附表 1

附表 1　　算法一　次梯度法求解带非负约束的数值算法

算法一伪代码：
1.输入： 观测数据矩阵 $\boldsymbol{X}$ ， $\boldsymbol{Y}$ ；正则化参数 λ 。
2.输出： 系数矩阵 $\boldsymbol{A}$ ；非负矩阵 $\boldsymbol{B}$ 。
3.初始化： $\boldsymbol{B}$ 为非负矩阵，根据 SVD 对 $\boldsymbol{A}$ 进行初始化。
4.迭代过程（直到式（1）收敛）：
（a）固定 $\boldsymbol{B}$ ，使用 SVD 求解 $\boldsymbol{A}$ ，如式（2）所示。
（b）当 $\boldsymbol{B}$ 未收敛时：
（i）对每个 i ，根据式（5）更新 $\boldsymbol{B}^i$ 。
（ii）检查 $\boldsymbol{B}$ 是否已收敛。
（c）检查目标函数是否已收敛。

附表 2

附表 2　　算法二　非负约束的稳健降秩回归算法

算法二伪代码：
输入：矩阵 $\boldsymbol{X}$,$\boldsymbol{Y}$ ，初始矩阵 $\boldsymbol{C}^{(0)}$ 和 $\boldsymbol{B}^{(0)}$（确保 $\boldsymbol{B}^{(0)} \geqslant 0$ ），阈值函数 Θ ，计数器 $t=0$ ，
重复，
（a） $t \leftarrow t+1$ 。
（b） $\boldsymbol{C}^{(t+1)} \leftarrow \overrightarrow{\Theta}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B}^{(t)};\lambda)$
（c）根据估计结果式（12）， $\acute{\boldsymbol{r}}^{(t+1)} \leftarrow \boldsymbol{R}\left(\boldsymbol{X},\boldsymbol{Y}-\boldsymbol{C}^{(t+1)},r\right)$。
直到收敛。

附表 3

附表 3 算法三 ADMM 算法

算法三伪代码：

(1)**初始化参数：** 主变量 $\boldsymbol{A}$, $\boldsymbol{D}$, $\boldsymbol{Z}$ 和 $\boldsymbol{W}$ 初始化为零矩阵，$\boldsymbol{A} \geqslant 0$ ；对偶变量 $\boldsymbol{B_D}$ ，$\boldsymbol{B_Z}$ 和 $\boldsymbol{B_W}$ 初始化为零矩阵；常数 $\rho > 0$ 和 ò>0 。

(2) **迭代直至满足停止准则** $\dfrac{\left\|\boldsymbol{a}_j^t - \boldsymbol{a}_j^{t-1}\right\|_2^2}{\left\|\boldsymbol{a}_j^{t-1}\right\|_2^2} \leqslant \varepsilon$ ，对于所有的 $j = 1,\ldots q$ ，$\boldsymbol{a}_j^t$ 是系数矩阵 $\boldsymbol{A}$ 的第 j 列在第 t 次迭代后的值。

（a）更新：$\boldsymbol{A}$ ，$\boldsymbol{Z}$ ，$\boldsymbol{W}$ ，$\boldsymbol{D}$ ；

（i）更新 $\boldsymbol{A}$ 并确保其非负：

$$\boldsymbol{A} = \max\left\{0, \left(\tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{X}}\right)^{-1}\tilde{\boldsymbol{X}}^{\mathrm{T}}\left(\underset{,}{\boldsymbol{n}} + \boldsymbol{B}\right)\right\},$$

（ii）$\boldsymbol{Z} = S\left(\boldsymbol{A} - \boldsymbol{B_Z}, \lambda\gamma/\rho\right)$。其中，$S$ 表示软阈值操作符，对矩阵逐元素应用。

$$S\left(\boldsymbol{A}_{ij}, b\right) = \operatorname{sign}\left(\boldsymbol{A}_{ij}\right)\max\left(\left|\boldsymbol{A}_{ij}\right| - b, 0\right)。$$

（iii）$\boldsymbol{W} = \sum_j \max\left(\omega_j - \lambda/\rho, 0\right)\boldsymbol{a}_j\boldsymbol{b}_j^{\mathrm{T}}$ ，其中 $\sum_j \omega_j \boldsymbol{a}_j \boldsymbol{b}_j^{\mathrm{T}}$ 是 $\boldsymbol{A} - \boldsymbol{B_W}$ 的奇异值分解。

（iv）$\boldsymbol{C} = \boldsymbol{XA} - \boldsymbol{B_D}$。令，

$$\boldsymbol{D}_{ij} = \begin{cases} \dfrac{\boldsymbol{Y}_{ij} + n\rho\boldsymbol{C}_{ij}}{1 + n\rho}, & \left|n\rho\left(\boldsymbol{Y}_{ij} - C_{ij}\right)\big/1 + n\rho\right| \leqslant \tau \\ \boldsymbol{Y}_{ij} - S\left(\boldsymbol{Y}_{ij} - C_{ij}, \tau/n\rho\right), & \text{其它} \end{cases}$$

（b）更新 $\boldsymbol{B_D}$ ，$\boldsymbol{B_Z}$ ，$\boldsymbol{B_W}$ ：

（i）$\boldsymbol{B_D} = \boldsymbol{B_D} + \boldsymbol{D} - \boldsymbol{XA}$ ；

（ii）$\boldsymbol{B_Z} = \boldsymbol{B_Z} + \boldsymbol{Z} - \boldsymbol{A}$ ；

（iii）$\boldsymbol{B_W} = \boldsymbol{B_W} + \boldsymbol{W} - \boldsymbol{A}$ 。

附录 1：

定义 1 与定义 2 的详细说明

定义 1 (阈值函数)阈值函数 $\Theta(t;\lambda)$ 是一个实值函数，定义在 $-\infty < t < \infty$ 和 $0 \leqslant \lambda < \infty$ 范围内，使得：

