低秩矩阵完整化问题的几种高效解法

林陈冉

2016年12月17日

1 简介

1.1 什么是低秩矩阵完整化(low-rank matrix completion)

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.

1.2 数学模型

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.通常我们最感兴趣的是这样的一个最优化问题:

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

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

在一些情况下,(1.1)等价于线性约束问题:

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

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, $\mathcal{A}(X) = (\langle A_1, X\rangle, \cdots, \langle A_m, X\rangle)$, $A_i \in \mathbb{R}^{p\times q}, \forall i = 1, \cdots, m$, $\langle A, X\rangle$是矩阵内积,

在此给出线性算子$\mathcal{A}$的共轭算子的定义: $\mathcal{A}^* : \mathbb{R}^m \to \mathbb{R}^{p\times q}$, $\mathcal{A}^*(y) = \sum_{i=1}^m y_i A_i, \forall y = (y_1, \cdots, y_m) \in \mathbb{R}^m$.
容易验证$\mathcal{A}$和$\mathcal{A}^*$是well-defined,即

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

但是(1.1),(1.2)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(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.当在某些条件下,若已知元素个数$|\Omega| = m = O(nr{\cdot}\text{polylog}(n))$,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.

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$\mathcal{A}(X) = A\text{vec}(X)$,其中$\text{vec}(X) \in \mathbb{R}^{pq}$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$m \geq C(r(p+q)log(pq)$,其中$C$是一个正的常数,则矩阵有很高概率可以通过(1.4)恢复.

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.

下面,将对(1.3),(1.4),(1.5)分别给出一种高效的解法.

2 交替方向增广Lagrange法

对于问题(1.3),我们考虑它的对偶问题

$$
\begin{aligned}
\max_{y \in \mathbb{R}^m} \quad & b^\top y \\
\text{s.t.} \quad & \|\mathcal{A}^*(y)\|_2 \leq 1
\end{aligned} \tag{2.1}
$$

引入一个形式上的变量S,将(2.1)变为如下的等价形式

$$
\begin{aligned}
\min_{y \in \mathbb{R}^m} \quad & -b^\top y \\
\text{s.t.} \quad & \mathcal{A}^*(y) - S = 0 \\
& \|S\|_2 \leq 1
\end{aligned} \tag{2.2}
$$

可以考虑(2.2)的增广Lagrange函数

$$
L(y,S,X,\mu) = -b^\top y + \langle X, \mathcal{A}^*(y) - S\rangle + \frac{1}{2\mu}\|\mathcal{A}^*(y) - S\|_F^2 \tag{2.3}
$$

其中$X \in \mathbb{R}^{p\times q}$是某个给定的矩阵(不同于原问题中的$X$),$\|\cdot\|_F$是矩阵的$F$-**范数**,定义为矩阵所有元素的平方和的平方根.

由此我们得到(2.1)的一个等价问题

$$\begin{aligned} \min_{y,S} \quad & L(y,S,X,\mu) \\ \text{s.t.} \quad & \|S\|_2 \le 1 \end{aligned} \tag{2.4}$$

通过选取最佳的μ和X可以求出(2.4)的最优解.

采用如下的迭代过程以求解问题(2.4)

$$(y^{k+1}, S^{k+1}) = \arg\min_{y,S} L(y,S,X^k,\mu^k) \tag{2.5}$$

$$\mu^{k+1} \in [\alpha\mu^k, \mu^k], \alpha \in (0,1) \tag{2.6}$$

$$X^{k+1} = X^k + \frac{\mathcal{A}^*(y^{k+1}) - S^{k+1}}{\mu^k} \tag{2.7}$$

但(2.5)并不容易解得,因此我们考虑使用交替方向法来求解这个方程,即

$$y^{k+1} = \arg\min_{y,S} L(y,S^k,X^k,\mu^k) \tag{2.8}$$

$$S^{k+1} = \arg\min_{y,S} L(y^{k+1},S,X^k,\mu^k) \tag{2.9}$$

μ^{k+1}和X^{k+1}的求法不变.

