R}^{r_i}$ 是 Y_i 的奇异值向量, $r_i = \mathrm{rank}\ Y_i$, 记 σ_i 中大于 ν 的项数为 k_i , 不失一般性, 我们认为奇异值是从大到小排列的, 即 $\sigma_i^{(1)} \ge \cdots \ge \sigma_i^{(k_i)} \ge \nu \ge \sigma_i^{(k_i+1)} \ge \cdots \ge \sigma_i^{(r_i)} \ge 0$, $i = 1, 2$.

$$\begin{aligned}
&\|Y_1 - Y_2\|_F - \|\bar{Y}_1 - \bar{Y}_2\|_F \\
=&tr((Y_1 - Y_2)^\top (Y_1 - Y_2)) - tr((\bar{Y}_1 - \bar{Y}_2)^\top (\bar{Y}_1 - \bar{Y}_2)) \\
=&tr(Y_1^\top Y_1 + Y_2^\top Y_2 - \bar{Y}_1^\top \bar{Y}_1 - \bar{Y}_2^\top \bar{Y}_2) + 2tr(Y_1^\top Y_2 - \bar{Y}_1^\top \bar{Y}_2) \\
=&\sum_{i=1}^{r_1} (\sigma_1^{(i)})^2 + \sum_{i=1}^{r_2} (\sigma_2^{(i)})^2 - \sum_{i=1}^{k_1} (\sigma_1^{(i)} - \nu)^2 - \sum_{i=1}^{k_2} (\sigma_2^{(i)} - \nu)^2 - 2tr(Y_1^\top Y_2 - \bar{Y}_1^\top \bar{Y}_2)
\end{aligned}$$

记 $U = U_1^\top U_2$, $V = V_1^\top V_2$, 将各项展开可得

$$tr(Y_1^\top Y_2 - \bar{Y}_1^\top \bar{Y}_2) = tr((D_1 - \bar{D}_1)^\top U (D_2 - \bar{D}_2) V^\top + (D_1 - \bar{D}_1)^\top U \bar{D}_2 V^\top + \bar{D}_1^\top U (D_2 - \bar{D}_2) V^\top) \tag{5.7}$$

有如下的两个定理成立:

- U 是正交矩阵, $tr(AU)$ 取到最大值, 当且仅当 AU 是正定矩阵.[4]
- 当 AB 正定, $tr(AB) \le \sum_i \sigma^{(i)}(A)\sigma^{(i)}(B)$, $\sigma^{(i)}$ 是矩阵的第 i 个奇异值.[4]

不失一般性, 我们设 $r_1 > r_2 > k_1 > k_2$. 对于(5.7)中等式右边的3项, 由上面的定理, 有

$$\begin{aligned}
tr((D_1-\bar{D}_1)^\top U(D_2-\bar{D}_2)V) &\le \sum_i \sigma^{(i)}(D_1-\bar{D}_1)\sigma^{(i)}(D_2-\bar{D}_2)\\
&\le \sum_{i=1}^{k_2}\nu^2+\sum_{i=k_2+1}^{k_1}\sigma_2^{(i)}\nu+\sum_{i=k_1+1}^{r_2}\sigma_1^{(i)}\sigma_2^{(i)}\\
&\le \sum_{i=1}^{k_1}\sigma_2^{(i)}\nu+\sum_{i=k_1+1}^{r_2}\sigma_1^{(i)}\sigma_2^{(i)}\\
tr((D_1-\bar{D}_1)^\top U\bar{D}_2V^\top) &\le \sum_i \sigma^{(i)}(D_1-\bar{D}_1)\sigma^{(i)}(\bar{D}_2)\\
&\le \sum_{i=1}^{k_1}(\sigma_2^{(i)}-\nu)\nu+\sum_{i=k_1+1}^{r_2}(\sigma_2^{(i)}-\nu)\sigma_1^{(i)}\\
&\le \sum_{i=1}^{r_2}(\sigma_2^{(i)}-\nu)\sigma_1^{(i)}\\
tr(\bar{D}_1^\top U(D_2-\bar{D}_2)V^\top) &\le \sum_i \sigma^{(i)}(\bar{D}_1)\sigma^{(i)}(D_2-\bar{D}_2)\\
&\le \sum_{i=1}^{k_2}(\sigma_1^{(i)}-\nu)\nu+\sum_{i=k_2+1}^{r_2}(\sigma_1^{(i)}-\nu)\sigma_2^{(i)}\\
&\le \sum_{i=1}^{r_2}(\sigma_1^{(i)}-\nu)\sigma_2^{(i)}
\end{aligned} \tag{5.8}$$