(1) $\Theta(-t;\lambda) = -\Theta(t;\lambda)$；

(2)对于 $t \leqslant t'$，有 $\Theta(t;\lambda) \leqslant \Theta(t';\lambda)$；

(3) $\lim\limits_{t\to\infty}\Theta(t;\lambda) = \infty$ ；

(4)对于 $0 \leqslant t < \infty$，有 $0 \leqslant \Theta(t;\lambda) \leqslant t$。

定义 2 (多元阈值函数)给定阈值函数 Θ 和任意向量 $\boldsymbol{d} \in {}^{m}$。多元阈值函数 $\vec{\Theta}(\boldsymbol{d};\lambda)$ 定义为：当 $\boldsymbol{d} \neq \boldsymbol{0}$ 时，$\vec{\Theta}(\boldsymbol{d};\lambda) = \boldsymbol{d}\Theta\left(\|\boldsymbol{d}\|_2;\lambda\right) \big/ \|\boldsymbol{d}\|_2$；当 $\boldsymbol{d} = \boldsymbol{0}$ 时，$\vec{\Theta}(\boldsymbol{d};\lambda) = \boldsymbol{0}$。对于任意矩阵：

$\boldsymbol{D} = (\boldsymbol{d}_1 \ldots \boldsymbol{d}_n)^{\mathrm{T}} \in {}^{n\times m}$，定义 $\vec{\Theta}(\boldsymbol{D};\lambda) = \left\{\vec{\Theta}(\boldsymbol{d}_1;\lambda) \ldots \vec{\Theta}(\boldsymbol{d}_n;\lambda)\right\}^{\mathrm{T}}$。

附录 2：

定理一的证明推导说明

首先介绍证明中出现的一些主要符号。

我们通过求解式（13）研究令 $\boldsymbol{A}=\left(\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_q\right)$ 的估计值 $\hat{\boldsymbol{A}}=\left(\hat{\boldsymbol{a}}_1,\hat{\boldsymbol{a}}_2,\ldots,\hat{\boldsymbol{a}}_q\right)$ 的理论性质。其中，$\boldsymbol{a}_i,\hat{\boldsymbol{a}}_i$ 为列向量，$i=1,2,\ldots,q$，$\boldsymbol{A}\in\mathbb{R}_+^{p\times q},\hat{\boldsymbol{A}}\in\mathbb{R}_+^{p\times q}$。令 $\mathcal{V}_{p,q}=\left\{\boldsymbol{U}\in\mathbb{R}^{p\times q}:\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{I}_q\right\}$ 表示 $p\times q$ 正交矩阵的 Stiefel 流形。在整个理论分析中，假设 $\hat{\boldsymbol{A}}$ 可以分解为 $\hat{\boldsymbol{A}}=\hat{\boldsymbol{U}}\hat{\Lambda}\left(\hat{\boldsymbol{V}}\right)^{\mathrm{T}}=\sum_{k=1}^{r}\hat{\lambda}_k\hat{\boldsymbol{u}}_k\left(\hat{\boldsymbol{v}}_k\right)^{\mathrm{T}}$，其中 $\hat{\boldsymbol{U}}\in\mathcal{V}_{p,r}$，$\hat{\boldsymbol{V}}\in\mathcal{V}_{q,r}$，$\max_k\left\|\hat{\boldsymbol{u}}_k\right\|_0\leq s_u$，以及 $\max_k\left\|\hat{\boldsymbol{v}}_k\right\|_0\leqslant s_v$，$s_v,s_v\ll n$，$r\ll n$，并且 $rs_us_v\ll n$，s_u 是行稀疏，s_v 是列稀疏。因此 $\hat{\boldsymbol{A}}$ 是稀疏且低秩的。令 $\mathcal{S}=\operatorname{supp}\left(\hat{\boldsymbol{A}}\right)$ 为 $\hat{\boldsymbol{A}}$ 的支持集,其基数为 $|\mathcal{S}|=s$，即 $\mathcal{S}$ 包含 $\hat{\boldsymbol{A}}$ 中非零元素的索引。同时，$s\leqslant rs_us_v$。

对于任意向量 $\boldsymbol{u}=(u_1,\ldots,u_p)^{\mathrm{T}}\in\mathbb{R}^p$ 且 $q\geqslant1$，定义 $\|\boldsymbol{u}\|_q=\left(\sum_{j=1}^{p}\left|\boldsymbol{u}_j\right|^q\right)^{1/q}$ 为 ℓ_q 范数。$\|\boldsymbol{u}\|_0=\sum_{j=1}^{p}1\left(u_j\neq0\right)$ 表示 $\boldsymbol{u}$ 中非零元素的个数，$\|\boldsymbol{u}\|_\infty=\max_{1\leqslant j\leqslant p}\left|\boldsymbol{u}_j\right|$。对于任意两个向量 $\boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^p$ 的内积记为 $\langle\boldsymbol{u},\boldsymbol{v}\rangle=\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}$。设 $\boldsymbol{M}$ 为 $m\times n$ 实矩阵，q 阶范数 $\|\boldsymbol{M}\|_q=\max_{\boldsymbol{u}\in\mathbb{R}^n}\frac{\|\boldsymbol{M}\boldsymbol{u}\|_q}{\|\boldsymbol{u}\|_q}$，核范数 $\|\boldsymbol{M}\|_*=\sum_{k=1}^{\min\{m,n\}}\lambda_k$，Frobenius 范数 $\|\boldsymbol{M}\|_F=\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}\boldsymbol{M}^2{}_{ij}}$，将矩阵 $\boldsymbol{M}$ 的列逐一堆叠，即可得到向量化形式 $vec\left(\boldsymbol{M}\right)$。

