低秩矩阵完整化问题的几种高效解法

林陈冉

中国科学院大学

2016年12月19日

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. 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.

数学模型

通常我们认为矩阵的秩比较小. 也就是这样的一个最优化问题:

$$\begin{aligned} \min \quad & rank(X) \\ \text{s.t.} \quad & X_{ij} = M_{ij}, \forall (i,j) \in \Omega \end{aligned} \tag{2.1}$$

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, $|\Omega| = m$.

在一些情况下, (2.1)等价于线性约束问题:

$$\begin{aligned} \min \quad & rank(X) \\ \text{s.t.} \quad & \mathcal{A}(X) = b \end{aligned} \tag{2.2}$$

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

数学模型

但是(2.1), (2.2)都是"**NP-难**"的, 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(2.1)与(2.2)转化为以下形式:

$$\begin{aligned} \min \quad & \|X\|_* \\ \text{s.t.} \quad & \mathcal{A}(X) = b \end{aligned} \tag{2.3}$$

$$\begin{aligned} \min \quad & \|X\|_* \\ \text{s.t.} \quad & X_{ij} = M_{ij}, \forall (i,j) \in \Omega \end{aligned} \tag{2.4}$$

数学模型

很自然的，一个重要的问题是：(2.1)与(2.3)，或者(2.2)与(2.4)什么时候等价？略过证明，直接描述以下重要的结论：

- 对于(2.3)，当在某些正则条件下，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恢复.
- 对于(2.4)，将线性映射 $\mathcal{A}$ 的矩阵形式记作 A，当 A 是一个随机高斯矩阵，若向量 b 的维数 $m \geq C(r(p+q)log(pq)$，其中 C 是一个正的常数，则矩阵有很高概率可以恢复.

数学模型

实际上，由线性代数知识容易知道，(2.4)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(2.4)的罚函数：

$$\min \quad \|X\|_* + \frac{1}{2\mu}\|\mathcal{A}(X) - b\|_2^2 \tag{2.5}$$

其中 μ 是某个给定常数.

下面，将对(2.3)，(2.4)，(2.5)分别给出一种高效的解法.

问题转化

对于问题(2.3), 我们考虑它的对偶问题

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

引入一个形式上的变量 S , 将(3.1)变为如下的等价形式

$$\begin{aligned} \min_{y\in\mathbb{R}^m} \quad & -b^\top y \\ \text{s.t.} \quad & \mathcal{A}^*(y) - S = 0 \\ & \|S\|_2 \leq 1 \end{aligned} \tag{3.2}$$

问题转化

可以考虑(3.2)的增广Lagrange函数

$$L(y,S,X,\mu) = -b^\top y + \langle X, \mathcal{A}^*(y) - S\rangle + \frac{1}{2\mu}\|\mathcal{A}^*(y) - S\|_F^2 \quad (3.3)$$

由此我们得到(3.1)的一个等价问题

$$\begin{aligned} \min_{y,S} \quad & L(y,S,X,\mu) \\ \text{s.t.} \quad & \|S\|_2 \le 1 \end{aligned} \quad (3.4)$$

通过选取最佳的 μ 和 X 可以求出(3.4)的最优解.

问题转化

对下面的几个量进行迭代, 可以求解问题(3.4)

$$\begin{aligned} \mu^{k+1} &= \alpha\mu^k, \quad \alpha \in (0,1) \\ X^{k+1} &= X^k + \frac{\mathcal{A}^*(y^{k+1}) - S^{k+1}}{\mu^k} \\ (y^{k+1}, S^{k+1}) &= \arg\min_{y,S} L(y, S, X^k, \mu^k) \end{aligned} \tag{3.5}$$

交替方向法

但实际上, (3.5)第三式并不容易解, 因此我们考虑使用交替方向法来求解这个方程, 即固定 S 求 y , 再固定 y 求 S .

$$y^{k+1} = \arg\min_{y,S} L(y, S^k, X^k, \mu^k) \tag{3.6}$$

$$S^{k+1} = \arg\min_{y,S} L(y^{k+1}, S, X^k, \mu^k) \tag{3.7}$$

交替方向法

定理

当固定 S^k , (3.6)的最优解为:

$$y^{k+1} = \mu^k(b - \mathcal{A}(X^k)) + \mathcal{A}(S^k) \tag{3.8}$$

当固定 y^{k+1} , (3.7)的最优解为:

$$S^{k+1} = U\,\mathit{Diag}(\min\{\sigma, 1\})V^\top \tag{3.9}$$

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

AAL法算法

算法

步1　给出 μ^0 , X^0 , y^0 , S^0 , ϵ , α , 设置计数器 $k=0$;

步2　若 $\frac{\|X^{k+1}-X^k\|_F}{\max\{1,\|X^k\|_F\}} \le \epsilon$, 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:

$Y = U Diag(\sigma) V^\top$;

步5　计算 $S^{k+1} = U Diag(min\{\sigma, 1\}) V^\top$;

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.

