A smoothing majorization method for l_2^2-l_p^p matrix minimization

Liwei Zhang

Dalian University of Technology

(A Joint Work with Yue Lu and Jia Wu)

2014 Workshop on Optimization for Modern Computation,BICMR,Peking University

September 2-4, 2014

Contents

1 Introduction

2 Lower bound analysis

3 The smoothing model

4 The majorization algorithm

5 Numerical experiments

Background

The aim of the matrix rank minimization problem is to find a matrix with minimum rank that satisfies a given convex constraint, i.e.,

$$\begin{array}{ll} \min & \mathrm{rank}(X) \\ \text{s.t.} & X \in \mathcal{C}, \end{array} \tag{1}$$

where $\mathcal{C}$ is a nonempty closed convex subset of $\mathfrak{R}^{m\times n}$ and $\mathfrak{R}^{m\times n}$ represents the space of $m \times n$ matrices.

Without loss of generality, we assume $m \leq n$ throughout this paper. For solving (1), Fazel et al. [13, 14] suggested using the matrix nuclear norm to approximate the rank function and proposed the following convex optimization problem

$$\begin{array}{ll} \min & \|X\|_* \\ \text{s.t.} & X \in \mathcal{C}, \end{array} \tag{2}$$

where $\|X\|_* := \sum_{i=1}^{m} \sigma_i(X)$, $\sigma_i(X)$ denotes the ith largest singular value of X.

Many important problems can be formulated as (2). For example, several authors have used (2) to solve the famous matrix completion problem with the following model

$$\begin{array}{ll} \min & \|X\|_* \\ \text{s.t.} & X_{ij} = M_{ij},\ (i,j) \in \Omega, \end{array} \tag{3}$$

where Ω is an index set of the entries of M.

- the singular value thresholding algorithm [5],
- the fixed-point continuation algorithm [23],
- the alternating-direction-type algorithm [15].

Recently, these methods have also been applied to the nuclear norm regularized linear least square problem

$$\min_{X \in \Re^{m\times n}} \left\{ \frac{1}{2}\|\mathcal{A}(X) - b\|_2^2 + \tau\|X\|_* \right\}, \tag{4}$$

where $\mathcal{A}$ is a linear operator from $\Re^{m\times n}$ to $\Re^q$. It is worthwhile to note that (4) is regarded as a convex approximation to the regularized version of the affine rank minimization problem

$$\min_{X \in \Re^{m\times n}} \left\{ \frac{1}{2}\|\mathcal{A}(X) - b\|_2^2 + \tau \cdot \text{rank}(X) \right\}. \tag{5}$$

The l_2^2-l_p^p model

We consider another approximation to (5), which uses the following l_2^2-l_p^p model

$$\min_{X \in \Re^{m \times n}} \left\{ F(X) := \frac{1}{2}\|\mathcal{A}(X) - b\|_2^2 + \frac{\tau}{p}\|X\|_p^p \right\}, \qquad (6)$$

where $\|X\|_p^p := \sum_{i=1}^{r} \sigma_i^p(X), r := \text{rank}(X), p \in (0,1)$ and $b \in \Re^q$.

Vector model

$$\min_{x\in\Re^m} \left\{\frac{1}{2}\|Cx-b\|_2^2+\frac{\tau}{p}\|x\|_p^p\right\}. \tag{7}$$

- X. J. Chen, F. Xu, and Y. Y. Ye, *Lower bound theory of nonzero entries in solutions of l_2-l_p minimization*, SIAM J. Sci. Comput., 32 (2011), pp. 2832–2852.
- X. J. Chen, *Smoothing methods for nonsmooth, nonconvex minimization*, Math. Program., 134 (2012), pp. 71–99.
- X. J. Chen, D. D. Ge, Z. Z. Wang and Y. Y. Ye, *Complexity of unconstrained l_2-l_p minimization*, Math. Program., 143 (2014), pp. 371–383.

On vector l_2^2-l_p^p problem

- Chen, Xu and Ye (2011) [10] gave a lower bound estimate of nonzero entries in solutions of (7).
- Chen (2012)[9] introduced the smoothing technique to tackle the term $\|x\|_p^p$ and proposed an SQP-type algorithm to solve (7).
- Chen, Ge, Wang and Ye (2014) [11] studied the complexity of (7) and proved that the vector l_2^2-l_p^p problem (7) is strongly NP-Hard.

The purpose of the work

- To check whether we can develop the parallel lower bound analysis in Chen, Xu and Ye (2011) [10] for the matrix l_2^2-l_p^p problem.
- To develop an numerical method for solving an approximate solution to the matrix l_2^2-l_p^p problem.

Features of the proposed method

- We present a smoothing majorization method in which the smoothing parameter ϵ is treated as a decision variable and introduce an automatic update mechanism of the smoothing parameter ϵ.
- The unconstrained subproblems based on the majorization functions are solved inexactly and the corresponding optimal solutions can be obtained explicitly.
- Numerical experiments show that our method is insensitive to the choice of the parameter p.

