Lower Bound Theory for Schatten-p Regularized Least Squares Problem

Qingna Li[1]
Joint Work with Shiqian Ma, CUHK

December 22, 2013

[1]Beijing Institute of Technology, qnl@bit.edu.cn

Introduction

Lower Bound and Necessary Conditions

Smoothing Problem

Numerical Results

Conclusions

Schatten-p regularized least squares problem $(0 < p < 1)$

$$\min_{X \in \Re^{m \times n}} \ \|\mathcal{A}(X) - b\|_2^2 + \lambda \|X\|_p^p := f(X), \tag{1}$$

$\|X\|_p^p := \sum_{i=1}^{\min\{m,n\}} \sigma_i(X)^p$.
The set of local minimizers of (21) is denoted as $\mathcal{X}_p^*$.

Affine matrix rank minimization problem:

$$\begin{array}{ll} \min_{X\in\Re^{m\times n}} & rank(X) \\ s.t. & \mathcal{A}(X)=b, \end{array} \tag{2}$$

Nuclear norm convex relaxation:

$$\begin{array}{ll} \min_{X\in\Re^{m\times n}} & \|X\|_* \\ s.t. & \mathcal{A}(X)=b. \end{array} \tag{3}$$

- Singular Value Thresholding method (SVT): Cai, Candès and Shen, SIAMOPT, 2010
- Fixed Point Continuation method (FPC): Ma, Goldfarb and Chen, Math. Program., 2011
- Atomic Decomposition for Minimum Rank Approximation (ADMiRA): Lee and Bresler, IEEE Transactions on Information Theory, 2010
- Iterative Reweighted Least Squares method(IRLS): 2010

Schatten-p nonconvex relaxation:

- Interior Point Method: Ji, Sze, Zhou, So, Ye, INFOCOM 2013
- Hard thresholding method and fixed point method: Peng, Xiu and Yu, 2013
- Iterative Reweighted Singular Value Minimization method: Lu 2013

$l_2 - l_p$ vector minimization problem

$$\min_{x \in \Re^m} \|Ax - b\|_2^2 + \lambda \|x\|_p^p, \tag{4}$$

- Chen, Xu and Ye 2010
- Chen, Zhou 2013
- Chen, Niu and Yuan 2013
  ...
- $p = \frac{1}{2}$: Xu, Chang, Xu and Zhang, 2012
  Xu 2010
- Lu, 2013
- ...

Progress in Matrix Optimization

- Chao Ding, Defeng Sun, and Kim Chuan Toh, An introduction to a class of matrix cone programming, PDF version. To appear in Mathematical Programming
- Bin Wu, Chao Ding, Defeng Sun, and Kim Chuan Toh, On the Moreau-Yosida regularization of the vector k-norm related functions, March 2011
- Sun D.F., Matrix Cone Programming, Lecture Notes presented in Dalian University of Technology, June 2011.

Smoothing Methods:

- Smoothing Gradient Method: Chen, Xu and Ye 2010
- Smoothing Projected Gradient Method: Zhang and Chen, 2009
- Smoothing Trust Region Method: Chen, Niu and Yuan, 2013
- Smoothing Newton Method: Gao and Sun 2009, 2012, Qi and L, 2013

Lower Bound and Necessary Conditions

Let $X \in \Re^{m\times n}$ have singular value decomposition (SVD) as

$$X = U[Diag(\sigma)\ 0]V^T, \quad U \in \mathcal{O}_m,\ V \in \mathcal{O}_n, \tag{5}$$

Suppose $rank(X) = k$.

$$A_{U,V} := [\mathcal{A}(u_1 v_1^T), \cdots, \mathcal{A}(u_k v_k^T)] \in \Re^{q\times k} \tag{6}$$

where u_i, v_i are the i-th column vectors of U and V respectively.