回顾优化问题的形式：

$$\min_{A\geqslant0}\left\{\mathcal{L}_\tau(\boldsymbol{A})+\lambda\left(\|\boldsymbol{A}\|_*+\gamma\|\boldsymbol{A}\|_{1,1}\right)\right\}$$

其中，$\mathcal{L}_\tau(\boldsymbol{A})=\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{q}\ell_\tau\left(Y_{ik}-\boldsymbol{X}_{i\cdot}^{\mathrm{T}}\boldsymbol{A}_{\cdot k}\right)$，以方便在证明中使用。令 $\boldsymbol{A}\in\mathbb{R}^{p\times q}$ 是一个秩为 r 的非负矩阵，具有奇异值分解 $\boldsymbol{U}\mathring{\boldsymbol{u}}\boldsymbol{V}^{\mathrm{T}}$，其中 $\boldsymbol{U}\in\mathbb{R}^{p\times r},\boldsymbol{V}\in\mathbb{R}^{q\times r},\mathring{\boldsymbol{u}}\in\mathbb{R}^{r\times r}$。核范数的次微分由下式给出：

$$\partial\|\boldsymbol{A}\|_*=\left\{\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}+\boldsymbol{W}:\boldsymbol{W}\in\mathbb{R}^{p\times q},\boldsymbol{U}^{\mathrm{T}}\boldsymbol{W}=\boldsymbol{0},\boldsymbol{W}\boldsymbol{V}=\boldsymbol{0},\|\boldsymbol{W}\|_2\leqslant1\right\}$$

设 $\mathcal{F}(r)=\left\{\boldsymbol{A}\in\mathbb{R}^{p\times q}:\operatorname{rank}\left(\boldsymbol{A}\right)\leqslant r\right\}$ 是一个代数多样体，其包含了所有秩至多为 r 的矩阵。与 $\mathcal{F}(r)$ 相关的切空间在 $\boldsymbol{A}$ 处可以表示为：

$$T(\boldsymbol{A})=\left\{\boldsymbol{U}\boldsymbol{W}_1^{\mathrm{T}}+\boldsymbol{W}_2\boldsymbol{V}^{\mathrm{T}}:\boldsymbol{W}_1\in\ {}^{q\times r},\boldsymbol{W}_2\in\ {}^{p\times r}\right\}$$

其中，$T(\boldsymbol{A})$ 可以解释为 ${}^{p\times q}$ 中的一个子空间。令 $\mathsf{P}_{T(\boldsymbol{A})}$ 表示到 $T(\boldsymbol{A})$ 的投影算子。则以下关系成立：

$\tilde{\boldsymbol{N}}\in\partial\|\boldsymbol{A}\|_*$，当且仅当，$\mathsf{P}_{T(\boldsymbol{A})}\left(\tilde{\boldsymbol{N}}\right)=\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}$ 且 $\left\|\mathsf{P}_{T(\boldsymbol{A})\perp}\tilde{\boldsymbol{N}}\right\|_2\leqslant 1$。

此外,定义了一些将在证明中使用的量。

对于任何凸损失函数 $\mathsf{L}_\tau(\cdot)$，$\hat{\boldsymbol{A}}$ 和 $\boldsymbol{A}$ 之间的非对称 Bregman 散度为：

$$D_{\mathsf{L}}\left(\hat{\boldsymbol{A}},\boldsymbol{A}\right)=\mathsf{L}_\tau\left(\hat{\boldsymbol{A}}\right)-\mathsf{L}_\tau(A)-\left\langle\nabla\mathsf{L}_\tau(\boldsymbol{A}),\hat{\boldsymbol{A}}-\boldsymbol{A}\right\rangle\geqslant 0$$

定义对称 Bregman 散度为：

$$D_{\mathsf{L}}^S(\hat{\boldsymbol{A}},\boldsymbol{A})=D_{\mathsf{L}}\left(\hat{\boldsymbol{A}},\boldsymbol{A}\right)+D_{\mathsf{L}}(\boldsymbol{A},\hat{\boldsymbol{A}})=\left\langle\nabla\mathsf{L}_\tau(\hat{\boldsymbol{A}})-\nabla\mathsf{L}_\tau(\boldsymbol{\acute{o}}),\hat{\boldsymbol{A}}-\boldsymbol{A}\right\rangle\geqslant 0$$

该证明涉及对称 Bregman 散度的上界和下界。为此，我们陈述了一些将在证明中使用的条件与技术性引理。

定义 4 （矩阵的受限特征值）设 $\xi>1$。给定矩阵 $\boldsymbol{S}$，其最小与最大 $(\xi,m)-$受限特征值分别定义为

$$\rho_-\left(\boldsymbol{S},\xi,m\right)=\inf_{\boldsymbol{U}}\left\{\frac{tr\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{U}\right)}{\|\boldsymbol{U}\|_{1,2}^2}:\boldsymbol{U}\in\ {}^{p\times q},\boldsymbol{U}\neq\boldsymbol{0},S\subseteq J,|J|\leqslant m,\left\|\boldsymbol{U}_{J^c}\right\|_{1,1}\leqslant\xi\left\|\boldsymbol{U}_J\right\|_{1,1}\right\}$$

$$\rho_+\left(\boldsymbol{S},\xi,m\right)=\sup_{\boldsymbol{U}}\left\{\frac{tr\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{U}\right)}{\|\boldsymbol{U}\|_{1,2}^2}:\boldsymbol{U}\in\ {}^{p\times q},\boldsymbol{U}\neq\boldsymbol{0},S\subseteq J,|J|\leqslant m,\left\|\boldsymbol{U}_{J^c}\right\|_{1,1}\leqslant\xi\left\|\boldsymbol{U}_J\right\|_{1,1}\right\}$$