SDP问题

对于问题(2.4), 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.

Schur补

定义

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

$$X = \begin{pmatrix} \xi & y^\top \\ y & B \end{pmatrix} = \begin{pmatrix} 1 & y^\top B^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} \xi - y^\top B^{-1} y & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} 1 & 0 \\ B^{-1} y & I \end{pmatrix} \tag{4.2}$$

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

$$X \succeq 0 \Leftrightarrow B \succeq 0, (X/B) \geq 0 \tag{4.3}$$

SOCP问题

约定以下记号

$$X_{\alpha,\beta} = \begin{cases} x_{\alpha\beta} & \alpha, \beta \in \mathbb{R} \\ (x_{\alpha\beta_1}, \cdots, x_{\alpha\beta_n}) & \alpha \in \mathbb{R}, \beta = \{\beta_1, \cdots, \beta_n\} \\ (x_{\alpha_1\beta}, \cdots, x_{\alpha_m\beta})^\top & \alpha = \{\alpha_1, \cdots, \alpha_m\}, \beta \in \mathbb{R} \\ \begin{pmatrix} x_{\alpha_1\beta_1} & \cdots & x_{\alpha_1\beta_n} \\ \cdots & \cdots & \cdots \\ x_{\alpha_m\beta_1} & \cdots & x_{\alpha_m\beta_n} \end{pmatrix} & \alpha = \{\alpha_1, \cdots, \alpha_m\}, \beta = \{\beta_1, \cdots, \beta_n\} \end{cases} \tag{4.4}$$

$$i^c = \{1, \cdots, n\} \backslash \{i\} = \{1, \cdots, i-1, i+1, \cdots, n\} \tag{4.5}$$

SOCP问题

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Schur补, 令 i 取遍 $\{1,\cdots,n\}$, 逐行解如下的SOCP问题来解决SDP问题(4.1)

$$\begin{aligned} \min_{[\xi;y]\in\mathbb{R}^n} \quad & \bar{c}^\top \begin{pmatrix} \xi \\ y \end{pmatrix} \\ \text{s.t.} \quad & \bar{X} \begin{pmatrix} \xi \\ y \end{pmatrix} = \bar{b} \\ & (X/B) \geq \delta \end{aligned} \tag{4.6}$$

SOCP问题

其中

$$\bar{c} = \begin{pmatrix} C_{i,i} \\ 2Ci^c, i \end{pmatrix}, \bar{X} = \begin{pmatrix} X^(1)_{i,i} & 2X^(1)_{i,i^c} \\ \cdots & \cdots \\ X^(m)_{i,i} & X^(m)_{i,i^c} \end{pmatrix}, \bar{b} = \begin{pmatrix} b_1 - \langle X^(1)_{i^c,i^c}, B \rangle \\ \cdots \\ b_m - \langle X^(m)_{i^c,i^c}, B \rangle \end{pmatrix} \tag{4.7}$$

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$\delta > 0$.

Powell罚函数

考虑(3.7)的Powell罚函数

$$F(X,\theta,\mu)=\bar{c}^{\top}\begin{pmatrix}\xi\\ y\end{pmatrix}+\frac{1}{2\mu}\|\bar{X}[\xi;y]-\bar{b}-\theta\|_2^2 \tag{4.8}$$

其中 $\theta\in\mathbb{R}^m$, $\mu>0$ 都是是给定的. 记 $\lambda=\theta+\bar{b}$

(4.6)等价于以下问题

$$\begin{aligned}\min_{X}\quad & F(X,\lambda,\mu)=\bar{c}^{\top}\begin{pmatrix}\xi\\ y\end{pmatrix}+\frac{1}{2\mu}\|\bar{X}[\xi;y]-\lambda\|_2^2\\ \text{s.t.}\quad & X\succeq 0\end{aligned} \tag{4.9}$$