首先来求解(2.8)

$$L(y,S^k,X^k,\mu^k) = -b^\top y + \mathcal{A}(X^k)^\top y - \langle X, S^k\rangle + \frac{1}{2\mu^k}\|\mathcal{A}^*(y) - S^k\|_F^2 \tag{2.10}$$

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

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

则

$$\begin{aligned} \frac{\partial\|T\|_F^2}{\partial y} = \frac{\partial \sum_{i,j}(\mathcal{A}_{ij}^\top y - S_{ij}^k)^2}{\partial y} &= \sum_{i,j}\frac{\partial(\mathcal{A}_{ij}^\top y - S_{ij}^k)^2}{\partial y} \\ &= 2\sum_{i,j}(\mathcal{A}_{ij}^\top y - S_{ij}^k)\mathcal{A}_{ij} = 2(\sum_{i,j} T_{ij}A_{k_{ij}}) = 2(\langle A_i, T\rangle) \\ &= 2\mathcal{A}(\mathcal{A}^*(y) - S^k) \end{aligned} \tag{2.11}$$

由(2.11)可以给出梯度

$$\frac{\partial L(y,S^k,X^k,\mu^k)}{\partial y} = -b + \mathcal{A}(X^k) + \frac{1}{\mu^k}\mathcal{A}(\mathcal{A}^*(y) - S^k) \tag{2.12}$$

令(2.12)式等于0,可得

$$y^{k+1} = \mu^k(b - \mathcal{A}(X^k)) + \mathcal{A}(S^k) \tag{2.13}$$

再考虑(2.9).将(2.10)如下变形

$$
\begin{aligned}
L(y^{k+1},S,X^k,\mu^k) &= -b^\top y^{k+1}+\mathcal{A}(X^k)^\top y^{k+1}-\langle X,S\rangle+\frac{1}{2\mu^k}\|\mathcal{A}^*(y^{k+1})-S\|_F^2 \\
&= \frac{1}{2\mu^k}(\|S\|_F^2-2\langle Y,S\rangle+\|\mathcal{A}^*(y^{k+1})\|_F^2)-b^\top y^{k+1}+\mathcal{A}(X^k)^\top y^{k+1}
\end{aligned}
\tag{2.14}
$$

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

$$
\frac{\partial L(y^{k+1},S,X^k,\mu^k)}{\partial S}=\frac{1}{2\mu^k}(\frac{\partial\|S\|_F^2}{\partial S}-2\frac{\partial\langle Y,S\rangle}{\partial S})=\frac{S-Y}{\mu^k} \tag{2.15}
$$

令(2.15)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

$$
S^{k+1}=UDiag(min\{\sigma,1\})V^\top \tag{2.16}
$$

其中$Y=UDiag(\sigma)V^\top$.根据Y的定义,可以简化X^{K+1}的计算方法

$$
X^{K+1}=\frac{\mu^kX^k+\mathcal{A}^*(y^{k+1})-S^{k+1}}{\mu^k}=\frac{Y-S^{k+1}}{\mu^k} \tag{2.17}
$$

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$\mu^0,X^0,y^0,S^0,\epsilon,\alpha$,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$Y=\mathcal{A}^*(y^{k+1})+\mu^kX^k$,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;

3 RBR法

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

$$
\begin{aligned}
\min_{X\in S^n}\quad & \langle C,X\rangle \\
\text{s.t.}\quad & \mathcal{A}(X)=b,X\succ 0
\end{aligned}
\tag{3.1}
$$

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

$$
X=\begin{pmatrix}\xi & y^\top\\ y & B\end{pmatrix} \tag{3.2}
$$

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

$$X = \begin{pmatrix} 1 & y^\top B^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} \xi - y^\top B^{-1} y & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} 1 & 0 \\ B^{-1}y & I \end{pmatrix} \tag{3.3}$$