Lower Bound Theory

Theorem 2.1

- (The second order bound) Let $L := \left(\frac{\lambda p(1-p)}{2\sum_{i=1}^{q}\|A_i\|^2}\right)^{\frac{1}{2-p}}$. Then for any $\overline{X} \in \mathcal{X}_p^*$ with rank k, let $A_{\overline{U},\overline{V}}$ be defined as in (6).

$$\text{For any } i \in \mathcal{N}, \bar{\sigma}_i \in [0, L) \Rightarrow \bar{\sigma}_i = 0. \tag{7}$$

- (The first order bound) Let $\overline{X} \in \mathcal{X}_p^*$ satisfying $f(\overline{X}) \leq f(X^0)$ for an arbitrarily given initial point X^0. Let $L_0 := \left(\frac{\lambda p}{2\|\mathcal{A}\|\sqrt{f(X^0)}}\right)^{\frac{1}{1-p}}$. Then

$$\textit{for any } i \in \mathcal{N}, \quad \bar{\sigma}_i \in [0, L_0) \Rightarrow \bar{\sigma}_i = 0.$$

Some Lemmas

Lemma 2.1(Characterization of Subgradients) Suppose that the function $f : \Re^m \to (-\infty, +\infty]$ is absolutely symmetric, and that the $m \times n$ matrix T has the singular value vector $\sigma(T)$ in $dom(f)$. Then the $m \times n$ matrix H lies in $\partial(f \circ \sigma)(T)$ if and only if $\sigma(H)$ lies in $\partial f(\sigma(T))$ and there exists a simultaneous singular value decomposition form

$$T = U[\mathrm{Diag}(\sigma(T))\ 0]V^T, \ H = U[\mathrm{Diag}(\sigma(H))\ 0]V^T,$$

where $U \in \mathcal{U}_m$, $V \in \mathcal{U}_n$ are unitary matrices. In fact,

$$\begin{aligned} \partial(f \circ \sigma)(T) \quad = \quad & \{U[\mathrm{Diag}(\mu)\ 0]V^T \mid \mu \in \partial f(\sigma(T)), U \in \mathcal{U}_m, \\ & V \in \mathcal{U}_n, \ T = U[\mathrm{Diag}(\sigma(T))\ 0]V^T\}. \end{aligned}$$

Some Lemmas

Let $\theta : \Re^+ \to \Re$ be a scalar function.
The corresponding non-symmetric Löwner operator is defined by

$$\Theta(X) := U[\mathrm{diag}(\theta(\sigma))\ 0]V^T. \tag{8}$$

Lemma 2.2 The non-symmetric Löwner operator Θ is well-defined if and only if $\theta(0) = 0$.

Suppose θ is differentiable at σ_i, $i = 1, \cdots, m$.
Define $\Gamma_1 \in \Re^{m\times m}$, $\Gamma_2 \in \Re^{m\times m}$, $\Gamma_3 \in \Re^{m\times(n-m)}$ as

$$(\Gamma_1)_{ij} = \begin{cases} \frac{\theta(\sigma_i)-\theta(\sigma_j)}{\sigma_i-\sigma_j}, & \text{if } \sigma_i \neq \sigma_j, \\ \theta'(\sigma_i), & \text{otherwise,} \end{cases} \qquad i,j \in \{1, \cdots, m\}. \tag{9}$$

$$(\Gamma_2)_{ij} = \begin{cases} \frac{\theta(\sigma_i)+\theta(\sigma_j)}{\sigma_i+\sigma_j}, & \text{if } \sigma_i + \sigma_j \neq 0, \\ \theta'(0), & \text{otherwise,} \end{cases} \qquad i,j \in \{1, \cdots, m\}. \tag{10}$$

$$(\Gamma_3)_{ij} = \begin{cases} \frac{\theta(\sigma_i)}{\sigma_i}, & \text{if } \sigma_i \neq 0, \\ \theta'(0), & \text{if } \sigma_i = 0, \end{cases} \qquad i \in \{1, \cdots, m\},\ j \in \{1, \cdots, n-m\}. \tag{11}$$

Some Lemmas

Lemma 2.3 The non-symmetric Löwner operator Θ is continuously differentiable at X if and only if θ is continuously differentiable at $\sigma_i(X)$, $i = 1, \cdots, m$. Moreover, the derivative $\Theta'(X)$, for any $H \in \Re^{m\times n}$ is given by

$$\Theta'(X)H = U[\Gamma_1 \circ S(A) + \Gamma_2 \circ T(A)\ \Gamma_3 \circ B]V^T, \qquad (12)$$

where $A = U^T H V_1$, $B = U^T H V_2$, $S(A) = \frac{1}{2}(A + A^T)$, $T(A) = \frac{1}{2}(A - A^T)$, $V = [V_1\ V_2]$, $V_1 \in \Re^{m\times m}$, $V_2 \in \Re^{m\times(n-m)}$.

