Some basic results of Low-Rank and Sparse

Yao Zhang

1 Overview

$$\min_X \tau \|X\|_* + \frac{1}{2}\|X - A\|_F^2 \tag{1}$$

$$\min_X \tau \|X\|_1 + \frac{1}{2}\|X - A\|_F^2 \tag{2}$$

2 ADMM Algorithm

Example 1 (RPCA)**.**

$$\min_{L,S} \|L\|_* + \lambda\|S\|_1 \quad s.t.\ X = L + S \tag{3}$$

where $X, L, S, Y \in \mathbb{R}^{m\times n}$.
The Lagrangian function $\mathcal{L}(L, S, Y)$ *is as following*

$$\mathcal{L}(L, S, Y) = \|L\|_* + \lambda\|S\|_1 + \langle Y, X - (L+S)\rangle + \frac{\mu}{2}\|X - (L+S)\|_F^2 \tag{4}$$

ADMM:

1. *update* L

$$\begin{aligned} L^* &= \arg\min_L \|L\|_* + \langle Y, X - (L+S)\rangle + \frac{\mu}{2}\|X - (L+S)\|_F^2 \\ &= \arg\min_L \|L\|_* + \frac{\mu}{2}\left\|X - (L+S) + \frac{Y}{\mu}\right\|_F^2 \\ &= \arg\min_L \frac{1}{\mu}\|L\|_* + \frac{1}{2}\left\|L - \left(X + \frac{Y}{\mu} - S\right)\right\|_F^2 \end{aligned} \tag{5}$$

2. *update* S

$$\begin{aligned} S^* &= \arg\min_S \lambda\|S\|_1 + \langle Y, X - (L+S)\rangle + \frac{\mu}{2}\|X - (L+S)\|_F^2 \\ &= \arg\min_S \lambda\|S\|_1 + \frac{\mu}{2}\left\|X - (L+S) + \frac{Y}{\mu}\right\|_F^2 \\ &= \arg\min_S \frac{\lambda}{\mu}\|S\|_1 + \frac{1}{2}\left\|S - \left(X + \frac{Y}{\mu} - L\right)\right\|_F^2 \end{aligned} \tag{6}$$

3. *update* Y

$$Y^{k+1} = Y^k + \mu(X - (L+S)) \tag{7}$$

Example 2 (LRR)**.**

$$\min_{Z,E} \|Z\|_* + \lambda\|E\|_1 \quad s.t.\ X = XZ + E \tag{8}$$

where $X, E \in \mathbb{R}^{m\times n}, Z \in \mathbb{R}^{n\times n}$.
Eq.8 equals to

$$\min_{Z,E,A} \|Z\|_* + \lambda\|E\|_1 \quad s.t.\ X = XA + E, \quad A = Z \tag{9}$$

where $X, E \in \mathbb{R}^{m\times n}, Z, A \in \mathbb{R}^{n\times n}$.
The Lagrangian function $\mathcal{L}(Z, E, A, Y_1, Y_2)$ *is as following*

$$\begin{aligned}&\mathcal{L}(Z, E, A, Y_1, Y_2) = \\ &\quad \|Z\|_* + \lambda\|E\|_1 + \langle Y_1, X - (XA + E)\rangle + \langle Y_2, A - Z\rangle + \frac{\mu}{2}\left(\|X - (XA + E)\|_F^2 + \|A - Z\|_F^2\right)\end{aligned} \tag{10}$$

ADMM:

1. *update Z*

$$\begin{aligned} Z^* &= \arg\min_Z \|Z\|_* + \langle Y_2, A - Z\rangle + \frac{\mu}{2}\|A - Z\|_F^2 \\ &= \arg\min_Z \|Z\|_* + \frac{\mu}{2}\left\|A - Z + \frac{Y_2}{\mu}\right\|_F^2 \\ &= \arg\min_Z \frac{1}{\mu}\|Z\|_* + \frac{1}{2}\left\|Z - \left(A + \frac{Y_2}{\mu}\right)\right\|_F^2 \end{aligned} \tag{11}$$