用SDP和RBR法补全矩阵

回头考虑(2.4), 我们需要把它转化为一个SDP问题(4.1), 考虑两种情况:

- 当 $X \in S^n$, $\|X\|_* = tr(X)$, (2.4)等价于下面的SDP问题

$$\begin{aligned} \min_{X} \quad & tr(X) = \langle E, X \rangle \\ \text{s.t.} \quad & X_{ij} = M_{ij}, \quad \forall (i,j) \in \Omega \end{aligned} \tag{4.10}$$

用SDP和RBR法补全矩阵

- 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

$$
\begin{aligned}
\min_{X} \quad & tr(X) \\
\text{s.t.} \quad & X = \begin{pmatrix} X_1 & W \\ W^\top & X_2 \end{pmatrix} \succ 0 \\
& W_{ij} = M_{ij}, \quad \forall (i,j) \in \Omega
\end{aligned} \tag{4.11}
$$

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

用SDP和RBR法补全矩阵

我们主要讨论一般的情况, 即问题(4.11). 对于某个 i , 我们把向量 y 分为两个部分, 即

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad y_1 = X_{\alpha_i, i}, \quad y_2 = X_{\beta_i, i} \tag{4.12}$$

其中 $\alpha_i = \begin{cases} \{j + p | (i, j \in \Omega)\}, i \leq p \\ \{j | (j, i - p) \in \Omega\}, p < i \leq n \end{cases}$, $\beta_i = \{1, \cdots, p\} \backslash (\alpha_i \cup \{i\})$, 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.

用SDP和RBR法补全矩阵

分解这个对称正定矩阵 X ，使其符合SOCP问题(4.6)的形式，令

$$B = \begin{pmatrix} X_{\alpha_i,\alpha_i} & X_{\alpha_i,\beta_i} \\ X_{\beta_i,\alpha_i} & X_{\beta_i,\beta_i} \end{pmatrix}, \quad \xi = X_{i,i} \ .$$

对照SOCP问题的罚函数形式(4.9)，可以给出 $\bar{X}\begin{pmatrix} \xi \\ y \end{pmatrix}$,λ 和 $\bar{c}$ 的显式表达

$$\lambda = \begin{cases} (M_{i,\alpha_i - p})^\top, & i \le p \\ M_{\alpha_i, i-p} \ , & p < i \le n \end{cases}, \quad \bar{X}\begin{pmatrix} \xi \\ y \end{pmatrix} = y_1, \quad \bar{c} = (1, \overbrace{0, \cdots, 0}^{n-1}) \tag{4.13}$$

用SDP和RBR法补全矩阵

故(4.9)化为以下的形式

$$\begin{aligned} \min \quad & \xi + \frac{1}{2\mu}\|y_1 - \lambda\|_2^2 \\ \text{s.t.} \quad & \xi - y^\top B^{-1} y \geq \delta \end{aligned} \tag{4.14}$$

现在只需要解这个问题的最优解, 即可得到矩阵 X 的第 i 行和第 i 列, 当 i 循环取遍 1 到 n , 就可以得到原问题(4.11)的最优解, 从而求得目标矩阵 W .

用SDP和RBR法补全矩阵

定理

问题*(4.14)*的最优解为

$$\begin{aligned} y_1 &= (2\mu I + X_{\alpha,\alpha})^{-1} X_{\alpha,\alpha} \tilde{b} \\ y_2 &= \frac{1}{2\mu} X_{\beta,\alpha}(\lambda - y_1) \\ \xi &= \frac{1}{2\mu} y_1^\top (\lambda - y_1) + \delta \end{aligned} \tag{4.15}$$

RBR法算法

算法

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 $X^k_{\alpha,\alpha} = x^l_{\alpha,\beta} = X^k_{\beta,\alpha} = 0$, $X^k_{\beta,\beta} = X^k$，否则按通常定义求出以上几个量.

步6　按(4.15), 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;

FPC法

FPC法是一种基于不动点定理的算法，可以解决问题(2.5).