其中，$\|\boldsymbol{U}\|_{2,2}^2=\sum_{i,j}\boldsymbol{U}^2{}_{i,j}$ 表示 Frobenius 范数平方，$\boldsymbol{U}_J$ 是 $\boldsymbol{U}$ 在列指标集 J 上的子矩阵。

条件 3 存在常数 $0<K_{\text{lower}}\leqslant K_{\text{upper}}<\infty$，使得 $\boldsymbol{S}$ 的受限特征值满足

$$K_{\text{lower}}\leqslant\rho_-\left(\boldsymbol{S},\xi,m\right)\leqslant\rho_+\left(\boldsymbol{S},\xi,m\right)\leqslant K_{\text{upper}}$$

引理 1 设矩阵 $\boldsymbol{A}^*\in\mathsf{C}(m,\xi,\eta)$，其中

$$\mathsf{C}(m,\xi,\eta)=\left\{(\boldsymbol{A}^*,\boldsymbol{U})\in\ {}^{p\times q}\times\ {}^{p\times q}:\boldsymbol{U}\neq\boldsymbol{0},\mathsf{S}\subseteq J,|J|\leqslant m,\left\|\boldsymbol{U}_{\mathsf{S}^c}\right\|_{1,1}\leqslant\xi\left\|\boldsymbol{U}_{\mathsf{S}}\right\|_{1,1},\left\|\boldsymbol{A}^*-\boldsymbol{A}\right\|_{1,1}\leqslant\eta\right\}$$

取 $\tau\geqslant\max\left\{8\eta,C\cdot\left(mv_\delta\right)^{1/1+\delta}\right\}$，$n>C'm^2\log\left(pq\right)$，$C,C'>0$ 为充分大的常数。条件 1 与条件 3 成立的前提下，存在常数 K_{lower} 与 K_{upper}，使得矩阵 $\boldsymbol{H}_\tau\left(\boldsymbol{A}^*\right)$ 的局部受限特征值满足

$$0<\frac{K_{\text{lower}}}{2}\leqslant K_-\left(\boldsymbol{H}_\tau\left(\boldsymbol{A}^*\right),\xi,\eta\right)\leqslant K_+\left(\boldsymbol{H}_\tau\left(\boldsymbol{A}^*\right),\xi,\eta\right)\leqslant K_{\text{upper}}<\infty$$

且以至少 $1-\left(pq\right)^{-1}$ 的概率成立。

引理 2 (梯度无穷范数边界) 假设协变量标准化，使得 $\max\limits_{i,j}\left|\boldsymbol{X}_{ij}\right|=1$，且 E_{ik} 满足 $v_\delta=\mathsf{E}\left(\left|E_{ik}\right|^{1+\delta}\right)<\infty$。令 $\boldsymbol{A}=\left(\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_q\right)$，$\hat{\boldsymbol{A}}=\left(\hat{\boldsymbol{a}}_1,\hat{\boldsymbol{a}}_2,\ldots,\hat{\boldsymbol{a}}_q\right)$ 为估计值，$\boldsymbol{a}_i,\hat{\boldsymbol{a}}_i$ 为列向量，$i=1,2,\ldots,q$。选择 $\tau\geqslant C_1\left\{nv_\delta/\log\left(pq\right)\right\}^{\min\{1/2,1/(1+\delta)\}}$，则有

$$\left\|\nabla\mathsf{L}_{\tau}(\boldsymbol{A})\right\|_{\infty,\infty}\leqslant C_2 v_{\delta}^{1/\min(1+\delta,2)}\left(\frac{\log(pq)}{n}\right)^{\min\{1/2,\delta/(1+\delta)\}}$$

的概率为$1-(pq)^{-1}$，其中C_1和C_2是常数。

引理 3 ($\ell_{1,1}$−锥约束性质) 假设$\left\|\nabla\mathsf{L}_{\tau}(\boldsymbol{A})\right\|_{\infty,\infty}\leqslant\frac{\lambda}{2}$，$\gamma>3/2$。设$\hat{\boldsymbol{A}}$是式(12)的一个解，$\boldsymbol{A}=\left(\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_q\right)$，其估计$\hat{\boldsymbol{A}}=\left(\hat{\boldsymbol{a}}_1,\hat{\boldsymbol{a}}_2,\ldots,\hat{\boldsymbol{a}}_q\right)$，$\boldsymbol{A}\in{}_{+}^{p\times q},\hat{\boldsymbol{A}}\in{}_{+}^{p\times q}$那么有$\hat{\boldsymbol{A}}$落在以下$\ell_{1,1}$−锥中，

$$\left\|\left(\hat{\boldsymbol{A}}-\boldsymbol{A}\right)_{\mathsf{S}^c}\right\|_{1,1}\leqslant\frac{2\gamma+3}{2\gamma-3}\left\|\left(\hat{\boldsymbol{A}}-\boldsymbol{A}\right)_{\mathsf{S}}\right\|_{1,1}$$

引理 3 的证明

$$D_{\mathsf{L}}^{s}\left(\hat{\boldsymbol{A}},\boldsymbol{A}\right)=\left\langle\nabla\mathsf{L}_{\tau}(\boldsymbol{A}),\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle+\lambda\left\langle\tilde{\boldsymbol{N}},\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle+\lambda\gamma\left\langle\tilde{\boldsymbol{s}},\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle+\left\langle\prod,\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle\leqslant 0$$

接下来分解并计算各组成部分的上界。

令：$\boldsymbol{I}_1=\left\langle\nabla\mathsf{L}_{\tau}(\boldsymbol{A}),\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle$ ，$\boldsymbol{I}_2=\lambda\left\langle\tilde{\boldsymbol{N}},\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle$，$\boldsymbol{I}_3=\lambda\gamma\left\langle\tilde{\boldsymbol{s}},\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle$，$\boldsymbol{I}_4=\left\langle\prod,\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle$。记$\Delta=\hat{\boldsymbol{A}}-\boldsymbol{A}$，则

$$\boldsymbol{I}_1=\left\langle\nabla\mathsf{L}_{\tau}(\boldsymbol{A}),\Delta\right\rangle,\quad\boldsymbol{I}_2=\lambda\left\langle\tilde{\boldsymbol{N}},\Delta\right\rangle,\quad\boldsymbol{I}_3=\lambda\gamma\left\langle\tilde{\boldsymbol{s}},\Delta\right\rangle,\quad\boldsymbol{I}_4=\left\langle\prod,\Delta\right\rangle$$