2. *update E*

$$\begin{aligned} E &= \arg\min_E \lambda\|E\|_1 + \langle Y_1, X - (XA + E)\rangle + \frac{\mu}{2}\|X - (XA + E)\|_F^2 \\ &= \arg\min_E \lambda\|E\|_1 + \frac{\mu}{2}\left\|X - (XA + E) + \frac{Y_1}{\mu}\right\|_F^2 \\ &= \arg\min_E \frac{\lambda}{\mu}\|E\|_1 + \frac{1}{2}\left\|E - \left(X + \frac{Y_1}{\mu} - XA\right)\right\|_F^2 \end{aligned} \tag{12}$$

3. *update A*

$$\begin{aligned} A^* &= \arg\min_A \langle Y_1, X - (XA + E)\rangle + \langle Y_2, A - Z\rangle + \frac{\mu}{2}\left(\|X - (XA + E)\|_F^2 + \|A - Z\|_F^2\right) \\ &= \arg\min_A \frac{\mu}{2}\left(\left\|X - (XA + E) + \frac{Y_1}{\mu}\right\|_F^2 + \left\|A - Z + \frac{Y_2}{\mu}\right\|_F^2\right) \\ &= \left\|X - (XA + E) + \frac{Y_1}{\mu}\right\|_F^2 + \left\|A - Z + \frac{Y_2}{\mu}\right\|_F^2 \end{aligned} \tag{13}$$

denote that

$$q = \left\|X - (XA + E) + \frac{Y_1}{\mu}\right\|_F^2 + \left\|A - Z + \frac{Y_2}{\mu}\right\|_F^2 \tag{14}$$

let

$$\frac{\partial q}{\partial A} = -2X^T\left(X-(XA+E)+\frac{Y_1}{\mu}\right)+2\left(A-Z+\frac{Y_2}{\mu}\right)=0 \tag{15}$$

so

$$\frac{\partial q}{\partial A}=0 \Rightarrow A=\left(X^TX+I\right)^{-1}\left(X^TX+Z-X^TE+\frac{X^TY_1-Y_2}{\mu}\right) \tag{16}$$

4. *update* Y_1

$$Y_1^{k+1}=Y_1^k+\mu\left(X-(XA+E)\right) \tag{17}$$

5. *update* Y_2

$$Y_2^{k+1}=Y_2^k+\mu\left(A-Z\right) \tag{18}$$

3 LADMM Algorithm

idea:

$$f(x)=f(x_0)+f'(\xi)(x-x_0) \tag{19}$$

where $\xi : x \leqslant \xi \leqslant x_0$ or $x_0 \leqslant \xi \leqslant x$, which dependents on the $x_0, x's$ value.

Example 3 (RPCA)**.**

$$\min_{L,S} \|L\|_* + \lambda\|S\|_1 \quad s.t.\ X = L+S \tag{20}$$

The Lagrangian function $\mathcal{L}(L,S,Y)$ *is as following:*

$$\mathcal{L}(L,S,Y)=\|L\|_*+\lambda\|S\|_1+\langle Y, X-(L+S)\rangle+\frac{\mu}{2}\|X-(L+S)\|_F^2 \tag{21}$$