First Order Necessary Condition for (1)

Proposition 2.1 We say that X satisfies the first order necessary condition for (1) if

$$0 = 2Diag(\sigma)U^T\mathcal{A}^*(\mathcal{A}(X)-b)V + \lambda p[\mathrm{Diag}(\sigma^p)\ \ 0]. \tag{13}$$

Recall: x is said to satisfy the first order necessary condition of (4) if

$$2\mathrm{Diag}(x)A^T(Ax-b) + \lambda p|x|^p = 0.$$

Define $\Omega_1 \in \Re^{m\times m}$, $\Omega_2 \in \Re^{m\times m}$, $\Omega_3 \in \Re^{m\times(n-m)}$ as

$$(\Omega_1)_{ij} = \begin{cases} \frac{p}{2}\sigma_i^p, & \sigma_i = \sigma_j; \\ \sigma_i^2 \frac{\sigma_i^p - \sigma_j^p}{\sigma_i^2 - \sigma_j^2}, & \sigma_i \neq \sigma_j, \end{cases} \tag{14}$$

$$(\Omega_2)_{ij} = \begin{cases} \frac{p-1}{2}\sigma_i^p, & \sigma_i = \sigma_j; \\ \frac{\sigma_i^p \sigma_j^2 - \sigma_j^p \sigma_i^2}{\sigma_i^2 - \sigma_j^2}, & \sigma_i \neq \sigma_j, \end{cases} \tag{15}$$

$$(\Omega_3)_{ij} = \begin{cases} \sigma_i^p, & \sigma_i > 0; \\ 0, & \sigma_i = 0. \end{cases} \tag{16}$$

Second Order Necessary Condition for (1)

Proposition 2.2 We say $X \in \Re^{m\times n}$ with SVD as in (5) satisfies the second order necessary condition for (1) if the following holds:

$$2\langle Diag(\sigma)Z, \overline{\mathcal{A}}^*\overline{\mathcal{A}}(Diag(\sigma)Z)\rangle + \lambda p\langle Z, [\Omega_1 \circ Z_1 + \Omega_2 \circ Z_1^T \;\; \Omega_3 \circ Z_2]\rangle \geq 0 \tag{17}$$

for any $Z = [Z_1 \;\; Z_2] \in \Re^{m\times n}$, where
$\overline{\mathcal{A}} := \mathcal{A}_\mu(H) := [\langle U_\mu^T A_1 V_\mu, H\rangle, \cdots, \langle U_\mu^T A_s V_\mu, H\rangle]^T \in \Re^s$
$\Omega_1, \; \Omega_2, \; \Omega_3$ are defined as in (14), (15) and (16).

Recall: x is said to satisfy the second order necessary condition of (4) if

$$2\mathrm{Diag}(x)A^T A\mathrm{Diag}(x) + \lambda p(p-1)\mathrm{Diag}(|x|^p) \succeq 0.$$

Lower Bound Theory and Necessary Conditions for Smoothing Problem

Smoothing Function

The smoothing function for $|t| (t \in \Re)$:

$$s_\mu(t) = \begin{cases} |t|, & |t| \geq \mu \\ \frac{t^2}{2\mu} + \frac{\mu}{2}, & |t| < \mu. \end{cases} \tag{18}$$

where $\mu > 0$. $s_\mu(t)$ is continuously differentiable.

$$0 \leq s_\mu(t)^p - |t|^p \leq \frac{\mu}{2}.$$

When $\mu \to 0$, $s_\mu(t)^p \to |t|^p$.

Smoothing Function

$$p \cdot \theta(t) := (s_\mu(t)^p)' = \begin{cases} p|t|^{p-1} sign(t) := p \cdot \theta_1(t), & |t| \geq \mu \\ p(\frac{t^2}{2\mu} + \frac{\mu}{2})^{p-1} \frac{t}{\mu} := p \cdot \theta_2(t), & |t| < \mu. \end{cases} \tag{19}$$

$s'(t)$ is not differentiable at $t = \pm\mu$, so is θ.

$$\theta'(t) = \begin{cases} (p-1)|t|^{p-2} := \theta_1'(t), & |t| > \mu \\ (p-1)(\frac{t^2}{2\mu} + \frac{\mu}{2})^{p-2} \frac{t^2}{\mu^2} + (\frac{t^2}{2\mu} + \frac{\mu}{2})^{p-1} \frac{1}{\mu} := \theta_2'(t), & |t| < \mu. \end{cases} \tag{20}$$