记，$S=\operatorname{supp}\left(\boldsymbol{A}\right)$，$s_0:=\left\|\Delta_S\right\|_{1,1}$，$s_1:=\left\|\Delta_{S^C}\right\|_{1,1}$，则

$$\boldsymbol{I}_1=\left\langle\nabla\mathsf{L}_{\tau}(\boldsymbol{A}),\Delta\right\rangle\geqslant-\left\|\nabla\mathsf{L}_{\tau}(\boldsymbol{A})\right\|_{\infty,\infty}\left\|\Delta\right\|_{1,1}\geqslant-\frac{\lambda}{2}\left\|\Delta\right\|_{1,1}=-\frac{\lambda}{2}\left(s_0+s_1\right)$$

$$\boldsymbol{I}_2=\lambda\left\langle\tilde{\boldsymbol{N}},\Delta\right\rangle\geqslant-\lambda\left\|\tilde{\boldsymbol{N}}\right\|_{\infty,\infty}\left\|\Delta\right\|_{1,1}\geqslant-\lambda\left(s_0+s_1\right)$$

其中，取$\tilde{\boldsymbol{N}}=\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}$，令$\boldsymbol{W}=0$，在分解$\tilde{\boldsymbol{N}}=\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}+\boldsymbol{W}$中，则$\left\|\tilde{\boldsymbol{N}}\right\|_{\infty,\infty}\leqslant 1$

$$\begin{aligned}\boldsymbol{I}_3&=\lambda\gamma\left\langle\tilde{\boldsymbol{s}},\Delta\right\rangle\\&\geqslant\lambda\gamma\left\langle\tilde{\boldsymbol{s}}_S,\Delta_S\right\rangle+\lambda\gamma\left\langle\tilde{\boldsymbol{s}}_{S^C},\Delta_{S^C}\right\rangle\\&\geqslant-\lambda\gamma\left\|\Delta_S\right\|_{1,1}+\lambda\gamma\left\|\Delta_{S^C}\right\|_{1,1}\\&=-\lambda\gamma s_0+\lambda\gamma s_1\end{aligned}$$

其中，$\left\|\tilde{\boldsymbol{s}}_S\right\|_{\infty,\infty}\leqslant 1$

$$\boldsymbol{I}_4=\left\langle\prod,\Delta\right\rangle\geqslant 0$$

将上述结果代入下式

$$D_{\mathsf{L}}^{s}\left(\hat{\boldsymbol{A}},\boldsymbol{A}\right)=\left\langle\nabla\mathsf{L}_{\tau}(\boldsymbol{A}),\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle+\lambda\left\langle\tilde{\boldsymbol{N}},\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle+\lambda\gamma\left\langle\tilde{\boldsymbol{s}},\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle+\left\langle\prod,\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle\leqslant 0$$

得到

$$-\frac{\lambda}{2}\left(s_0+s_1\right)-\lambda\left(s_0+s_1\right)-\lambda\gamma s_0+\lambda\gamma s_1\leqslant 0$$

$$\left(\gamma-\frac{3}{2}\right)s_1\leqslant\left(\gamma+\frac{3}{2}\right)s_0$$

$$\frac{s_1}{s_0}=\frac{2\gamma+3}{2\gamma-3}$$

即$\left\|\left(\hat{\boldsymbol{A}}-\boldsymbol{A}\right)_{\mathsf{S}^c}\right\|_{1,1}\leqslant\frac{2\gamma+3}{2\gamma-3}\left\|\left(\hat{\boldsymbol{A}}-\boldsymbol{A}\right)_{\mathsf{S}}\right\|_{1,1}$。

引理 4 (限制强凸性) 对于矩阵$(\boldsymbol{A}^*,\boldsymbol{U})\in\mathsf{C}(m,\xi,\eta)$，$\boldsymbol{A}=\left(\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_q\right)$，　$\boldsymbol{a}_i,\hat{\boldsymbol{a}}_i$为列向量，$\boldsymbol{A}^*=\left(\boldsymbol{a}^*{}_1,\boldsymbol{a}^*{}_2,\ldots,\boldsymbol{a}^*{}_q\right)$，

$$\mathsf{C}(m,\xi,\eta)=\left\{(\boldsymbol{A}^*,\boldsymbol{U})\in{}^{p\times q}\times{}^{p\times q}:\boldsymbol{U}\neq\boldsymbol{0},\mathsf{S}\subseteq J,|J|\leqslant m,\left\|\boldsymbol{U}_{\mathsf{S}^c}\right\|_{1,1}\leqslant\xi\left\|\boldsymbol{U}_{\mathsf{S}}\right\|_{1,1},\left\|\boldsymbol{A}^*-\boldsymbol{A}\right\|_{1,1}\leqslant\eta\right\}$$

在和引理 1 相同的条件下有：

$\mathsf{D}_{\mathsf{L}}^{s}\left(\boldsymbol{A}^*,\boldsymbol{A}\right)\geqslant\frac{K_{\text{lower}}}{2}\left\|\boldsymbol{A}^*-\boldsymbol{A}\right\|_F^2$的概率至少为$1-(pq)^{-1}$。