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

显然的,

$$X \succeq 0 \Leftrightarrow B \succeq 0, (X/B) \geq 0 \tag{3.4}$$

约定以下记号

$$X_{\alpha,\beta} = \begin{cases} x_{\alpha\beta} & \alpha, \beta \in \mathbb{R} \\ (x_{\alpha\beta_1}, \cdots, x_{\alpha\beta_n}) & \alpha \in \mathbb{R}, \beta = \{\beta_1, \cdots, \beta_n\} \\ (x_{\alpha_1\beta}, \cdots, x_{\alpha_m\beta})^\top & \alpha = \{\alpha_1, \cdots, \alpha_n\}, \beta \in \mathbb{R} \\ \begin{pmatrix} x_{\alpha_1\beta_1} & \cdots & x_{\alpha_1\beta_n} \\ \cdots & \cdots & \cdots \\ x_{\alpha_m\beta_1} & \cdots & x_{\alpha_m\beta_n} \end{pmatrix} & \alpha = \{\alpha_1, \cdots, \alpha_n\}, \beta = \{\beta_1, \cdots, \beta_n\} \end{cases} \tag{3.5}$$

$$i^c = \{1, \cdots, n\} \backslash \{i\} = \{1, \cdots, i-1, i+1, \cdots, n\} \tag{3.6}$$

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(3.4), 令i取遍$\{1, \cdots, n\}$,逐行解如下的SOCP问题来解决SDP问题(3.1)

$$\begin{aligned} \min_{[\xi;y] \in \mathbb{R}^n} \quad & \bar{c}^\top \begin{pmatrix} \xi \\ y \end{pmatrix} \\ \text{s.t.} \quad & \bar{X} \begin{pmatrix} \xi \\ y \end{pmatrix} = \bar{b}, \\ & (X/B) \geq \delta \end{aligned} \tag{3.7}$$

其中

$$\bar{c} = \begin{pmatrix} C_{i,i} \\ 2Ci^c, i \end{pmatrix}, \quad \bar{X} = \begin{pmatrix} X^(1)_{i,i} & 2X^(1)_{i,i^c} \\ \cdots & \cdots \\ X^(m)_{i,i} & X^(m)_{i,i^c} \end{pmatrix}, \quad \bar{b} = \begin{pmatrix} b_1 - \langle X^(1)_{i^c,i^c}, B \rangle \\ \cdots \\ b_m - \langle X^(m)_{i^c,i^c}, B \rangle \end{pmatrix} \tag{3.8}$$

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$\delta > 0$

这种逐行求解的方法就是RBR方法.

考虑(3.7)的Powell罚函数

$$F(X, \theta, \mu) = \bar{c}^\top \begin{pmatrix} \xi \\ y \end{pmatrix} + \frac{1}{2\mu} \| \bar{X}[\xi; y] - \bar{b} - \theta \|_2^2 \tag{3.9}$$

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

(3.7)等价于以下问题

$$\begin{aligned}\min_X \quad & F(X,\lambda,\mu)=\bar{c}^\top\begin{pmatrix}\xi\\ y\end{pmatrix}+\frac{1}{2\mu}\|\bar{X}[\xi;y]-\lambda\|_2^2\\ \text{s.t.}\quad & X\succeq 0\end{aligned}\tag{3.10}$$

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

$$\begin{aligned}\min_X \quad & tr(X)=\langle E,X\rangle\\ \text{s.t.}\quad & X_{ij}=M_{ij},\forall(i,j)\in\Omega\end{aligned}\tag{3.11}$$

当$X\in\mathbb{R}^{p\times q}$不是对称正定的,可以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)

$$\begin{aligned}\min_X \quad & tr(X)\\ \text{s.t.}\quad & X=\begin{pmatrix}X_1 & W\\ W^\top & X_2\end{pmatrix}\succ 0\\ & W_{ij}=M_{ij},\forall(i,j)\in\Omega\end{aligned}\tag{3.12}$$

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(3.12).采用RBR法,对于某个i,我们把向量y分为两个部分,即

$$y=\begin{pmatrix}y_1\\ y_2\end{pmatrix},\quad y_1=X_{\alpha_i,i},y_2=X_{\beta_i,i}\tag{3.13}$$

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$\bar{X}\begin{pmatrix}\xi\\ y\end{pmatrix}$,$\lambda$和$\bar{c}$的显式表达

$$\lambda=\begin{cases}(M_{i,\alpha_i-p})^\top, & i\le p\\ M_{\alpha_i,i-p}\,, & p<i\le n\end{cases},\quad \bar{X}\begin{pmatrix}\xi\\ y\end{pmatrix}=y_1,\bar{c}=(1,\overbrace{0,\cdots,0}^{n-1})\tag{3.14}$$