At μ, the generalized gradient of $\theta(t)$ is

$$\partial(\theta(\mu)) = \{v \in \Re : (p-1)\mu^{p-2} \leq v \leq p\mu^{p-2}\}.$$

Smoothing Problem

The smoothing problem

$$\min_{X \in \Re^{m \times n}} \ f_\mu(X) := \|\mathcal{A}(X) - b\|_2^2 + \lambda \|S_\mu(X)\|_p^p. \qquad (21)$$

where

$$S_\mu(X) = U[Diag(s_\mu(\sigma)) \ 0]V.$$

The set of local minimizers of (21) is denoted as $\mathcal{X}^*_{p,\mu}$.

Lower Bound Theory for Smoothing Problem

Theorem 3.1 Let $L := \left(\frac{\lambda p(1-p)}{2\sum_{i=1}^{q}\|A_i\|^2}\right)^{\frac{1}{2-p}}$ and $L_0 = \left(\frac{\lambda p}{2\|\mathcal{A}\|\sqrt{f(X^0)}}\right)^{\frac{1}{1-p}}$ for an arbitrarily given initial point X^0.

(i) (The second order bound) For any $\mu > 0$, and any $\overline{X}_\mu \in \mathcal{X}^*_{p,\mu}$, we have

$$\forall\ i \in \mathcal{N},\quad (\bar{\sigma}_\mu)_i \in [0, L) \Rightarrow (\bar{\sigma}_\mu)_i \in [0, \mu].$$

(ii) (The first order bound) For any $\mu > 0$ and any $\overline{X}_\mu \in \mathcal{X}^*_{p,\mu}$ satisfying $f(\overline{X}_\mu) \le f(X^0)$, we have

$$\forall\ i \in \mathcal{N},\quad (\bar{\sigma}_\mu)_i \in [0, L_0) \Rightarrow (\bar{\sigma}_\mu)_i \in [0, \mu].$$

First and Second Order Necessary Condition for (21)

Proposition 3.1
(The first order necessary condition for (21))
We say X_μ satisfies the first order necessary condition if

$$2\mathcal{A}^*(\mathcal{A}(X_\mu) - b) + \lambda U_\mu[Diag(\Psi(\sigma_\mu)) \ 0]V_\mu^T = 0 \qquad (22)$$

where

$$\Psi_\mu(x) = ((s_\mu(x_1)^p)', \cdots, (s_\mu(x_m)^p)')^T \in \Re^m.$$

Proposition 3.2 (The second order necessary condition for (21))
For X_μ, suppose $(\sigma_\mu)_i \neq \mu$, $i = 1, \cdots, m$. We say X_μ satisfies the second order necessary condition if

$$\langle H, 2\mathcal{A}^*\mathcal{A}(H) + \lambda p\Theta'(X_\mu)H\rangle \geq 0, \quad \text{for any } H \in \Re^{m\times n}, \qquad (23)$$

where

$$\Theta'(X_\mu)H = U[\Gamma_1 \circ S(A) + \Gamma_2 \circ T(A) \ \Gamma_3 \circ B]V^T.$$

First and Second Order Necessary Condition for (21)

Proposition 3.3 Let $\theta(\cdot)$ be defined as in (19). We say that X_μ satisfies the second order necessary condition if

$$\langle H, 2\mathcal{A}^*\mathcal{A}(H) + \lambda p MH\rangle \geq 0, \;\; M \in \partial\Theta(X) \;\; \text{for any } H \in \Re^{m\times n}, \tag{24}$$

where the generalized Jacobian of $\Theta(\cdot)$ at X is given by

$$\begin{aligned} \partial\Theta(X)H &= \{MH = U[\overline{\Gamma}_1 \circ S(A) + \Gamma_2 \circ T(A) \; \Gamma_3 \circ B]V^T, \\ & \qquad (\overline{\Gamma}_1)_{ij} = \begin{cases} \frac{\theta(\sigma_i)-\theta(\sigma_j)}{\sigma_i-\sigma_j}, & \text{if } \sigma_i \neq \sigma_j, \\ \theta'(\sigma_i), & \sigma_i = \sigma_j \neq \mu, \\ \gamma_{ij} \in \partial\theta(\mu), & \sigma_i = \sigma_j = \mu. \end{cases} \;\; i,j \in \{1,\cdots,m\}.\} \end{aligned}$$

where $A = U^T HV_1$, $B = U^T HV_2$, $V = [V_1 \;\; V_2]$, Γ_2, Γ_3 are defined as in (10), (11), and