引理 1，引理 2 与引理 4 的证明详见 Tan 等（2023）的引理 1，引理 2 与引理 5。

证明定理 1：

首先，定义一阶优化条件

存在$\tilde{\boldsymbol{N}}\in\partial\left\|\hat{\boldsymbol{A}}\right\|_*$，$\tilde{\boldsymbol{s}}\in\partial\left\|\hat{\boldsymbol{A}}\right\|_{1,1}$，$\prod\in{}_{+}^{p\times q}$，对$\hat{\boldsymbol{A}}$有

$$\nabla\mathsf{L}_\tau\left(\hat{\boldsymbol{A}}\right)+\lambda\left(\tilde{\boldsymbol{N}}+\gamma\tilde{\boldsymbol{s}}\right)-\prod=\boldsymbol{0}$$

$$\prod\quad\hat{\boldsymbol{A}}=\boldsymbol{0}\text{，}\prod\geqslant\boldsymbol{0}$$

其中，$\prod$为非负拉格朗日乘子。互补条件$\prod\quad\hat{\boldsymbol{A}}=\boldsymbol{0}$，确保$\hat{\boldsymbol{A}}_{ij}=0$时$\prod_{ij}\geqslant0$，而$\hat{\boldsymbol{A}}_{ij}>0$时$\prod_{ij}=0$。

定义对称 Bregman 散度：

$$D_{\mathsf{L}}^{s}\left(\hat{\boldsymbol{A}},\boldsymbol{A}\right)=\left\langle-\lambda\tilde{\boldsymbol{N}}-\lambda\gamma\tilde{\boldsymbol{s}}-\nabla\mathsf{L}_\tau\left(\boldsymbol{A}\right)+\prod,\hat{\boldsymbol{A}}-\boldsymbol{A}\right\rangle$$

进一步

$$D_{\mathsf{L}}^{s}\left(\hat{\boldsymbol{A}},\boldsymbol{A}\right)=\left\langle\nabla\mathsf{L}_\tau(\boldsymbol{A}),\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle+\lambda\left\langle\tilde{\boldsymbol{N}},\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle+\lambda\gamma\left\langle\tilde{\boldsymbol{s}},\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle+\left\langle\prod,\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle$$

接下来分解并计算各组成部分的上界。

令：$\boldsymbol{I}_1=\left\langle\nabla\mathsf{L}_\tau(\boldsymbol{A}),\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle$，$\boldsymbol{I}_2=\lambda\left\langle\tilde{\boldsymbol{N}},\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle$，$\boldsymbol{I}_3=\lambda\gamma\left\langle\tilde{\boldsymbol{s}},\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle$，$\boldsymbol{I}_4=\left\langle\prod,\boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle$。

根据 Holder 不等式，

对$\boldsymbol{I}_1$的上界估计：

$$\begin{aligned}\boldsymbol{I}_1&\leqslant\left\|\nabla\mathsf{L}_\tau\left(\boldsymbol{A}\right)\right\|_{\infty,\infty}\left\|\hat{\boldsymbol{A}}-\boldsymbol{A}\right\|_{1,1}\\&\leqslant\frac{\lambda}{2}\left\|\hat{\boldsymbol{A}}-\boldsymbol{A}\right\|_{1,1}\\&\leqslant\frac{\lambda}{2}\left(\left\|\left(\hat{\boldsymbol{A}}-\boldsymbol{A}\right)_{\mathsf{S}}\right\|_{1,1}+\left\|\left(\hat{\boldsymbol{A}}-\boldsymbol{A}\right)_{\mathsf{S}^c}\right\|_{1,1}\right)\\&\leqslant\frac{2\lambda\gamma}{2\gamma-3}\left\|\left(\hat{\boldsymbol{A}}-\boldsymbol{A}\right)_{\mathsf{S}}\right\|_{1,1}\end{aligned}$$

对$\boldsymbol{I}_2$的上界估计：

$$
\begin{aligned}
\boldsymbol{I}_2 &\leqslant \lambda\left\|\tilde{\boldsymbol{N}}\right\|_{\infty,\infty}\left\|\hat{\boldsymbol{A}}-\boldsymbol{A}\right\|_{1,1} \\
&\leqslant \lambda\left\|\hat{\boldsymbol{A}}-\boldsymbol{A}\right\|_{1,1} \\
&\leqslant \frac{4\lambda\gamma}{2\gamma-3}\left\|\left(\hat{\boldsymbol{A}}-\boldsymbol{A}\right)_{\mathsf{S}}\right\|_{1,1}
\end{aligned}
$$

其中，$\left\|\tilde{\boldsymbol{N}}\right\|_{\infty,\infty}\leqslant 1$

对 $\boldsymbol{I}_3$ 的上界估计：

$$
\begin{aligned}
\boldsymbol{I}_3 &\leqslant \lambda\gamma\left\|\tilde{\boldsymbol{s}}\right\|_{\infty,\infty}\left\|\hat{\boldsymbol{A}}-\boldsymbol{A}\right\|_{1,1} \\
&\leqslant \lambda\gamma\left\|\hat{\boldsymbol{A}}-\boldsymbol{A}\right\|_{1,1} \\
&\leqslant \frac{4\lambda\gamma^2}{2\gamma-3}\left\|\left(\hat{\boldsymbol{A}}-\boldsymbol{A}\right)_{\mathsf{S}}\right\|_{1,1}
\end{aligned}
$$

其中，$\left\|\tilde{\boldsymbol{s}}\right\|_{\infty,\infty}\leqslant 1$

对 $\boldsymbol{I}_4$ 的上界估计：

$$
\begin{aligned}
\boldsymbol{I}_4 &= \left\langle \Pi, \boldsymbol{A}-\hat{\boldsymbol{A}}\right\rangle \\
&\leqslant \left\|\Pi\right\|_{\infty,\infty}\left\|\hat{\boldsymbol{A}}-\boldsymbol{A}\right\|_{1,1} \\
&\leqslant \frac{\lambda\left(4\gamma^2+8\gamma\right)}{2\gamma-3}\left\|\left(\hat{\boldsymbol{A}}-\boldsymbol{A}\right)_{S}\right\|_{1,1}
\end{aligned}
$$