故(3.10)化为以下形式

$$\begin{aligned}\min\quad & \xi+\frac{1}{2\mu}\|y_1-\lambda\|_2^2\\ \text{s.t.}\quad & \xi-y^\top B^{-1}y\ge\delta\end{aligned}\tag{3.15}$$

容易发现, 当 $\xi=y^\top B^{-1}y+\delta$ 时, 才可能取到最小值, 则这个问题等价于

$$\min_y y^\top B^{-1}y+\frac{1}{2\mu}\|y_1-\lambda\|_2^2\tag{3.16}$$

显示表达为

$$\begin{pmatrix} X_{\alpha,\alpha} & X_{\alpha,\beta} \\ X_{\beta,\alpha} & X_{\beta,\beta} \end{pmatrix}^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \frac{1}{2\mu} \begin{pmatrix} y_1 - \lambda \\ 0 \end{pmatrix} = 0 \tag{3.17}$$

变形可得

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \frac{1}{2\mu} \begin{pmatrix} X_{\alpha,\alpha} \\ X_{\beta,\alpha} \end{pmatrix} (y_1 - \lambda) = 0 \tag{3.18}$$

则

$$\begin{aligned} y_1 =& (2\mu I + X_{\alpha,\alpha})^{-1} X_{\alpha,\alpha} \\ y_2 =& \frac{1}{2\mu} X_{\beta,\alpha}(y_1 - \lambda) \end{aligned} \tag{3.19}$$

那么由(3.17)和(3.19), 有

$$\xi = y^\top B^{-1} y + \delta = -\frac{1}{2\mu} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}^\top \begin{pmatrix} y_1 - \lambda \\ 0 \end{pmatrix} + \delta = \frac{1}{2\mu} y_1^\top (\lambda - y_1) + \delta \tag{3.20}$$

即我们证明了, (3.15)的最优解为:

$$\begin{aligned} y_1 =& (2\mu I + X_{\alpha,\alpha})^{-1} X_{\alpha,\alpha} \tilde{b} \\ y_2 =& \frac{1}{2\mu} X_{\beta,\alpha}(\lambda - y_1) \\ \xi =& \frac{1}{2\mu} y_1^\top (\lambda - y_1) + \delta \end{aligned} \tag{3.21}$$

我们可以让 i 循环取遍 $\{1, \cdots, n\}$, 逐行求解(3.15), 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 $X^k_{\alpha,\alpha} = x^l_{\alpha,\beta} = X^k_{\beta,\alpha} = 0$, $X^k_{\beta,\beta} = X^k$, 否则按通常定义求出以上几个量.

步6 按(3.21), 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7 $i = i + 1$, 转步3;

4 FPCA法

FPCA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, 那么只要求梯度为0的点. 但 $\|X\|_*$ 是不可微的, 故考虑次梯度, 即要求

$$0 \in \mu\partial\|X\|_* + g(X) \tag{4.1}$$

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

设 $Y=X-\psi g(X)$, $\psi>0$ 是给定常数, 则(4.1)等价于

$$0\in\psi\mu\partial\|X\|_*+X-(X-\psi g(X))=\psi\mu\partial\|X\|_*+X-Y \tag{4.2}$$

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

$$\min\psi\mu\|X\|_*+\frac{1}{2}\|X-Y\|_F^2 \tag{4.3}$$

当认为 Y 是给定的情况下, 该问题的解为 $S_{\psi\mu}(Y)$, 其中 $S_{\psi\mu}(Y)$ 是Shinkage算子, 定义如下

对 $X\in\mathbb{R}^{m\times n}$ 做奇异值分解, $X=U\mathrm{Diag}(\sigma)V^\top$, $r=\mathrm{rank}(X)$, $U\in\mathbb{R}^{m\times r}$, $V\in\mathbb{R}^{n\times r}$, 对 $\forall\nu>0$, 定义向量 $\bar{\sigma}=(\bar{\sigma_1},\cdots,\bar{\sigma_r})$, 其中$\bar{\sigma_i}=\max\{\sigma_i-\nu,0\}$, 那么Shinkage算子定义为 $S_\nu(X)=U\mathrm{Diag}(\bar{\sigma})V^\top$