1. *update* L

$$\begin{aligned} L^{k+1} &= \arg\min_L \|L\|_* + \langle Y, X-(L+S)\rangle + \frac{\mu}{2}\|X-(L+S)\|_F^2 \\ &= \arg\min_L \|L\|_* + \frac{\mu}{2}\left\|\left(X-S+\frac{Y}{\mu}\right)-L\right\|_F^2 \end{aligned} \tag{22}$$

denote q *as*

$$q=\frac{\mu}{2}\left\|\left(X-S+\frac{Y}{\mu}\right)-L\right\|_F^2 \tag{23}$$

then, we can get that

$$\frac{\partial q}{\partial L}=-\mu\left(X-S+\frac{Y}{\mu}-L\right) \tag{24}$$

and

$$\frac{\partial q}{\partial L}\left(L^k\right)=-\mu\left(X-S+\frac{Y}{\mu}-L^k\right) \tag{25}$$

so Eq.22 can be replaced as

$$\begin{aligned} L^{k+1} &= \arg\min_{L} \|L\|_{*} + \left\langle \frac{\partial q}{\partial L}\left(L^{k}\right), L - L^{k} \right\rangle + \frac{\eta}{2}\left\|L - L^{k}\right\|_{F}^{2} \\ &= \arg\min_{L} \|L\|_{*} + \frac{\eta}{2}\left\|L - L^{k} + \frac{1}{\eta}\cdot\frac{\partial q}{\partial L}\left(L^{k}\right)\right\|_{F}^{2} \\ &= \arg\min_{L} \frac{1}{\eta}\|L\|_{*} + \frac{1}{2}\left\|L - L^{k} + \frac{1}{\eta}\cdot\frac{\partial q}{\partial L}\left(L^{k}\right)\right\|_{F}^{2} \\ &= \arg\min_{L} \frac{1}{\eta}\|L\|_{*} + \frac{1}{2}\left\|L - \left(L^{k} - \frac{1}{\eta}\cdot\frac{\partial q}{\partial L}\left(L^{k}\right)\right)\right\|_{F}^{2} \end{aligned} \tag{26}$$

Remark 1.

Eq.22 → Eq.26, simple → complexed, so we prefer to Eq.22 rather than Eq.26 to update L.

2. *update S to be continued*
3. *update Y to be continued*

Example 4 (LRR)**.**

$$\min_{Z,E} \|Z\|_{*} + \lambda\|E\|_{1} \quad s.t.\ X = XZ + E \tag{27}$$

The Lagrangian function $\mathcal{L}(Z, E, Y)$ is as following:

$$\mathcal{L}(Z, E, Y) = \|Z\|_{*} + \lambda\|E\|_{1} + \langle Y, X - (XZ + E)\rangle + \frac{\mu}{2}\|X - (XZ + E)\|_{F}^{2} \tag{28}$$

1. *update Z*

$$\begin{aligned} Z^{k+1} &= \arg\min_{Z} \|Z\|_{*} + \langle Y, X - (XZ + E)\rangle + \frac{\mu}{2}\|X - (XZ + E)\|_{F}^{2} \\ &= \arg\min_{Z} \|Z\|_{*} + \frac{\mu}{2}\left\|X - (XZ + E) + \frac{Y}{\mu}\right\|_{F}^{2} \end{aligned} \tag{29}$$

now, we denote $q = \frac{\mu}{2}\left\|X - (XZ + E) + \frac{Y}{\mu}\right\|_{F}^{2} = \frac{\mu}{2}\left\|XZ + E - X - \frac{Y}{\mu}\right\|_{F}^{2}$, then we can get that

$$\frac{\partial q}{\partial Z} = \mu X^{T}\left(XZ + E - X - \frac{Y}{\mu}\right) \tag{30}$$

then

$$\frac{\partial q}{\partial Z}\left(Z^{k}\right) = \mu X^{T}\left(XZ^{k} + E - X - \frac{Y}{\mu}\right) \tag{31}$$

so Eq.29 can be replaced as

$$\begin{aligned} Z^{k+1} &= \arg\min_{Z} \|Z\|_{*} + \left\langle \frac{\partial q}{\partial Z}\left(Z^{k}\right), Z - Z^{k} \right\rangle + \frac{\eta}{2}\left\|Z - Z^{k}\right\|_{F}^{2} \\ &= \arg\min_{Z} \|Z\|_{*} + \frac{\eta}{2}\left\|Z - Z^{k} + \frac{1}{\eta}\cdot\frac{\partial q}{\partial Z}\left(Z^{k}\right)\right\|_{F}^{2} \\ &= \arg\min_{Z} \frac{1}{\eta}\|Z\|_{*} + \frac{1}{2}\left\|Z - \left(Z^{k} - \frac{1}{\eta}\cdot\frac{\partial q}{\partial Z}\left(Z^{k}\right)\right)\right\|_{F}^{2} \end{aligned} \tag{32}$$