其中，$\left\|\Pi\right\|_{\infty,\infty}\leqslant \lambda\left(\gamma+2\right)$

接着将 $\boldsymbol{I}_1$、$\boldsymbol{I}_2$、$\boldsymbol{I}_3$ 和 $\boldsymbol{I}_4$ 的上界代入到 $D_{\mathsf{L}}^{s}\left(\hat{\boldsymbol{A}},\boldsymbol{A}\right)$ 的计算中，得到：

$$
\begin{aligned}
D_{\mathsf{L}}^{s}\left(\hat{\boldsymbol{A}},\boldsymbol{A}\right) &\leqslant \frac{8\gamma^2+14\gamma}{2\gamma-3}\lambda\left\|\left(\hat{\boldsymbol{A}}-\boldsymbol{A}\right)_{\mathsf{S}}\right\|_{1,1} \\
&\leqslant \frac{8\gamma^2+14\gamma}{2\gamma-3}\lambda\sqrt{\mathrm{s}}\left\|\left(\hat{\boldsymbol{A}}-\boldsymbol{A}\right)_{S}\right\|_{F}
\end{aligned}
$$

其中，$s\le rs_u s_v$ 是 $\boldsymbol{A}$ 的稀疏性参数，$\left\|\left(\hat{\boldsymbol{A}}-\boldsymbol{A}\right)_{\mathsf{S}}\right\|_{1,1}\leqslant\sqrt{s}\left\|\left(\hat{\boldsymbol{A}}-\boldsymbol{A}\right)_{S}\right\|_{F}$，$s=\left|\mathrm{supp}\left(\boldsymbol{A}\right)\right|$。

接下来，利用引理 4 来获得对称 Bregman 散度的下界。引理 4 要求向量 $\boldsymbol{A}^*$ 在 $\mathsf{C}(m,\xi,\eta)$ 中。因此构造向量 $\hat{\boldsymbol{A}}_\eta=\boldsymbol{A}+\zeta\left(\hat{\boldsymbol{A}}-\boldsymbol{A}\right)$，对于 $\eta>0$，有 $\left\|\hat{\boldsymbol{A}}_\eta-\hat{\boldsymbol{A}}\right\|_{1,1}\leqslant\eta$。如果 $\left\|\hat{\boldsymbol{A}}-\boldsymbol{A}\right\|\leqslant\eta$，则设 $\zeta=1$，则 $\hat{\boldsymbol{A}}_\eta=\hat{\boldsymbol{A}}$。否则取 $\zeta\in(0,1)$，使得 $\left\|\hat{\boldsymbol{A}}_\eta-\hat{\boldsymbol{A}}\right\|_{1,1}\leqslant\eta$。因此，根据引理 3 可以证明 $\hat{\boldsymbol{A}}_\eta$ 落在一个 ℓ_1 锥中，因此 $\hat{\boldsymbol{A}}_\eta\in\mathsf{C}(m,\xi,\eta)$，并且有

$$
\left\|\left(\hat{\boldsymbol{A}}_\eta-\boldsymbol{A}\right)_{\mathsf{S}^c}\right\|_{1,1}\leqslant\frac{2\gamma+3}{2\gamma-3}\left\|\left(\hat{\boldsymbol{A}}_\eta-\boldsymbol{A}\right)_{\mathsf{S}}\right\|_{1,1} \text{ 和 } \quad \left\|\hat{\boldsymbol{A}}_\eta-\boldsymbol{A}\right\|_{1,1}\leqslant\eta \tag{17}
$$

再通过引理 4 有：

$$
D_{\mathsf{L}}^{s}(\hat{\boldsymbol{A}}_\eta,\boldsymbol{A})\geqslant\frac{K_{\text{lower}}}{2}\left\|\hat{\boldsymbol{A}}_\eta-\boldsymbol{A}\right\|_F^2 \tag{18}
$$

根据 Sun 等（2020）的引理 **A.1** 有：

$$
D_{\mathsf{L}}^{s}\left(\hat{\boldsymbol{A}}_\eta,\boldsymbol{A}\right)\leqslant\zeta D_{\mathsf{L}}^{s}\left(\hat{\boldsymbol{A}},\boldsymbol{A}\right) \tag{19}
$$

结合上述(18)式和(19)式，得到：

$$\left\|\hat{\boldsymbol{A}}_{\eta}-\boldsymbol{A}\right\|_F^2 \leqslant \zeta K_{\text{lower}}^{-1}\frac{16\gamma^2+28\gamma}{2\gamma-3}\lambda\left\|\left(\hat{\boldsymbol{A}}-\boldsymbol{A}\right)_{\mathrm{S}}\right\|_{1,1}$$

由于$\left\|\left(\hat{\boldsymbol{A}}-\boldsymbol{A}\right)_{\mathrm{S}}\right\|_{1,1} \leqslant \sqrt{s}\left\|\left(\hat{\boldsymbol{A}}-\boldsymbol{A}\right)_{S}\right\|_F$，因此

$$\left\|\hat{\boldsymbol{A}}_{\eta}-\boldsymbol{A}\right\|_F^2 \leqslant \zeta K_{\text{lower}}^{-1}\frac{16\gamma^2+28\gamma}{2\gamma-3}\lambda\sqrt{s}\left\|\hat{\boldsymbol{A}}-\boldsymbol{A}\right\|_F$$

由于$\hat{\boldsymbol{A}}-\boldsymbol{A}=\zeta^{-1}\left(\hat{\boldsymbol{A}}_{\eta}-\boldsymbol{A}\right)$，则

$$\left\|\hat{\boldsymbol{A}}_{\eta}-\boldsymbol{A}\right\|_F \leqslant K_{\text{lower}}^{-1}\frac{16\gamma^2+28\gamma}{2\gamma-3}\lambda\sqrt{s}$$