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 $\partial\|X\|_*=\{UV^\top+W|U^\top W=0,WV=0,\|W\|_2\le1\}$, 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 $\bar{\sigma}\in\mathbb{R}_+^{m-r}$, $\|\sigma\|_\infty\le1$, 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

$$\begin{aligned}
&\nu(UV^\top+W)+X-Y\\
=&(\nu UIV^\top+\bar{U}[\mathrm{Diag}(\bar{\sigma}),0]\bar{V}^\top+U\mathrm{Diag}(\sigma)V^\top)-Y\\
=&\begin{pmatrix}U\\ \bar{U}\end{pmatrix}\left[\begin{pmatrix}\nu I&0&0\\0&0&0\end{pmatrix}+\begin{pmatrix}0&0&0\\0&\nu\mathrm{Diag}(\bar{\sigma})&0\end{pmatrix}+\begin{pmatrix}\mathrm{Diag}(\sigma)&0&0\\0&0&0\end{pmatrix}\right]\begin{pmatrix}V\\ \bar{V}\end{pmatrix}^\top-U_Y\mathrm{Diag}(\gamma)V_Y^\top\\
=&\begin{pmatrix}U\\ \bar{U}\end{pmatrix}\begin{pmatrix}\nu I+\mathrm{Diag}(\sigma)&0&0\\0&\nu\mathrm{Diag}(\bar{\sigma})&0\end{pmatrix}\begin{pmatrix}V\\ \bar{V}\end{pmatrix}^\top-U_Y\mathrm{Diag}(\gamma)V_Y^\top=0
\end{aligned} \tag{4.4}$$

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 $W=0$, (4.4)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(4.4)也成立.

故 $S_{\psi\mu}(Y)$ 是(4.3)的解.

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

$$X^*=S_{\psi\mu}(X^*-\mu g(X^*))=S_{\psi\mu}(X^*-\mu\mathcal{A}^*(\mathcal{A}(X^*)-b)) \tag{4.5}$$

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

$$\|S_\nu(Y_1)-S_\nu(Y_2)\|_F \le \|Y_1-Y_2\|_F \tag{4.6}$$

以及, 当 $\mu\in(0,2/\lambda_{max}(A^\top A))$, 其中 A 满足 $\mathcal{A}(X)=A\mathrm{vec}(X)$

$$\|h(X_1)-h(X_2)\|_F \le \|X_1-X_2\|_F \tag{4.7}$$

那么当 $\mu\in(0,2/\lambda_{max}(A^\top A))$

$$\|S_{\psi\mu}(h(X_1))-S_{\psi\mu}(h(X_2))\|_F \le \|X_1-X_2\|_F \tag{4.8}$$

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

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.

$X\in\mathbb{R}^{p\times q}$, 取 c , k

5 补充知识

此部分会给出一些阅读本文必备的结论,许多结论将略过证明.

5.1 奇异值分解

$\forall X\in\mathbb{R}^{p\times q}, X=UDV^\top$, 其中$U\in\mathbb{R}^{p\times p}, V\in\mathbb{R}^{q\times q}$ 是正交矩阵, $D=Diag(\sigma)\in\mathbb{R}^{p\times q}$ 对角线上为非负实数,其他位置为0的矩阵.

这个分解称为**奇异值分解(SVD)**, 对角矩阵$\mathrm{Diag}(\sigma)$的每个对角元素σ_i都是非负的, 称为矩阵的**奇异值**, 正交矩阵U的列向量称为矩阵的**左奇异向量**, 正交矩阵V的列向量称为矩阵的**右奇异向量**.

有时也认为$X=UDV^\top$, 其中$U\in\mathbb{R}^{p\times r}, V\in\mathbb{R}^{q\times r}$ 是列正交矩阵, $D\in\mathbb{R}^{r\times r}$ 是非负实对角矩阵, $r=rank(X)$.