2. *update E to be continued*

3. *update Y to be continued*

4 Optimization

Theorem 1. *The solution to*

$$\min_{X\in\mathbb{R}^{m\times n}} \frac{1}{2}||X-Y||_2^2+\tau||X||_* \tag{33}$$

is $\mathcal{D}_\tau(Y)$, obtained by soft-thresholding the singular values of $Y=U\Sigma V^t$

$$\mathcal{D}_\tau(Y)=U\mathcal{S}_\tau(\Sigma)V^t \tag{34}$$

$$\mathcal{S}_\tau(\Sigma)_{ii}=\begin{cases}\Sigma_{ii}-\tau & if \quad \Sigma_{ii}>\tau\\ 0 & otherwise\end{cases} \tag{35}$$

Proof. It will be proved as below. □

Theorem 2. *The solution to*

$$\min_X \frac{1}{2}\left\|X-A\right\|_F^2+\tau\|X\|_1 \tag{36}$$

is $\mathcal{S}_\tau\left(A\right)$, where

$$\mathcal{S}_\tau\left(a\right)=sign\left(a\right)\max\left(|a|-\tau,0\right)=\begin{cases}a-\tau & a>\tau\\ a+\tau & a<-\tau.\\ 0 & else\end{cases} \tag{37}$$

Proof. It will be proved as below. □

5 Proof of Theorem 1

Note : $||\cdot||$ is the operator norm, $||\cdot||_*$ is the nuclear norm.

Definition 1. *Let $\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_n$ be the singular values of $A\in\mathbb{R}^{m\times n},\ m\geq n$. The operator norm is*

$$||A||=\max_{||u||_2=1}||Au||_2 \tag{38}$$

where $u\in\mathbb{R}^n$.

Proposition 1.

$$||A||=\sigma_1 \tag{39}$$

Proof. Let $B = A^t A$, by linear algebra , B is a Hermitian matrix.
Then $\lambda_1, ..., \lambda_n \geq 0$ are the eigenvalues of B and $e_1, ..., e_n$ are the responding eigenvector of B, it's well known that $\{e_1, ..., e_n\}$ is an orthonormal basis of $\mathbb{R}^n$.
let

$$x = a_1 e_1 + \cdots + a_n e_n \tag{40}$$

$||x||_2^2 = a_1^2 + \cdots a_n^2$. then

$$\begin{aligned} ||Ax||_2^2 = \langle Ax, Ax \rangle &= \left\langle x, A^t Ax \right\rangle \\ &= \langle x, Bx \rangle \\ &= \left\langle \sum_{i=1}^{n} a_i e_i, B \sum_{i=1}^{n} a_i e_i \right\rangle \\ &= \left\langle \sum_{i=1}^{n} a_i e_i, \sum_{i=1}^{n} a_i \lambda_i e_i \right\rangle \\ &= \sum_{i=1}^{n} \lambda_i {a_i}^2 \\ &\leq \lambda_1 \sum_{i=1}^{n} {a_i}^2 \\ &= \lambda_1 ||x||_2^2 \end{aligned} \tag{41}$$

we can get

$$\begin{aligned} ||Ax||_2^2 \leq \lambda_1 ||x||_2^2 & \\ &\Leftrightarrow ||A \frac{x}{||x||_2^2}||_2^2 \leqslant \lambda_1 \\ &\Leftrightarrow \sup_{||u||_2=1} ||Au||_2^2 \leqslant \lambda_1 \\ &\Leftrightarrow = \sup_{||u||_2=1} ||Au||_2 \leqslant \sqrt{\lambda_1} = \sigma_1 \end{aligned} \tag{42}$$