Notations and definitions

- Given any $X, Y \in \Re^{m\times n}$, $\langle X, Y\rangle := \mathrm{Tr}(X^TY)$ and the Frobenius norm of X is denoted by $\|X\|_F := \sqrt{\mathrm{Tr}(XX^T)}$.
- Given any vector $x \in \Re^m$, let $x^\beta := (x_1^\beta, x_2^\beta, \cdots, x_m^\beta)^T$. For $X \in \Re^{m\times m}$, $\mathrm{Diag}(X) := (X_{11}, X_{22}, \cdots, X_{mm})^T$.
- Given an index set $\mathcal{I} \subseteq \{1, 2, \cdots, m\}$, $x_{\mathcal{I}}$ denotes the sub-vector of x indexed by $\mathcal{I}$. Similarly, $X_{\mathcal{I}}$ denotes the sub-matrix of X whose columns are indexed by $\mathcal{I}$. Denote the index $I(x) := \{j : j \in \{1, 2, \cdots, m\} \text{ and } |x_j| > 0\}$ for any $x \in \Re^m$.

- Let X admit the singular value decomposition (SVD):

$$X := U \left[\mathrm{Diag}(\sigma(X)) \; 0_{m\times(n-m)} \right] V^T, (U, V) \in \mathcal{O}^{m,n}(X),$$

where $\sigma_1(X) \geq \sigma_2(X) \geq \cdots \geq \sigma_m(X) \geq 0$.

- $\mathcal{O}^{m,n}(X)$ is given by

$$\mathcal{O}^{m,n}(X) := \left\{ \begin{array}{l} (U, V) \in \mathcal{O}^m \times \mathcal{O}^n : X = \\ U \left[\mathrm{Diag}(\sigma(X)) \; 0_{m\times(n-m)} \right] V^T \end{array} \right\},$$

where $\mathcal{O}^m$ represents the set of all $m \times m$ orthogonal matrices.

- The definitions of $\mathcal{A}$ and $\mathcal{A}^*$:

$$\mathcal{A}(X) := (\langle A_1, X\rangle, \langle A_2, X\rangle, \cdots, \langle A_q, X\rangle)^T$$
$$\mathcal{A}^*(y) := \sum_{i=1}^q y_i A_i,$$

  where $A_i \in \Re^{m\times n}, y \in \Re^q$.

- Let $G : \Re^{m\times n} \to \Re$ and $X, H \in \Re^{m\times n}$, the second-order Gâteaux derivative $\mathrm{D}^2 G(X)$ at X is defined as follows:

$$\mathrm{D}^2 G(X)H := \lim_{t\downarrow 0} \frac{\mathrm{D}G(X+tH) - \mathrm{D}G(X)}{t}.$$

Smoothing function

Let $\Phi : \Re^{m\times n} \to \Re$ be a continuous function. We call $\bar{\Phi} : \Re_+ \times \Re^{m\times n} \to \Re$ a smoothing function of Φ, if $\bar{\Phi}(\mu, \cdot)$ is continuously differentiable in $\Re^{m\times n}$ for any fixed $\mu > 0$, and for any $X \in \Re^{m\times n}$, we have

$$\lim_{\mu\downarrow 0, Z\to X} \bar{\Phi}(\mu, Z) = \Phi(X).$$

Necessary optimality conditions

Definition

For $X \in \Re^{m\times n}$ and $p \in (0,1)$, X is said to satisfy the first-order necessary condition of (6) if

$$\mathcal{A}(X)^T(\mathcal{A}(X)-b)+\tau\|X\|_p^p=0. \tag{8}$$

Also, X is said to satisfy the second-order necessary condition of (6) if

$$\|\mathcal{A}(X)\|_2^2+\tau(p-1)\|X\|_p^p \geq 0. \tag{9}$$

Lemma

Let $X^\star$ be a local minimizer of (6). Then, for any pair $(U^\star, V^\star) \in \mathcal{O}^{m,n}(X^\star)$, the vector $z^\star := \sigma(X^\star) \in \mathfrak{R}^m$ is a local minimizer of the following problem

$$\begin{array}{ll} \min & \varphi(z) := F(U^\star[Diag(z)\ 0_{m\times(n-m)}](V^\star)^T) \\ s.t. & z \in \mathfrak{R}^m. \end{array} \tag{10}$$

Theorem

Let $X^\star$ be any local minimizer of (6). Then $X^\star$ satisfies the conditions (8) and (9).