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

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

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

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

FPC法

(5.2)恰好是 $\|X\|_* + \frac{1}{2}\|X-Y\|_F^2$ 的次梯度, 即(2.4)等价于

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

定义

对 $X \in \mathbb{R}^{m\times n}$ 做奇异值分解, $X = U\mathit{Diag}(\sigma)V^\top$, $r = \mathit{rank}(X)$, $U \in \mathbb{R}^{m\times r}$, $V \in \mathbb{R}^{n\times r}$, 对 $\forall\nu > 0$, 定义向量 $\bar{\sigma} = (\bar{\sigma}_1, \cdots, \bar{\sigma}_r)$, 其中$\bar{\sigma}_i = \max\{\sigma_i - \nu, 0\}$.

*Shinkage*算子 $S_\nu : \mathcal{R}^{m\times n} \to \mathcal{R}^{m\times n}$, $S_\nu(X) = U\mathit{Diag}(\bar{\sigma})V^\top$

FPC法

定理

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*Shinkage*算子

FPC法的收敛性

上面的定理告诉我们, 只需要求出合适的 Y , $S_{\psi\mu}(Y)$ 就是(2.5)的解.

由 X 与 Y 的关系可以看出, 若 X^* 是(2.5)的解, 则

$$X^* = S_{\psi\mu}(X^* - \mu g(X^*)) = S_{\psi\mu}(X^* - \mu \mathcal{A}^*(\mathcal{A}(X^*) - b)) \qquad (5.4)$$

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.

FPC法的收敛性

定理

*Shinkage*算子 $S_\nu(\cdot)$ 是非扩张的, 即

$$\|S_\nu(Y_1) - S_\nu(Y_2)\|_F \leq \|Y_1 - Y_2\|_F \tag{5.5}$$

定理

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

$$\|h(X_1) - h(X_2)\|_F \leq \|X_1 - X_2\|_F \tag{5.6}$$

FPC法的收敛性

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

$$\|S_{\psi\mu}(h(X_1)) - S_{\psi\mu}(h(X_2))\|_F \leq \|h(X_1) - h(X_2)\|_F \leq \|X_1 - X_2\|_F \tag{5.7}$$

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.

优化奇异值计算

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(2.5)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.

这种优化FPC法中奇异值计算的方法, 就是FPCA法.

优化奇异值计算

$A \in \mathbb{R}^{p\times q}$ ，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

$$C = \begin{pmatrix} C^{(1)} \\ \cdots \\ C^{(c)} \end{pmatrix} = \begin{pmatrix} A^{(i_i)}/\sqrt{cP_{i_1}} \\ \cdots \\ A^{(i_c)}/\sqrt{cP_{i_c}} \end{pmatrix} \in \mathbb{R}^{p\times c} \tag{5.8}$$

优化奇异值计算

求 $C^\top C$ 的特征值分解

$$C^\top C = \sum_{i=1}^{c} \sigma_i^2(C) y_i y_i^\top \tag{5.9}$$

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

$$H = \begin{pmatrix} H^{(1)} \\ \cdots \\ H^{(d)} \end{pmatrix} = \begin{pmatrix} Cy_1/\sigma_1(C) \\ \cdots \\ Cy_d/\sigma_d(C) \end{pmatrix} \in \mathbb{R}^{p\times d} \tag{5.10}$$

优化奇异值计算

令

$$A_d = H\mathrm{Diag}(\sigma(C))\left(A^\top H\mathrm{Diag}(1/\sigma(C))\right)^\top \qquad (5.11)$$

可以证明 A_d 是 A 的一个比较好的近似

$$\|A - A_d\|_\xi^2 \leq \min_{\mathrm{rank}(D)\leq d} \|A - D\|_\xi^2 + ploylog(d, 1/c)\|A\|_\xi^2 \qquad (5.12)$$

其中 $\xi = 2$ 或 F , $polylog$ 是多重对数函数.

FPCA法算法

算法

步1 给出 X^1 , $\mu^0 > 0$, $1 > \theta > 0$, $\epsilon > 0$, 设置计数器 $k = 1$;

步2 若 $\mu^k = \mu^{k-1}\theta \leq \epsilon$, 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(5.8)至(5.11), 计算近似的奇异值分解 $Y_d = H\mathit{Diag}(\sigma(C))\left(Y^\top H\mathit{Diag}(1/\sigma(C))\right)^\top$

步6 计算 $X^{k+1} = S_{\psi\mu^k}(Y_d)$;

步7 $k = k + 1$, 转步2;

Thank you!