sup can replace max because $|| \cdot ||$ is a continious function over a conpacet unit ball.
if $u = e_i$,

$$||Ae_1||_2^2 = \langle Ae_1, Ae_1 \rangle = \langle e_1, Be_1 \rangle = \langle e_1, \lambda_1 e_1 \rangle = \lambda_1 \Rightarrow ||A|| = ||Ae_1||_2 = \sqrt{\lambda_1} = \sigma_1 \tag{43}$$

so the proposition holds. □

Lemma 1. *For any $m \times n$ matrix A and B*

$$trace(AB) \quad = \quad trace(BA) \tag{44}$$

Lemma 2. *For any $Q \in \mathbb{R}^{m \times m}, V \in \mathbb{R}^{n \times n}$, if $U^t U = I$ and $V^t V = I$ then*

$$||UQV|| = ||Q|| \tag{45}$$

Lemma 3. *For any $Q \in \mathbb{R}^{n\times n}$*

$$\max_{1\le i\le n} |Q_{ii}| \le ||Q|| \tag{46}$$

Proof. We denote the standard basis vectors by $e_i, 1 \le i \le n$ in $\mathbb{R}^n$, then

$$Qe_i = \begin{bmatrix} Q_{11} & \cdots & Q_{1i} & \cdots & Q_{1n} \\ \vdots & & & & \\ Q_{i1} & \cdots & Q_{ii} & \cdots & Q_{in} \\ \vdots & & & & \\ Q_{n1} & \cdots & Q_{ni} & \cdots & Q_{nn} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} Q_{1i} \\ \vdots \\ Q_{ii} \\ \vdots \\ Q_{ni} \end{bmatrix} \tag{47}$$

and

$$\{e_i\} \subset \{x : ||x||_2 = 1\} \quad (\because x = \sum_{i=1}^{n} a_i e_i, \sum_{i=1}^{n} {a_i}^2 = 1) \tag{48}$$

so

$$\begin{aligned} \max_{1\le i\le n} |Q_{ii}| &\le \max_{1\le i\leqslant n} \sqrt{\sum_{j=1}^{n} {Q_{ji}}^2} \\ &= \max_{1\le i\le n} \langle Qe_i, Qe_i \rangle \\ &= \max_{1\leqslant i\leqslant n} ||Qe_i||_2 \\ &\le \sup_{||x||_2=1} ||Qx||_2 = ||Q|| \end{aligned} \tag{49}$$

□

Lemma 4. *If the singular value decomposition of A is $U\Sigma V^t$, then*

$$tr\left\langle A^t B\right\rangle = tr\left\langle V\Sigma^t U^t B\right\rangle = tr\left\langle \Sigma^t B\right\rangle \tag{50}$$

Definition 2. *The nuclear norm is equal to the l_1 norm of the singular values*

$$||A||_* = \sum_{i=1}^{n} \sigma_i \tag{51}$$

Proposition 2. *For any matrix $A \in \mathbb{R}^{m\times n}$,*

$$||A||_* = \sup_{||B||\le 1} \langle A, B\rangle \tag{52}$$

Proof. We recall that

$$||B|| = \sum {B_{ij}}^2 \le 1 \Rightarrow |B_{ij}| \le 1 \Rightarrow \mathop{max}_{1\le i\le n} |B_{ii}| \le 1 \quad (\text{otherwise} \;\; |B_{ij}| > 1 \Rightarrow ||B|| > 1)$$