Lower bound result 1

Theorem

Let $X^\star$ be any local minimizer of (6) satisfying $F(X^\star) \le F(X_0)$ for any given point $X_0 \in \Re^{m\times n}$ and $\mu_A := \sqrt{q}\max_{1\le i\le q} \|A_i\|_F$. Then, for any $i \in \{1, 2, \cdots, m\}$, we have

$$\sigma_i(X^\star) < L(\tau, \mu_A, X_0, p) := \left(\frac{\tau}{\mu_A\sqrt{2F(X_0)}}\right)^{\frac{1}{1-p}} \Rightarrow \sigma_i(X^\star) = 0.$$

In addition, the rank of $X^\star$ is bounded by

$$\min\left(m, \frac{pF(X_0)}{\tau L(\tau, \mu_A, X_0, p)^p}\right).$$

Hence, if $X_0 = 0$ and $\|A_i\|_F = 1$ $(i = 1, 2, \cdots, q)$, we obtain the following corollary:

Corollary

Let $X^\star$ be any local minimizer of (6). Then, for any $i \in \{1, 2, \cdots, m\}$, we have

$$\sigma_i(X^\star) < L_1(\tau, p) := \left(\frac{\tau}{\sqrt{q}\|b\|_2}\right)^{\frac{1}{1-p}} \quad \Rightarrow \quad \sigma_i(X^\star) = 0.$$

In addition, the rank of $X^\star$ is bounded by $\min\left(m, \frac{p\|b\|_2^2}{2\tau L_1(\tau,p)^p}\right)$.

Lower bound result 2

Theorem

Let $X^\star$ be any local minimizer of (6) and $\mu_A := \sqrt{q}\max_{1\le i\le q}\|A_i\|_F$. Then, for any $i \in \{1,2,\cdots,m\}$, we have

$$\sigma_i(X^\star) < L_2(\tau,\mu_A,p) := \left(\frac{\tau(1-p)}{\mu_A^2}\right)^{\frac{1}{2-p}} \quad \Rightarrow \quad \sigma_i(X^\star) = 0.$$

We give a sufficient condition on the parameter τ of (6) to obtain a desirable low-rank solution, which is a natural extension of that introduced in [11, Theorem 2] for (7).

Theorem

Let $X^\star$ be any local minimizer of (6) satisfying $F(X^\star) \leq F(X_0)$ for any given point $X_0 \in \mathfrak{R}^{m\times n}$ and $\mu_A := \sqrt{q}\max_{1\leq i\leq q}\|A_i\|_F$. Let

$$\tau(\mu_A, s, X_0, p) := \left(\frac{p}{s}\right)^{1-p}(F(X_0))^{1-\frac{p}{2}}2^{\frac{p}{2}}\mu_A^p.$$

If $\tau \geq \tau(\mu_A, s, X_0, p)$, then $\mathrm{rank}(X^\star) < s$ for $s \geq 1$.

If $X_0 = 0$ and $\|A_i\|_F = 1$ $(i = 1, 2, \cdots, q)$, the following corollary holds at $X^\star$:

Corollary

Let $X^\star$ be any local minimizer of (6). Let

$$\tau_1(s, p) := \left(\frac{p}{2s}\right)^{1-p} \|b\|_2^{2-p} q^{\frac{p}{2}}.$$

If $\tau \geq \tau_1(s, p)$, then $\mathrm{rank}(X^\star) < s$ for $s \geq 1$.

We define the smoothing model as follows:

$$\begin{array}{ll} \min & \bar{F}(\epsilon, X) \\ \text{s.t.} & X \in \Re^{m\times n}, \end{array} \tag{11}$$

where $\bar{F}(\epsilon, X)$ is defined by

$$\bar{F}(\epsilon, X) = \frac{1}{2}\|\mathcal{A}(X) - b\|_2^2 + \frac{\tau}{p}\sum_{i=1}^{m}(\sigma_i^2(X) + \epsilon^2)^{\frac{p}{2}}. \tag{12}$$

According to the definitions of $F(X)$ and $\bar{F}(\epsilon, X)$, we obtain

$$0 \le \bar{F}(\epsilon, X) - F(X) \le \frac{\tau m|\epsilon|^p}{p}. \tag{13}$$

Let $X_\epsilon^\star$ be a local minimizer of (11) for a given $\epsilon > 0$. Then for any $H \in \mathfrak{R}^{m\times n}$, the following conditions hold at $X_\epsilon^\star$:

$$\langle \mathrm{D}_X F(\epsilon, X_\epsilon^\star), H\rangle = 0, \tag{14}$$

$$\left\langle \mathrm{D}_X^2 F(\epsilon, X_\epsilon^\star) H, H\right\rangle \geq 0. \tag{15}$$

Convergence of the smoothing method

Theorem

(1) *Let* $\{X^\star_{\epsilon^k}\}$ *be a sequence of matrices satisfying (14) with* $\epsilon = \epsilon^k$. *Then any accumulation point of* $\{X^\star_{\epsilon^k}\}$ *satisfies the first-order necessary condition of (6).*