则可得

$$\begin{aligned}
&\|Y_1-Y_2\|_F-\|\bar{Y}_1-\bar{Y}_2\|_F\\
=&\sum_{i=1}^{r_1}(\sigma_1^{(i)})^2+\sum_{i=1}^{r_2}(\sigma_2^{(i)})^2-\sum_{i=1}^{k_1}(\sigma_1^{(i)}-\nu)^2-\sum_{i=1}^{k_2}(\sigma_2^{(i)}-\nu)^2-2tr(Y_1^\top Y_2-\bar{Y}_1^\top\bar{Y}_2)\\
\ge&\sum_{i=1}^{r_1}(\sigma_1^{(i)})^2+\sum_{i=1}^{r_2}(\sigma_2^{(i)})^2-\sum_{i=1}^{k_1}(\sigma_1^{(i)}-\nu)^2-\sum_{i=1}^{k_2}(\sigma_2^{(i)}-\nu)^2\\
&\quad-2\Big(\sum_{i=1}^{k_1}\sigma_2^{(i)}\nu+\sum_{i=k_1+1}^{r_2}\sigma_1^{(i)}\sigma_2^{(i)}+\sum_{i=1}^{r_2}(\sigma_2^{(i)}-\nu)\sigma_1^{(i)}+\sum_{i=1}^{r_2}(\sigma_1^{(i)}-\nu)\sigma_2^{(i)}\Big)\\
=&\sum_{i=k_1+1}^{k_2}\Big(2\sigma_2^{(i)}\nu-\nu^2+(\sigma_1^{(i)})^2-2\sigma_1^{(i)}\sigma_2^{(i)}\Big)\\
&\quad+\Big(\sum_{i=k_1+1}^{r_2}(\sigma_1^{(i)})^2+\sum_{i=k_1+1}^{r_1}(\sigma_2^{(i)})^2-2\sum_{i=k_1+1}^{r_1}\sigma_1^{(i)}\sigma_2^{(i)}\Big)\\
\ge&\sum_{i=k_1+1}^{k_2}\Big(2\sigma_2^{(i)}\nu-\nu^2+(\sigma_1^{(i)})^2-2\sigma_1^{(i)}\sigma_2^{(i)}\Big)+\sum_{i=k_1+1}^{r_2}(\sigma_1^{(i)}-\sigma_2^{(i)})^2\\
\ge&\sum_{i=k_1+1}^{k_2}\Big(2\sigma_2^{(i)}\nu-\nu^2+(\sigma_1^{(i)})^2-2\sigma_1^{(i)}\sigma_2^{(i)}\Big)
\end{aligned} \tag{5.9}$$

当 $k_1 < i \le k_2$, $\sigma_2^{(i)} < \nu \le \sigma_1^{(i)}$, 则

$$2\sigma_2^{(i)}\nu-\nu^2+(\sigma_1^{(i)})^2-2\sigma_1^{(i)}\sigma_2^{(i)}=(\sigma_1^{(i)}-\sigma_2^{(i)})^2-(\nu-\sigma_2^{(i)})^2\ge 0$$