Convergence of Smoothing Algorithm

Theorem 3.2

- (1) Let $\{X_{\mu_k}\}$ be a sequence of matrices satisfying the first order necessary condition of (21). Then any accumulation point of $\{X_{\mu_k}\}$ satisfies the first order necessary condition of (1).
- (2) Let $\{X_{\mu_k}\}$ be a sequence of matrices satisfying the second order necessary condition of (21). Then any accumulation point of $\{X_{\mu_k}\}$ satisfies the second order necessary condition of (1).
- (3) Let $\{X_{\mu_k}\}$ be a sequence of matrices being global minimizer of (21). Then any accumulation point of $\{X_{\mu_k}\}$ is a global minimizer of (1).

Convergence of Smoothing Algorithm

Theorem 3.3 Let $\{X_{\mu_k}\}$ be a sequence of matrices satisfying the first order necessary conditions of (21) and $f(X_{\mu_k}) \leq f(X^0)$ for an arbitrary given initial point X^0. Suppose X_{μ_k} has SDV as in (5). Then there is a $K > 0$ such that, for any $k \geq K$, there is $\overline{X} \in \mathcal{X}_p$ with SVD as in (5) such that

$$I_{\mu_k} := \{i \in \mathcal{N} \mid (\sigma_{\mu_k})_i \leq \mu_k\} = \{i \in \mathcal{N} \mid \bar{\sigma}_i = 0\} := I \qquad (25)$$

Algorithm 1

Smoothing Gradient Method(SG) for (1)

Step 1 Choose a starting point $X^0 \in \Re^{m\times n}$. Calculate L_0 by Theorem 2.1.

Step 2 Start from the initial point X^0, solve (21) using smoothing gradient algorithm to get X_μ.

Step 2.1 Choose constants η, $\rho \in (0,1)$, $tol > 0$, and an initial point X^0. $k := 0$.

Step 2.2 Compute the step size ν_k by the Armijo line search where $\nu_k = \max\{\rho^0, \rho^1, \cdots\}$ and ρ^i satisfies

$$f_{\mu_k}(X^k - \rho g_k) \le f_{\mu_k}(X^k) - \eta\rho^i \|g_k\|^2.$$

Set $X^{k+1} = X^k - \nu_k g_k$. Here $g_k = \nabla f_{\mu_k}(X^k)$.

Step 2.3 If $\|g_k\| \le tol$, stop, go to Step 3; otherwise, update μ_k such that $\mu_{k+1} \le \mu_k$, go to Step 2.2.

Step 3 Output $X_\mu^* = U_\mu[\mathrm{Diag}(\sigma_\mu^*) \;\; 0]V_\mu^T$, where σ_μ^* is defined by

$$(\sigma_\mu^*)_i = \begin{cases} (\sigma_\mu)_i, & (\sigma_\mu)_i \ge L_0, \\ 0, & \text{otherwise.} \end{cases}$$

U_μ and V_μ comes from the SVD of $X_\mu = U_\mu[\mathrm{Diag}(\sigma_\mu) \;\; 0]V_\mu^T$.

Numerical Results

Matrix Completion Problem:

$$\min_{X \in \Re^{m \times n}} \|P_\Omega(X) - b\|_2^2 + \lambda \|X\|_p^p, \tag{26}$$

$\lambda = 0.01,\ \mu_0 = 0.1,\ \eta = 10^{-4},\ \rho = 0.5,\ tol = \min(0.01,\ \max(10^{-4}, 10^{-3}\|g_0\|))$.
μ_k is updated by

$$\mu_{k+1} = \begin{cases} 0.96\mu_k, & \text{if } mod(k,2) = 0 \\ \mu_k, & otherwise. \end{cases}$$

Stopping Criteria: $\|g_k\| \leq tol$ or $\frac{\|X^{k+1} - X^k\|}{\|X^k\|} \leq 10^{-5}$.