$$\{M : \mathop{max}_{1\le i\le n} ||M_{ii}|| \le 1\}$$

Then,

$$\{B\} \subset \{M\} \tag{53}$$

$$\begin{aligned}
\sup_{||B||\leq 1} tr(A^t B) = \sup_{||B||\leq 1} tr(\Sigma^t B) \leq & \sup_{\{M: \max\limits_{1\leq i\leq n} ||M_{ii}||\leq 1\}} tr(\Sigma^t M) \\
= & \sup_{\{M: \max\limits_{1\leq i\leqslant n} ||M_{ii}||\leq 1\}} M_{ii}\sigma_i \\
\leq & \sum_{i=1}^{n} \sigma_i \\
= & ||A||^*
\end{aligned} \tag{54}$$

then,

$$\begin{aligned}
\langle A, UV^t \rangle &= tr(A^t UV^t) \\
&= tr(V\Sigma U^T UV^t) \\
&= tr(\Sigma) \\
&= ||A||_*
\end{aligned} \tag{55}$$

thus, $\sup\limits_{||x||\leq 1} ||UV^t x|| = 1$

so $Eq.52$ holds.

□

Proposition 3. *For any $m \times n$ matrices A and B*

$$||A + B||_* \leq ||A||_* + ||B||_* \tag{56}$$

Proof.

$$\begin{aligned}
||A + B||_* &= \sup_{||C||\leqslant 1} \langle A + B, C \rangle \\
&\leq \sup_{||C||\leqslant 1} \langle A, C \rangle + \sup_{||D||\leqslant 1} \langle B, D \rangle. \\
&= ||A||_* + ||B||_*
\end{aligned} \tag{57}$$

□

Proposition 4 (subgradients of the nuclear norm)**.** *Let $A = U\Sigma V^t$ be the singular value decomposition of A, Any matrix of the form*

$$\begin{aligned}
G : UV^t + W \quad & ||W|| \leq 1 \\
& U^t W = 0 \\
& WV = 0
\end{aligned} \tag{58}$$

is a subgradient of the nuclear norm of A.

Proof. By the Pythagorean theorem for inner product spaces, we have that

$$(\mathcal{P}_s x)^2 + (\mathcal{P}_{s^\perp} x)^2 = ||x||_2^2 \tag{59}$$

where $\mathcal{P}_s x$ denotes the vector x project to the space s, $s^\perp$ is the orthogonal space to s. Recall that the rows of UV^t are all in row(A), and the columns of U are all in column(A). by the definition the space $UV^t \perp$ the space W
then

$$\begin{aligned}
||G||^2 &= \sup_{||u||_2=1} ||Gu||_2^2 \\
&= \sup_{||u||_2=1} ||UV^t u||_2^2 + ||Wu||_2^2 \\
&= \sup_{||u||_2=1} ||UV^t(\mathcal{P}_{row(A)} + \mathcal{P}_{row(A)^\perp})u||_2^2 + \sup_{||u||_2=1} ||W(\mathcal{P}_{row(A)} + \mathcal{P}_{row(A)^\perp})u||_2^2 \\
&= \sup_{||u||_2=1} ||UV^t \mathcal{P}_{row(A)} u||_2^2 + \sup_{||u||_2=1} ||W\mathcal{P}_{row(A)^\perp} u||_2^2 \\
&= \sup_{||u||_2=1} ||UV^t \frac{\mathcal{P}_{row(A)} u}{||\mathcal{P}_{row(A)} u||_2} ||\mathcal{P}_{row(A)} u||_2||_2^2 + \sup_{||u||_2=1} ||W \frac{\mathcal{P}_{row(A)^\perp} u}{||\mathcal{P}_{row(A)^\perp} u||_2} ||\mathcal{P}_{row(A)^\perp} u||_2||_2^2 \\
&\le \sup_{||u_1||_2=1} ||UV^t u_1||^2 \cdot ||\mathcal{P}_{row(A)} u||_2^2 + \sup_{||u_2||_2=1} ||Wu_2||^2 \cdot ||\mathcal{P}_{row(A)^\perp} u||_2^2 \\
&= ||UV^t||^2 \cdot ||\mathcal{P}_{row(A)} u||_2^2 + ||W||^2 \cdot ||\mathcal{P}_{row(A)^\perp} u||_2^2 \\
&\le 1
\end{aligned} \tag{60}$$

then we can get $||G|| \le 1$;

$$\langle w, A \rangle = \left\langle w, U\Sigma V^t \right\rangle = \left\langle U^t w, \Sigma V^t \right\rangle = 0 \tag{61}$$