那么有 $\|Y_1 - Y_2\|_F - \|\bar{Y}_1 - \bar{Y}_2\|_F \ge \sum_{i=k_1+1}^{k_2} \left(2\sigma_2^{(i)}\nu - \nu^2 + (\sigma_1^{(i)})^2 - 2\sigma_1^{(i)}\sigma_2^{(i)}\right) \ge 0$, 即 $\|Y_1 - Y_2\|_F \ge \|\bar{Y}_1 - \bar{Y}_2\|_F$. □

定理 5.3. 当 $\mu \in (0, 2/\lambda_{max}(A^\top A))$, 其中 A 满足 $\mathcal{A}(X) = Avec(X)$, h 是非扩张的, 即

$$\|h(X_1) - h(X_2)\|_F \le \|X_1 - X_2\|_F \tag{5.10}$$

证明. [5] 取矩阵 A , 使得 $\mathcal{A}(X) = Avec(X)$, 则 $h(X) = X - \mu A^\top(Avec(X) - b)$,则

$$\|h(X_1 - X_2)\|_F = \|(I - \mu A^\top A)(vec(X_1) - vec(X_2)\|_2 \le \|I - \mu A^\top A\|_2 \|vec(X_1) - vec(X_2\|_2$$

由 $\mu < 2/\lambda_{max}(A^\top A)$, $\|I - \mu A^\top A\|_2 < 1$, 则

$$\|h(X_1 - X_2)\|_F \le \|vec(X_1) - vec(X_2\|_2 \le \|X_1 - X_2\|_F$$

这证明了 h 是非扩张的映射. □

那么很自然的推论是, 当 $\mu \in (0, 2/\lambda_{max}(A^\top A))$

$$\|S_{\psi\mu}(h(X_1)) - S_{\psi\mu}(h(X_2))\|_F \le \|h(X_1) - h(X_2)\|_F \le \|X_1 - X_2\|_F \tag{5.11}$$

即 $S_{\psi\mu} \circ h$ 是一个非扩张的映射, 由Brouwer不动点定理可知, 由任意 X 开始进行迭代, 都能收敛到某个不动点.

5.3 奇异值计算的近似算法FPCA

实际计算中, 上述迭代过程中, 每步都要求做奇异值分解, 这并不是一件容易的事. 我们认为问题(1.5)中的原矩阵 X 是一个低秩矩阵, 那么其奇异值大多都是 0 , 故可以考虑只求很少的几个奇异值.

$A \in \mathbb{R}^{p\times q}$, 取整数 c , d , $1 \le d \le c \le q$, $(P_1, \cdots, P_q)$, $P_i \ge 0$, $\sum_{i=1}^q P_i = 1$. 构造随机向量 $(i_1, \cdots, i_c)$, $P(i_t = j) = P_j$, $t \in \{1, \cdots, c\}$, $j \in \{1, \cdots, q\}$. 再令随机矩阵

$$C = \begin{pmatrix} C^{(1)} \\ \cdots \\ C^{(c)} \end{pmatrix} = \begin{pmatrix} A^{(i_i)}/\sqrt{cP_{i_1}} \\ \cdots \\ A^{(i_c)}/\sqrt{cP_{i_c}} \end{pmatrix} \in \mathbb{R}^{p\times c} \tag{5.12}$$

求 $C^\top C$ 的特征值分解

$$C^\top C = \sum_{i=1}^{c} \sigma_i^2(C) y_i y_i^\top \tag{5.13}$$

其中 $\sigma_i(C) \ge 0$, 为 C 的奇异值 , 再构造矩阵

$$H = \begin{pmatrix} H^{(1)} \\ \cdots \\ H^{(d)} \end{pmatrix} = \begin{pmatrix} Cy_1/\sigma_1(C) \\ \cdots \\ Cy_d/\sigma_d(C) \end{pmatrix} \in \mathbb{R}^{p\times d} \tag{5.14}$$

令

$$A_d = H\mathrm{Diag}(\sigma(C))\big(A^\top H\mathrm{Diag}(1/\sigma(C))\big)^\top \tag{5.15}$$

可以证明 A_d 是 A 的一个比较好的近似 [2]

$$\|A - A_d\|_\xi^2 \le \min_{\mathrm{rank}(D)\le d} \|A - D\|_\xi^2 + ploylog(d, 1/c)\|A\|_\xi^2 \tag{5.16}$$

其中 $\xi = 2$ 或 F , $polylog$ 是多重对数函数.