- SG: Smoothing Gradient method
- SVT: Singular Value Thresholding method ,Cai, Candès and Shen, SIAMOPT, 2010
- FPC: Fixed Point Continuation method, Ma, Goldfarb and Chen, Math. Program., 2011
- HFPA: half norm fixed point algorithm, Peng, Xiu and Yu, 2013

$MSE := \frac{\|X_\mu^* - X^*\|}{\|X^*\|}$.

An instance is said to be successful if $rank(X_\mu^*) = rank(X^*)$ and its corresponding $MSE \leq 10^{-3}$.

The Role of Lower Bound Theory

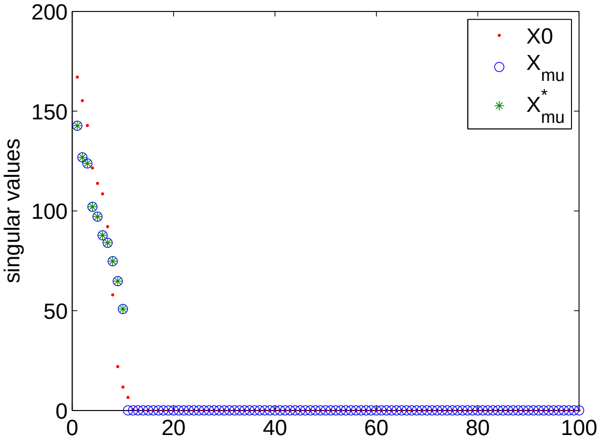

Figure: $m = n = 100,\ r = 10,\ OS = 5,\ L_0 = 8.493e-002,$
$MSE = 3.97e-005$

Comparison of Four Algorithms

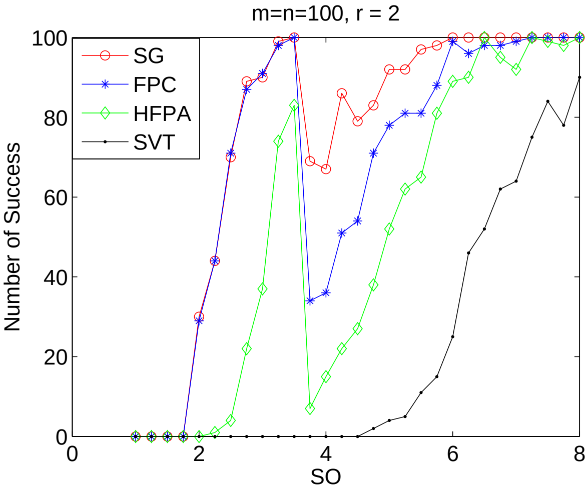

Figure: $m = n = 100,\ r = 2,\ \overline{OS} \approx 3.5$

Small Easy Problems

		SG			FPC			SVT			HFPA		
r	OS	NS	MSE	t	NS	MSE	t	NS	MSE	t	NS	MSE	t
2	4	7	1.18E-3	2.3	5	1.58E-3	0.2	0	2.29E-1	38.3*	0	2.71E-3	0.2
2	4.5	8	7.56E-3	1.9	2	1.27E-3	0.2	0	1.11E-1	30.6*	2	1.01E-2	0.2
2	5	10	3.15E-4	1.6	6	9.37E-4	0.1	0	8.65E-2	22.9	6	1.19E-3	0.1
2	6	10	1.06E-4	1.2	10	3.33E-4	0.1	5	1.85E-2	10.2	10	5.05E-4	0.1
2	7	10	6.95E-5	1.3	10	2.17E-4	0.1	7	9.60E-3	4.3	10	2.94E-4	0.1
2	8	10	7.71E-5	0.8	10	1.02E-4	0.1	9	3.45E-3	2.7	10	1.32E-4	0.1
5	3	10	1.16E-4	1.4	10	3.25E-4	0.2	0	6.67E-2	48.8	10	5.17E-4	0.2
5	4	10	8.51E-5	0.7	10	9.35E-5	0.2	4	1.03E-3	9.1	9	4.64E-3	0.1
5	5	10	5.53E-5	0.6	10	3.63E-5	0.3	10	1.72E-4	2.9	10	3.99E-5	0.1
5	6	10	3.15E-5	0.6	10	1.93E-5	0.3	10	1.55E-4	1.9	10	7.43E-6	0.1
5	8	10	1.88E-5	0.8	10	4.84E-5	0.9	10	1.27E-4	1.6	10	4.09E-7	0.1
10	3	10	7.82E-5	0.6	10	4.38E-5	0.4	1	3.74E-4	11.8	10	4.19E-5	0.1
10	3.5	10	4.24E-5	0.8	10	4.87E-5	0.9	10	1.71E-4	4.1	10	9.59E-6	0.1
10	4	10	4.16E-5	1.5	0	5.62E-3	2.5	10	1.53E-4	2.3	10	2.51E-6	0.1
10	4.5	10	6.69E-5	1.5	0	8.61E-2	2.5	10	1.35E-4	1.9	10	8.56E-7	0.1
10	5	10	4.19E-5	1.4	0	7.50E-1	2.5	10	1.15E-4	1.5	10	6.27E-7	0.1