根据式(17)，当$n>Cs^2\log\left(pq\right)$时，$C$是某个足够大的常数。则有，

$$\begin{aligned}\left\|\hat{\boldsymbol{A}}_{\eta}-\boldsymbol{A}\right\|_{1,1} &\leqslant \frac{4\gamma}{2\gamma-3}\left\|\left(\hat{\boldsymbol{A}}_{\eta}-\boldsymbol{A}\right)_{\mathrm{S}}\right\|_{1,1}\\ &\leqslant \frac{4\gamma\sqrt{s}}{2\gamma-3}\left\|\left(\hat{\boldsymbol{A}}_{\eta}-\boldsymbol{A}\right)_{\mathrm{S}}\right\|_F\\ &\leqslant K_{\text{lower}}^{-1}\frac{4\gamma}{2\gamma-3}\frac{16\gamma^2+28\gamma}{2\gamma-3}\lambda s\\ &<\eta\end{aligned}$$

由于，$\left\|\hat{\boldsymbol{A}}_{\eta}-\boldsymbol{A}\right\|_{1,1}<\eta$ ，则$\hat{\boldsymbol{A}}_{\eta}=\hat{\boldsymbol{A}}$，因此

$$\left\|\hat{\boldsymbol{A}}-\boldsymbol{A}\right\|_F \leqslant K_{\text{lower}}^{-1}\frac{16\gamma^2+28\gamma}{2\gamma-3}\lambda\sqrt{s}$$

由于$\hat{\boldsymbol{A}}=\left(\hat{\boldsymbol{a}}_1,\hat{\boldsymbol{a}}_2,\ldots,\hat{\boldsymbol{a}}_q\right)$，$\boldsymbol{A}=\left(\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_q\right)$，$\boldsymbol{a}_i,\hat{\boldsymbol{a}}_i$为列向量，$i=1,2,\ldots,q$。则对于每个向量$\hat{\boldsymbol{a}}_i$

$$\left\|\hat{\boldsymbol{A}}-\boldsymbol{A}\right\|_F^2=\sum_{i=1}^{q}\left\|\hat{\boldsymbol{a}}_i-\boldsymbol{a}_i\right\|_2^2$$

于是对于任意i都有，

$$\left\|\hat{\boldsymbol{a}}_i-\boldsymbol{a}_i\right\|_2 \leqslant \left\|\hat{\boldsymbol{A}}-\boldsymbol{A}\right\|_F$$

则

$$\left\|\hat{\boldsymbol{a}}_i-\boldsymbol{a}_i\right\|_2 \leqslant \left\|\hat{\boldsymbol{A}}-\boldsymbol{A}\right\|_F \leqslant K_{\text{lower}}^{-1}\frac{16\gamma^2+28\gamma}{2\gamma-3}\lambda\sqrt{s}$$

进一步得到

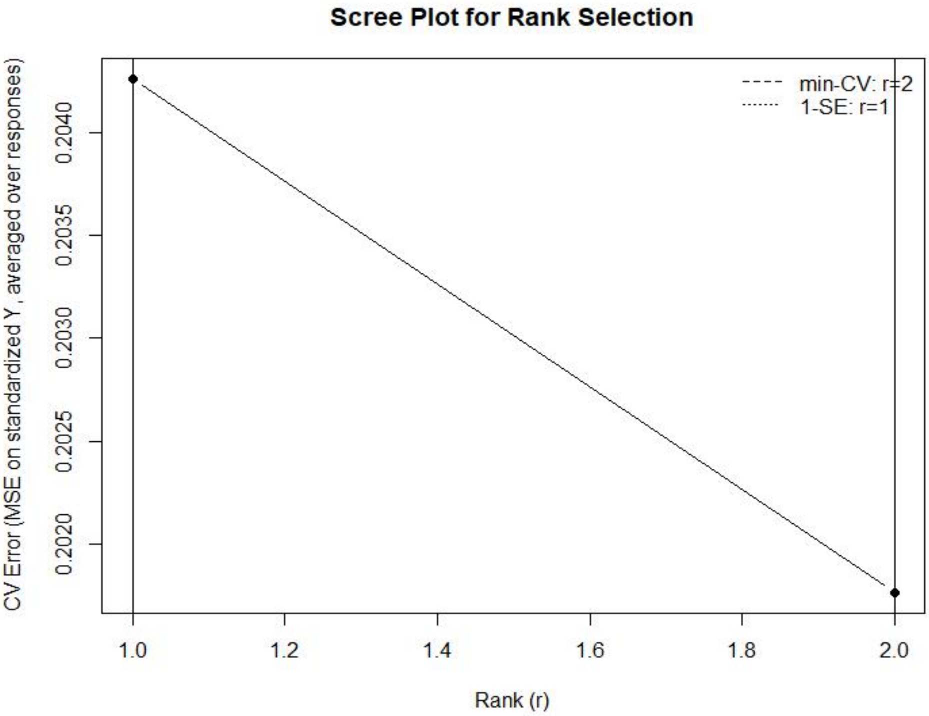
$$\left\|\hat{\boldsymbol{a}}_i-\boldsymbol{a}_i\right\|_2 \leqslant K_{\text{lower}}^{-1}\frac{16\gamma^2+28\gamma}{2\gamma-3}\lambda\sqrt{s_i}\ ,\quad i=1,2,\ldots,q$$

综上，定理 1 得证。

参考文献

[1] Sun Q, Zhou WX, Fan J. Adaptive Huber Regression [J]. Journal of the American Statistical Association, 2020, 115(529): 254-265.

附图 1



附图 1 $r=1$ 和 $r=2$ 时的最小交叉验证误差轨迹图