(2) *Let* $\{X^\star_{\epsilon^k}\}$ *be a sequence of matrices satisfying (15) with* $\epsilon = \epsilon^k$. *Then any accumulation point of* $\{X^\star_{\epsilon^k}\}$ *satisfies the second-order necessary condition of (6).*

(3) *Let* $\{X^\star_{\epsilon^k}\}$ *be a sequence of matrices being global minimizer of (11). Then any accumulation point of* $\{X^\star_{\epsilon^k}\}$ *is the global minimizer of (6).*

Lower bound result 3

Theorem

Let $X^\star_{\epsilon^k}$ be any local minimizer of (11) satisfying $\bar{F}(\epsilon^k, X^\star_{\epsilon^k}) \le F(X_0)$ for any given point $X_0 \in \mathfrak{R}^{m\times n}$ and $\mu_A := \sqrt{q}\max_{1\le i\le q}\|A_i\|_F$. Then, for any $i \in \{1,2,\cdots,m\}$ and any scalar $\lambda \in (0,+\infty)$, we have

$$\sigma_i(X^\star_{\epsilon^k}) < \bar{L}(\tau,\mu_A,X_0,p,\lambda) := \left(\frac{\lambda^2}{1+\lambda^2}\right)^{\frac{2-p}{2(1-p)}}\left(\frac{\tau}{\mu_A\sqrt{2F(X_0)}}\right)^{\frac{1}{1-p}}$$
$$\Rightarrow\ \sigma_i(X^\star_{\epsilon^k}) \le \lambda|\epsilon^k|.$$

Denote

$$\bar{F}_1(X) := \frac{1}{2}\|\mathcal{A}(X) - b\|_2^2, \;\; \bar{F}_2(\epsilon, X) := \frac{\tau}{p}\sum_{i=1}^{m}(\sigma_i^2(X) + \epsilon^2)^{\frac{p}{2}},$$

then

$$\bar{F}(\epsilon, X) = \bar{F}_1(X) + \bar{F}_2(\epsilon, X).$$

$$\mathrm{D}_X \bar{F}_1(X) = \mathcal{A}^*(\mathcal{A}(X) - b), \ \mathrm{D}_X \bar{F}_2(\epsilon, X) = \tau W(\epsilon, X)X,$$
$$\mathrm{D}_X \bar{F}(\epsilon, X) = \mathrm{D}_X \bar{F}_1(X) + \mathrm{D}_X \bar{F}_2(\epsilon, X),$$
$$\mathrm{D}_\epsilon \bar{F}(\epsilon, X) = \tau\epsilon \ \mathrm{Tr}(W(\epsilon, X)), \quad \text{if } \epsilon > 0,$$

where

$$W(\epsilon, X) := U\mathrm{Diag}\left((\sigma_1^2(X) + \epsilon^2)^{\frac{p}{2}-1}, \cdots, (\sigma_m^2(X) + \epsilon^2)^{\frac{p}{2}-1}\right) U^T,$$

and $(U, V) \in \mathcal{O}^{m,n}(X)$.

A majorization to $\bar{F}(\epsilon, X)$

$$\begin{array}{ll} \min & \hat{F}^k(\epsilon, X) \\ \text{s.t.} & (\epsilon, X) \in \Re \times \Re^{m\times n}, \end{array} \tag{16}$$

where

$$\hat{F}^k(\epsilon, X) = \bar{F}_1(X) + \tilde{F}_2^k(\epsilon, X, \eta^k) + \frac{\tau\rho^k}{2}\left[\|X - X^k\|_F^2 + (\epsilon - \epsilon^k)^2\right],$$

$$\tilde{F}_2^k(\epsilon, X, \eta^k) = \frac{\tau}{2}\sum_{i=1}^{m}\left[(\sigma_i^2(X) + \epsilon^2)(\eta^k)_i - \frac{p-2}{p}(\eta^k)_i^{\frac{p}{p-2}}\right],$$

$$\eta^k = \left((\sigma_1^2(X^k) + (\epsilon^k)^2)^{\frac{p}{2}-1}, \cdots, (\sigma_m^2(X^k) + (\epsilon^k)^2)^{\frac{p}{2}-1}\right)^T,$$

and $\rho^k > 0$ denotes the proximal parameter.

Solving (16) approximately

Instead of solving the stationary condition for (16) exactly, we consider

$$\begin{aligned} &\mathrm{D}_X \bar{F}_1(X) + \tau W(\epsilon^k, X^k) X^k + \tau\rho^k (X - X^k) = 0, \\ &\epsilon \mathrm{Tr}(W(\epsilon^k, X^k)) + \rho^k(\epsilon - \epsilon^k) = 0. \end{aligned} \tag{17}$$

$$\begin{aligned} \hat{X}^k &= \mathcal{G}^{-1}(\tau\rho^k X^k - \tau W(\epsilon^k, X^k) X^k + \mathcal{A}(b)), \\ \hat{\epsilon}^k &= \frac{\rho^k}{\rho^k + \mathrm{Tr}(W(\epsilon^k, X^k))} \epsilon^k, \end{aligned} \tag{18}$$

where $\mathcal{G}(X) := \mathcal{A}^*(\mathcal{A}(X)) + \tau\rho^k X$. A reasonable bound for ρ^k is obtained under the update rule (18).