用这种近似的方法改进的FPC算法就是FPCA法, 给出算法

算法 3.
步1 给出 X^1 , $\mu^0 > 0$, $1 > \theta > 0$, $\epsilon > 0$, 设置计数器 $k = 1$;
步2 若 $\mu^k = \mu^{k-1}\theta \le \epsilon$, 则停止;
步3 选取合适的 $\psi > 0$;
步4 计算 $Y^k = X^k - \psi\mathcal{A}^*(\mathcal{A}(X^k) - b)$;
步5 选取合适的 d , $\{P_i\}$, 按(5.12)至(5.15), 计算近似的奇异值分解 $Y_d = HDiag(\sigma(C))\big(Y^\top HDiag(1/\sigma(C))\big)^\top$
步6 计算 $X^{k+1} = S_{\psi\mu^k}(Y_d)$;
步7 $k = k + 1$, 转步2.

注 3. 本章除表明引用处, 均为原文 [3] 内容

6 结语

在我第一次听说NetFlix大奖赛后, 就对矩阵优化产生了兴趣, 因而选择了这篇文章. 通过本文的阅读, 我比较初步地了解了矩阵完整化问题是什么以及几种常见的算法, 熟悉了矩阵范数, 罚函数, 次梯度等基础知识.

这是第一次精读一篇文献, 也是第一次比较正式地准备报告, 花了比较多时间, 但是在阅读过程中明显感觉自己的提升, 阅读其他参考文献时阻力越来越小. 我也通过这次掌握了许多工具的使用, 比如latex, beamer, Google学术搜索等, 不过同样也因为放了太多精力在latex排版学习等地方, 对文献内容的深入思考不足. 另外最后的报告, 做的也不是那么让自己满意.

本学期的研讨课的学习让我学习到了不少, 也还有许多做得不好的地方, 不过总之还是收获颇丰. 最后感谢袁老师和其他各位老师的指导, 以及各位同学的帮助, 谢谢!

参考文献

[1] Emmanuel J Candès and Terence Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080, 2010.

[2] Petros Drineas, Ravi Kannan, and Michael W Mahoney. Fast monte carlo algorithms for matrices ii: Computing a low-rank approximation to a matrix. *SIAM Journal on Computing*, 36(1):158–183, 2006.

[3] Donald Goldfarb, Shiqian Ma, and Zaiwen Wen. Solving low-rank matrix completion problems efficiently. In *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, pages 1013–1020. IEEE, 2009.

[4] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[5] Shiqian Ma, Donald Goldfarb, and Lifeng Chen. Fixed point and bregman iterative methods for matrix rank minimization. *Mathematical Programming*, 128(1-2):321–353, 2011.

[6] Carl D Meyer. *Matrix analysis and applied linear algebra*, volume 2. Siam, 2000.

[7] Balas Kausik Natarajan. Sparse approximate solutions to linear systems. *SIAM journal on computing*, 24(2):227–234, 1995.

[8] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.

[9] Robert Schatten. *Norm ideals of completely continuous operators*, volume 27. Springer-Verlag, 2013.

[10] G Alistair Watson. Characterization of the subdifferential of some matrix norms. *Linear algebra and its applications*, 170:33–45, 1992.

[11] Zaiwen Wen, Donald Goldfarb, Shiqian Ma, and Katya Scheinberg. Row by row methods for semidefinite programming. *Industrial Engineering*, pages 1–21, 2009.

[12] 袁亚湘. *非线性优化计算方法*. 科学出版社, 2008.