Table: Comparison with FPC, SVT and HFPA on easy problems: $m = n = 100$. $*$ means that the algorithm reaches the maximum number of iteration.

Small Hard Problems

		SG			FPC			SVT			HFPA		
r	OS	NS	MSE	t	NS	MSE	t	NS	MSE	t	NS	MSE	t
2	2	2	4.70E-2	4.2	2	4.50E-2	2.7	0	1.32E+5	5.7	0	7.37E-2	1
2	2.25	3	1.40E-2	3.8	4	1.59E-2	2.5	0	6.05E+4	25.5	0	3.03E-2	1
2	2.5	7	1.47E-2	3.7	7	1.47E-2	2.4	0	3.33E+4	31.3	1	2.70E-2	0
2	2.75	7	6.76E-3	3.2	6	6.80E-3	2.0	0	5.35E-1	*	3	1.05E-2	0
2	3	9	4.00E-4	3.3	9	2.63E-4	2.1	0	4.97E-1	*	2	3.54E-3	0
2	3.25	9	4.25E-4	3.3	9	3.47E-4	2.0	0	4.04E-1	*	8	3.58E-3	0
2	3.5	10	1.89E-4	3.4	10	1.49E-5	2.1	0	3.32E-1	*	7	8.42E-4	0
5	1	0	9.89E-1	8.0	0	1.04E+0	3.7	0	1.40E+5	1.8	0	9.47E-1	1
5	1.25	0	8.49E-2	9.14	0	5.27E-2	3.8	0	1.00E+4	55.6	0	8.25E-1	2
5	1.5	4	1.66E-2	6.2	4	1.59E-2	3.7	0	8.73E-1	*	1	2.07E-2	1
5	1.75	9	3.53E-3	4.7	8	3.48E-3	3.5	0	6.72E-1	*	5	5.42E-3	1
5	2	10	1.95E-4	4.2	10	1.22E-5	3.0	0	5.49E-1	*	9	5.65E-4	1
5	2.25	10	1.46E-4	11.4	10	1.02E-5	8.6	0	4.31E-1	*	10	4.24E-4	2
5	2.5	10	1.09E-4	12.1	10	6.61E-6	9.5	0	2.79E-1	*	10	3.00E-4	2
10	1	0	7.05E-1	9.3	0	7.26E-1	5.8	0	9.58E-1	*	0	8.00E-1	2
10	1.25	10	3.30E-4	6.1	9	6.23E-5	5.6	0	8.47E-1	*	8	8.96E-4	2
10	1.5	10	2.31E-4	5.9	10	1.49E-5	5.2	0	6.27E-1	*	9	6.90E-4	1
10	1.75	10	1.82E-4	6.5	10	5.16E-6	6.0	0	4.16E-1	*	10	2.37E-4	1
10	2	10	1.63E-4	7.3	10	4.03E-6	6.9	0	2.29E-1	*	10	1.57E-4	1
10	2.25	10	1.29E-4	8.7	10	2.79E-6	8.3	0	8.43E-2	*	10	1.04E-4	1
10	2.5	10	1.10E-4	9.5	10	2.12E-6	9.3	0	1.72E-2	*	10	7.60E-5	2

Table: Comparison with FPC, SVT and HFPA on hard problems: $m = n = 100$. $*$ means that the algorithm reaches the maximum number of iteration.