By the *proposition* 2, $\quad ||A||_* = \sup_{||B|| \le 1} \langle A, B \rangle$, for any matrix Y,

$$\begin{aligned}
||Y||_* &\geqslant \langle G, Y \rangle \\
&= \langle G, A \rangle + \langle G, Y - A \rangle \\
&= \left\langle UV^t + W, A \right\rangle + \langle G, Y - A \rangle \\
&= \left\langle UV^t, A \right\rangle + \langle G, Y - A \rangle
\end{aligned} \tag{62}$$

by $Eq.55$

$$\begin{aligned}
||Y||_* &\geqslant \langle G, Y \rangle \\
&= \left\langle UV^t, A \right\rangle + \langle G, Y - A \rangle \\
&= ||A||_* + \langle G, Y - A \rangle
\end{aligned} \tag{63}$$

Finally, G is the subgradients of the nuclear norm of A. □

Now we begin to prove the Theorem 1

Proof. The subgradients of $Eq.33$ are

$$X - Y + \tau G \tag{64}$$

where G is the subgradient of $||X||_*$.
If we can show that

$$\frac{1}{\tau}(Y - \mathcal{D}_\tau(Y)) \tag{65}$$

is a subgradient of the the nuclear norm at $\mathcal{D}_\tau(Y)$ the $X^* = \mathcal{D}_\tau(Y)$ is the solution of $Eq.33(\because \mathcal{D}_\tau(Y) - Y + \tau((\frac{1}{\tau}(Y - \mathcal{D}_\tau(Y)) = 0)$.

Let us separate the singular value decomposition of Y into the singular vectors corresponding to singular values greater that τ, denoted by U_0 and V_0, the rest as U_1 and V_1,

$$Y = U\Sigma V^t = [U_0\ U_1] \begin{bmatrix} \Sigma_0 & 0 \\ 0 & \Sigma_1 \end{bmatrix} [V_0\ V_1]^t \tag{66}$$

$\mathcal{D}_\tau(Y) = U_0(\Sigma_0 - \tau I)V_0^t$, so

$$\begin{aligned} \frac{1}{\tau}(Y - \mathcal{D}_\tau(Y)) &= \frac{1}{\tau}[U_0\Sigma_0V_0^t + U_1\Sigma_1V_1^t - (U_0(\Sigma_0 - \tau I)V_0^t)] \\ &= \frac{1}{\tau}[U_0\Sigma_0V_0^t + U_1\Sigma_1V_1^t - U_0\Sigma_0V_0^t + U_0\tau IV_0^t] \\ &= U_0V_0^t + \frac{1}{\tau}U_1\Sigma_1V_1^t \end{aligned} \tag{67}$$

$\because$ all the singular values of $U_1\Sigma_1V_1^t$ are smaller than τ,

$$||\frac{1}{\tau}U_1\Sigma_1V_1^t|| = ||\frac{1}{\tau}\Sigma_1|| \leq 1 \tag{68}$$

when $X = \mathcal{D}_\tau(Y) = U_0(\Sigma_0 - \tau I)V_0^t$, and by the definition of the SVD $U_0^tU_1 = 0\&V_0^t = 0$, so $Eq.67$ is a subgradient of the the nuclear norm at $X = \mathcal{D}_\tau(Y)$. □

6 Proof of Theorem 2

to be continued, may after the Spring Festival.

Happy Spring Festival 2019!