Algorithm Smajor

Algorithm: Smajor (Smoothing majorization algorithm)

Step 0: *Choose the initial pair* (ϵ^0, X^0) *and set the counter* $k := 0$.

Step 1: *Set the parameter* $\rho^k \geq 0$. *Construct the problem (16) at* (ϵ^k, X^k) *(namely* $\min \hat{F}^k(\epsilon, X)$*).*

Step 2: *Set* $(\epsilon^{k+1}, X^{k+1}) := (\hat{\epsilon}^k, \hat{X}^k)$, $k := k + 1$ *and go to Step 1.*

Lemma

Let $\{(\epsilon^k, X^k)\}$ be the pairs of sequence generated by Algorithm Smajor. Then, we have

(1) *For any positive integer k, $\hat{F}^k(\epsilon^k, X^k) = \bar{F}(\epsilon^k, X^k)$.*

(2) *For any positive integer k, we have*

$$\begin{aligned}\hat{F}^k(\epsilon^{k+1}, X^{k+1}) \geq \; & \bar{F}(\epsilon^{k+1}, X^{k+1}) \\ & + \frac{\tau \rho^k}{2}\left[\|X^{k+1} - X^k\|_F^2 + (\epsilon^{k+1} - \epsilon^k)^2\right].\end{aligned}$$

Lemma

Let $\{(\epsilon^k, X^k)\}$ be the pairs of sequence generated by Algorithm Smajor. The parameter ρ^k satisfies the following condition:

$$\rho^k \geq \max_{1 \leq i \leq m} \eta_i^k, \tag{19}$$

where η^k is defined as

$$\eta^k := \left((\sigma_1^2(X^k) + (\epsilon^k)^2)^{\frac{p}{2}-1}, \cdots, (\sigma_m^2(X^k) + (\epsilon^k)^2)^{\frac{p}{2}-1} \right)^T,$$

then

$$\hat{F}^k(\epsilon^k, X^k) \geq \hat{F}^k(\epsilon^{k+1}, X^{k+1}). \tag{20}$$

Theorem

Let $\{(\epsilon^k, X^k)\}$ *be generated by Smajor,* ρ^k *satisfies (19).*

(1) $\{\bar{F}(\epsilon^k, X^k)\}$ *is a monotonically decreasing sequence:*

$$\bar{F}(\epsilon^k, X^k) - \frac{\tau\rho^k}{2}\left[\|X^{k+1} - X^k\|_F^2 + (\epsilon^{k+1} - \epsilon^k)^2\right] \geq \bar{F}(\epsilon^{k+1}, X^{k+1}).$$

(2) *The sequence* $\{(\epsilon^k, X^k)\}$ *contained in the level set* $\{(\epsilon, X) : \bar{F}(\epsilon, X) \leq F(X_0)\}$ *for some* $X_0 \in \Re^{m\times n}$ *is bounded. Let* $(\epsilon^\star, X^\star)$ *be any accumulation point of the sequence* $\{(\epsilon^k, X^k)\}$*. Then* $X^\star$ *satisfies the first-order necessary condition of (6).*

Numerical results

We report numerical results for solving a series of matrix completion problems of the form:

$$\begin{aligned} \min \quad & \frac{1}{2}\|(X - X_R)_\Omega\|_2^2 + \frac{\tau}{p}\|X\|_p^p \\ \text{s.t.} \quad & X \in \mathfrak{R}^{m\times n}, \end{aligned} \tag{21}$$

where Ω is an index set of the original matrix X_R and $(X - X_R)_\Omega \in \mathfrak{R}^q$ is obtained from $(X - X_R)$ by selecting entries whose indices are in Ω.

From (18), we present the update formulas of X and ϵ for (21) as follows:

$$\begin{aligned}
\mathcal{P}_\Omega(X^{k+1}) &= \mathcal{P}_\Omega\left(\tfrac{\tau\rho^k}{1+\tau\rho^k}X^k - \tfrac{\tau}{1+\tau\rho^k}W(\epsilon^k, X^k)X^k\right) \\
&\quad + \frac{1}{1+\tau\rho^k}\mathcal{P}_\Omega(X_R), \\
\mathcal{P}_{\Omega^c}(X^{k+1}) &= \mathcal{P}_{\Omega^c}\left(X^k - \tfrac{1}{\rho^k}W(\epsilon^k, X^k)X^k\right), \\
\epsilon(\rho^k) &= \frac{\rho^k}{\rho^k + \mathrm{Tr}(W(\epsilon^k, X^k))}\epsilon^k,
\end{aligned} \tag{22}$$

where Ω^c denotes the complement of Ω and

$$(\mathcal{P}_\Omega(X))_{ij} = \begin{cases} 0 & \text{if } (i,j) \notin \Omega, \\ X_{ij} & \text{otherwise.} \end{cases}$$

Random matrix completion problems

We begin by examining the behavior of Smajor on random matrix completion problems and its sensitivity to the parameters p, $SR = |\Omega|/mn$ and X^0.