Higher Dimension Problems

		SG			FPC			HFPA		
r	OS	NS	MSE	t	NS	MSE	t	NS	MSE	t
20	2.5	10	6.51E-04	17.6	9	6.48E-04	14.72	9	6.67E-04	13.36
20	2.75	10	6.01E-04	15.4	9	6.00E-04	12.52	9	6.16E-04	11.12
20	3	10	5.65E-04	15.03	10	5.64E-04	11.8	10	5.72E-04	10.4
20	3.5*	10	4.76E-04	13.14	10	4.98E-04	10.19	9	1.21E-03	8.76
20	4	10	3.83E-04	12.81	10	3.85E-04	9.69	10	3.49E-04	7.12
20	5	10	1.94E-04	13.27	10	1.87E-04	9.97	10	1.46E-04	6.81
20	7	10	8.54E-05	20.62	10	5.15E-05	16.67	10	2.17E-05	9.68
30	1.75	10	6.88E-04	32.69	10	6.86E-04	29.66	8	7.05E-04	28.4
30	2	10	5.84E-04	23.71	10	5.83E-04	20.74	10	5.89E-04	19.33
30	2.25	10	5.26E-04	19.76	10	5.18E-04	16.67	9	5.33E-04	15.03
30	2.5	10	4.75E-04	17.78	10	4.78E-04	14.69	10	4.75E-04	12.92
30	2.75	10	4.32E-04	16.5	10	4.29E-04	13.49	10	4.36E-04	11.62
30	3*	10	3.90E-04	16.06	10	3.89E-04	12.96	10	3.78E-04	10.28
30	4	10	1.65E-04	17.1	10	1.58E-04	13.45	10	1.25E-04	9.1
30	5	10	9.20E-05	22.49	10	6.88E-05	17.78	10	3.01E-05	11.04
30	6	10	6.33E-05	32.94	10	2.15E-05	28.01	10	6.09E-06	16.14

Table: Comparison with FPC and HFPA on high dimension problems: $m = n = 1000$. $*$ is the number approximately equal to $\overline{OS}(m, n, r)$

Higher Dimension Problems

		SG			FPC			HFPA		
r	OS	NS	MSE	t	NS	MSE	t	NS	MSE	t
10	5	5	9.91E-04	61.15	4	1.03E-03	44.65	2	1.03E-03	34.4
10	5.5	8	9.50E-04	57.54	8	9.45E-04	41.27	7	9.71E-04	30.7
10	6	9	8.90E-04	54.44	8	8.91E-04	37.87	10	8.97E-04	27.2
10	6.5	9	8.11E-04	53.42	10	8.38E-04	36.93	8	8.45E-04	26.0
10	7*	10	7.71E-04	51.53	9	7.98E-04	35.26	10	8.33E-04	24.4
10	8	10	6.95E-04	50.09	10	7.13E-04	33.45	9	1.57E-03	23.5
10	10	10	4.15E-04	48.73	10	4.38E-04	31.96	10	4.37E-04	19.0
10	11	10	3.48E-04	48.12	10	3.63E-04	31.05	10	3.30E-04	18.0
30	2.5	10	7.45E-04	92.53	10	7.53E-04	75.75	8	7.62E-04	65.0
30	3	10	6.42E-04	76.65	9	6.54E-04	60.65	10	6.67E-04	48.7
30	3.25	10	6.02E-04	72.64	8	6.15E-04	55.79	10	6.29E-04	44.1
30	4*	10	5.07E-04	66.9	10	5.20E-04	49.55	10	5.26E-04	37.3
30	5	10	2.82E-04	67.96	10	3.00E-04	50.27	10	2.90E-04	33.0
30	6	10	1.45E-04	73.23	10	1.52E-04	53.01	10	1.42E-04	32.8
30	8	10	4.98E-05	108.5	10	4.81E-05	83.82	10	3.20E-05	45.2
30	9	10	3.50E-05	159.91	10	2.80E-05	133.3	10	1.16E-05	69.7

Table: Comparison with FPC and HFPA on high dimension problems: $m = n = 2000$. $*$ is the number approximately equal to $\overline{OS}(m, n, r)$

Conclusions

- Lower bound theory for (1) and smoothing problem (21)
- First and second order necessary condition for (1) and smoothing problem (21)
- Smoothing gradient method for (1), using the developed lower bound result.

Future work

- Algorithm using second order information
- Optimality condition for other Shatten-p-like regularization problems

Thank You!