5.2 罚函数

对于约束优化问题:

$$\begin{aligned} \min_x \quad & f(x), \\ \text{s.t.} \quad & c_i(x)=0, \quad i=1,\cdots,m_e; \\ & c_i(x)\ge 0, \quad i=m_e+1,\cdots,m \end{aligned} \tag{5.1}$$

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**:

$$P(x,\mu)=f(x)+\mu\|\bar{c}(x)\|_2^2 \tag{5.2}$$

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$\mu \to \infty$,(5.3)和原问题(5.1)等价.

$$\min_{x} \quad P(x,\mu) \tag{5.3}$$

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**(以下叙述仅考虑等值约束的情况,即$m_e = m$):

$$P(x,\lambda,\mu) = f(x) - \lambda^{\top} c(x) + \frac{1}{2}\mu\|c(x)\|_2^2 \tag{5.4}$$

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

(5.4)等价于以下的**Powell罚函数**:

$$P(x,\theta,\mu) = f(x) + \frac{1}{2}\mu\|c(x) - \theta\|_2^2 \tag{5.5}$$

其中$c(x), \theta \in \mathbb{R}^m$, (5.4)与(5.5)只相差一个与$x$无关的常数$\frac{1}{2}\sigma\|\theta\|_2^2$

5.3 对偶问题

对于约束优化问题(5.1),它的**Langrange函数**为:

$$L(x,\mu,\lambda) = f(x) + \sum_{i=1}^{m_e} \mu_i c_i(x) + \sum_{i=m_e+1}^{m} \lambda_i c_i(x) \tag{5.6}$$

定义$g(\mu,\lambda) = \inf_x L(x,\mu,\lambda)$, 通过这个函数我们可以把(5.1)转化为它的对偶问题:

$$\begin{aligned} \max_{\mu,\lambda} \quad & g(\mu,\lambda) \\ \text{s.t.} \quad & \mu \succeq 0 \end{aligned} \tag{5.7}$$

当满足一定条件(KKT条件)时,(5.7)和(5.1)是等价的.

5.4 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

$$f(y) - f(x) \geq u(y-x), \forall y \in E \tag{5.8}$$

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**.

次梯度集通常用$\partial f(x)$表示

$$\partial f(x) = \{u \in \mathbb{R}^m | f(y) - f(x) \geq u(y-x), \forall y \in E\} \tag{5.9}$$

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 5.1 *x^*是(5.1)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.*

5.5 矩阵范数

以下默认$A \in \mathbb{R}^{m\times n}, x \in \mathbb{R}^n$,$k = \max\{m,n\}, \lambda(A) = (\lambda_1(A), \cdots, \lambda_k(A)) \in \mathbb{R}^k$是矩阵$A$的奇异值向量,$\lambda_i(A)$是矩阵$A$的第i个奇异值.

5.5.1 诱导-p范数

矩阵的**诱导-p范数**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**:

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

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**:

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

5.5.2 F-范数

矩阵的**F-范数**定义类似于向量的2-范数:

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

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,可以看作内积诱导的范数,即$\|A\|_F = \langle A, A\rangle^{\frac{1}{2}}$.

5.5.3 Schaten-p范数

矩阵的**Schaten-p范数**是p-范数应用与矩阵奇异值向量所得的:

$$\|A\|_p = (\sum_i^k \lambda_i(A)^p)^{\frac{1}{p}} = \|\lambda(A)\|_p$$

矩阵的**核范数**,又称**迹范数**,是在$p = 1$时Schaten-p范数:

$$\|A\|_* = \sum_i^k \lambda_i(X) = tr(\sqrt{AA^\top}) = \|A\|_{tr}$$

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

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

核范数$\|A\|_* = \|\lambda(A)\|_1$,而矩阵的秩$rank(A) = \|\lambda(A)\|_0$,由此可以看出,核范数与秩的关系就是1-范数与0-范数的关系,故采用核范数来近似秩.

核范数是非光滑函数的,次梯度集如下

$$\partial\|X\|_* = \{UV^\top + W | U^\top W = 0, WV = 0, \|W\|_2 \leq 1\} \tag{5.10}$$

其中$X = UDV^\top$