- Sensitivity to the parameter p.
- Sensitivity to the parameter SR.
- Sensitivity to the initial point X^0

.

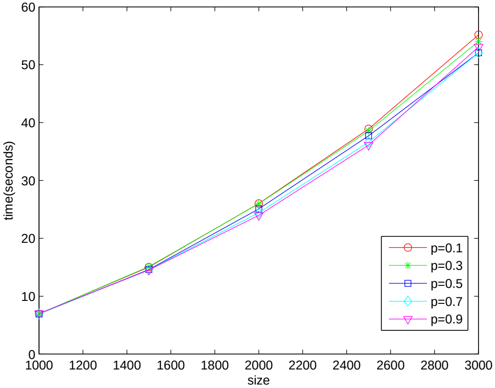

Figure: (a) The total computing time for different p.

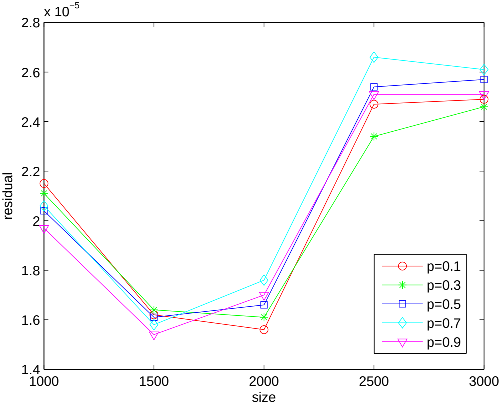

Figure: (b) The residual for different p.

Numerical results for SR= 0.39.

n	r	FR	rr	it.	time	Res
1200	5	0.021	5	47	7.48	7.73e-6
1400	5	0.018	5	46	9.77	8.05e-6
1600	5	0.016	5	46	12.66	8.41e-6
1800	5	0.014	5	45	15.19	9.21e-6
2000	5	0.013	5	44	18.94	9.80e-6
2200	5	0.012	5	44	23.49	1.27e-5
2400	5	0.011	5	43	27.01	1.39e-5
2600	5	0.010	5	42	31.03	1.65e-5
2800	5	0.009	5	42	36.26	1.96e-5
3000	5	0.009	5	41	40.38	2.28e-5

Numerical results for different SR= 0.57

n	r	FR	rr	it.	time	Res
1200	5	0.015	5	47	7.54	2.60e-6
1400	5	0.013	5	46	9.69	2.51e-6
1600	5	0.011	5	46	12.73	2.52e-6
1800	5	0.010	5	45	15.34	2.39e-6
2000	5	0.009	5	44	19.36	2.53e-6
2200	5	0.008	5	44	23.51	2.44e-6
2400	5	0.007	5	43	27.18	2.50e-6
2600	5	0.007	5	42	32.07	2.42e-6
2800	5	0.006	5	42	36.41	2.42e-6
3000	5	0.006	5	41	40.60	2.50e-6

Numerical results for different initial point X^0.

n	rr	it.	A-time	V-time	A-Res	V-Res
1000	10	48	7.26	2.00e-4	2.03e-5	3.80e-3
1500	10	46	15.15	1.00e-4	1.63e-5	1.60e-3
2000	10	44	25.09	7.50e-3	1.69e-5	1.00e-4
2500	10	43	39.38	2.30e-3	2.49e-5	1.00e-4
3000	10	41	52.85	1.40e-3	2.60e-5	2.00e-4

Comparison for matrix completion problems

Now, we report numerical results from two groups of experiments. In the first group of test, we compare our algorithm Smajor with SVT and sIRLS under the assumption that the information of rank(X_R) is known in advance. In the second group, the rank of X_R is assumed to be unknown.

Using the lower bound

The procedure for estimating the true rank:

Step 0: *Initialize the rank* $s := 1$ *and set the step* $s_{inc} := 3$. *Choose* τ^0 *to satisfy the condition*

$$\left(\frac{p}{2(s+1)}\right)^{1-p} \|(X_R)_\Omega\|_2^{2-p} q_0^{\frac{p}{2}} \leq \tau^0 \leq \left(\frac{p}{2s}\right)^{1-p} \|(X_R)_\Omega\|_2^{2-p} q_0^{\frac{p}{2}},$$

Set $\tau := \tau^0$, $k = 0$ *and run the s-truncated SVD of* X^k *using the package PROPACK, i.e.,*

$$X^k := \bar{U}^k Diag\left(\sigma_1(X^k), \cdots, \sigma_s(X^k)\right) (\bar{V}^k)^T.$$

Step 1: *Calculate the lower bound $L_1(\tau^0, p)$ as follows:*

$$L_1(\tau^0, p) := \left(\frac{\tau^0}{\sqrt{q_0}\|(X_R)_\Omega\|_2} \right)^{\frac{1}{1-p}}.$$

Step 2: *Compute $W(\epsilon^k, X^k)$ and $Tr(W(\epsilon^k, X^k))$.*
Step 3: *Using (22) to obtain the iteration $(\epsilon^{k+1}, X^{k+1})$ and run the $(s + s_{inc})$-truncated SVD of X^{k+1}, i.e,*
$X^{k+1} := \bar{U}^{k+1} Diag\left(\sigma_1(X^{k+1}), \cdots, \sigma_{(s+s_{inc})}(X^{k+1})\right)(\bar{V}^{k+1})^T$.

Step 4:*Set the estimated rank s as follows:*

$$s := \max\left\{i \in \{1, \cdots, (s + s_{inc})\} \mid \sigma_i(X^{k+1}) > L_1(\tau^0, p)\right\}$$

and set

$$X^{k+1} := \hat{U}^{k+1} Diag\left(\sigma_1(X^{k+1}), \cdots, \sigma_s(X^{k+1})\right)(\hat{V}^{k+1})^T,$$

where $\hat{U}^{k+1}$ *and* $\hat{V}^{k+1}$ *are the sub-matrix of* $\bar{U}^{k+1}$ *and* $\bar{V}^{k+1}$ *whose columns are the first s columns of* $\bar{U}^{k+1}$ *and* $\bar{V}^{k+1}$*, respectively. Set* $\bar{U}^{k+1} := \hat{U}^{k+1}$ *and* $\bar{V}^{k+1} := \hat{V}^{k+1}$*.*

Step 5:*If the termination criterion*

$$e(\epsilon, X) := \max\{\mathcal{A}(X)^T(\mathcal{A}(X) - b) + \tau\|X\|_p^p,\ \epsilon^2\} \leq tol,$$

holds at $(\epsilon^{k+1}, X^{k+1})$*, stop; otherwise, update the parameter* τ *as*

$$\tau^{k+1} = \max\{\gamma_\tau \tau^k, \bar{\tau}\},$$

and choose τ^0 *satisfy the condition in Step 0, set* $k := k + 1$ *and go to Step 1.*

Smajor-SVT-sIRLS

Numerical results for random matrix completion problems when the rank of X_R is known

n	FR	rr	$it.$	$time$	Res	rr	$it.$	$time$	Res	rr	$it.$	$time$	Res
1000	0.026	5	48	5.18	7.66e-6	5	50	6.06	3.28e-5	5	80	6.33	3.39e-4
1500	0.017	5	46	10.60	8.45e-6	5	47	12.97	3.08e-5	5	80	14.07	3.33e-4
2000	0.013	5	44	18.77	9.73e-6	5	43	25.03	2.95e-5	5	80	25.66	3.28e-4
2500	0.010	5	43	28.99	1.54e-5	5	41	34.03	2.79e-5	5	80	43.14	3.19e-4
3000	0.009	5	41	40.25	2.43e-5	5	39	43.96	3.07e-5	5	80	60.21	3.11e-4
1000	0.051	10	48	7.34	2.04e-5	10	54	8.49	3.60e-5	10	80	7.41	3.85e-4
1000	0.034	10	46	15.04	1.60e-5	10	47	18.75	3.21e-5	10	80	15.48	3.60e-4
2000	0.026	10	44	25.13	1.67e-5	10	43	31.61	2.92e-5	10	80	27.55	3.47e-4
2500	0.020	10	43	38.46	2.50e-5	10	41	43.70	2.57e-5	10	80	45.75	3.40e-4
3000	0.017	10	41	52.99	2.62e-5	10	39	68.28	2.72e-5	10	80	63.51	3.33e-4
1000	0.102	20	48	8.97	4.86e-5	20	67	12.75	1.01e-4	20	80	9.81	4.80e-4
1500	0.068	20	46	17.94	5.09e-5	20	56	24.87	9.24e-5	20	80	19.15	4.17e-4
2000	0.051	20	44	27.95	5.37e-5	20	51	42.73	9.22e-5	20	80	32.95	3.97e-4
2500	0.041	20	43	43.28	6.22e-5	20	47	59.84	8.07e-5	20	80	52.98	3.62e-4
3000	0.034	20	41	61.16	8.57e-5	20	45	82.15	8.83e-5	20	80	72.85	3.58e-4

Numerical results for random matrix completion problems when the rank of X_R is unknown.

n	FR	rr	$it.$	$time$	Res	rr	$it.$	$time$	Res	rr	$it.$	$time$	Res
1000	0.026	5	48	6.19	7.67e-6	5	54	6.80	5.07e-5	5	80	10.73	3.49e-4
1500	0.017	5	46	14.12	8.62e-6	5	50	14.36	4.83e-5	5	80	34.77	3.38e-4
2000	0.013	5	44	22.94	9.80e-6	5	46	27.78	4.81e-5	5	80	79.95	3.30e-4
2500	0.010	5	43	37.79	1.68e-5	5	44	38.46	4.66e-5	5	80	141.24	3.24e-4
3000	0.009	5	41	52.75	2.66e-5	5	42	59.02	5.23e-5	5	80	254.01	3.16e-4
1000	0.051	10	48	7.55	2.15e-5	10	64	9.89	1.17e-4	10	80	11.59	4.14e-4
1500	0.034	10	46	15.57	1.61e-5	10	56	19.79	1.11e-4	10	80	35.94	3.72e-4
2000	0.026	10	44	25.68	1.90e-5	10	51	33.97	9.70e-5	10	80	81.81	3.50e-4
2500	0.020	10	43	41.86	2.67e-5	10	48	47.31	9.52e-5	10	80	155.47	3.44e-4
3000	0.017	10	41	57.11	3.89e-5	10	44	69.57	9.84e-5	10	80	256.12	3.35e-4
1000	0.102	20	48	9.01	6.18e-5	20	69	13.45	1.22e-4	20	80	16.10	4.80e-4
1500	0.068	20	46	17.58	7.72e-5	20	58	27.62	1.16e-4	20	80	40.97	4.26e-4
2000	0.051	20	44	29.35	8.30e-5	20	52	47.27	1.05e-4	20	80	85.71	4.02e-4
2500	0.041	20	43	46.31	9.80e-5	20	49	64.69	1.04e-4	20	80	161.66	3.75e-4
3000	0.034	20	41	64.91	1.02e-4	20	46	93.27	1.07e-4	20	80	264.12	3.66e-4

Experiments on Movielens 100k data sets

We implement Smajor, sIRLS, IHT [24] and Optspace [18] to tackle the matrix completion problem whose data is taken from the well-known MovieLens data sets. In our numerical experiments, we consider MovieLens 100k data sets, which is available on the website http://www.grouplens.org/node/73. The MovieLens 100k data sets include four small data pairs (u1.base,u1.test), (u2.base,u2.test), (u3.base,u3.test), (u4.base,u4.test). For each data set, we train Smajor sIRLS, IHT and Optspace on the training set and compare their performance on the corresponding test set.

Define the mean absolute error (MAE) of the output matrix X generated by the algorithm as follows:

$$\mathrm{MAE} := \frac{\sum_{(i,j)\in\Omega} |X_{ij} - M_{ij}|}{|\Omega|}.$$

The matrices M_{ij} and X_{ij} are the original and computed ratings of movie j by user i, respectively. The normalized mean absolute error (NMAE) is used to measure the accuracy of the approximated completion X,

$$\mathrm{NMAE} := \frac{\mathrm{MAE}}{r_{\max} - r_{\min}},$$

where $r_{\max}, r_{\min}$ are upper and lower bounds for the ratings of movies.

NMAE for different algorithms.

Data sets	Smajor	sIRLS	IHT	Optspace
(u1.base, u1.test)	0.1924	0.1924	0.1925	0.1887
(u2.base, u2.test)	0.1871	0.1872	0.1884	0.1877
(u3.base, u3.test)	0.1883	0.1873	0.1874	0.1882
(u4.base, u4.test)	0.1888	0.1898	0.1897	0.1883

Conclusions

- We present the lower bound analysis for nonzero singular values in solutions of the l_2^2-l_p^p and the smoothing model.
- A smoothing model is proposed to approximate l_2^2-l_p^p, the convergence of stationary points and the global solutions of the smoothing model is demonstrated.
- A majorization algorithm in which the smoothing parameter ε is treated as a variable, is used to solve the smoothing model.
- The smoothing majorization algorithm is implemented to solve matrix completion problems and numerical results are reported.

THANK YOU !

- J. Abernethy, F. Bach, T. Evgeniou, and J. P. Vert, *Low-rank matrix factorization with attributes*, Technical Report, Ecole des Mines de Paris, 2006.
- Y. Amit, M. Fink, N. Srebro, and S. Ullman, *Uncovering shared structures in multiclass classification*, In *Proceedings of the 24th International Conference on Machine Learning*, ACM, Providence, RI, 2007, pp. 17–24.
- A. Argyriou, T. Evgeniou, and M. Pontil, *Multi-task feature learning*, Adv. Neural Inform. Process. Syst., 19 (2007), pp. 41–48.
- J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.

- J. F. Cai, E. J. Candes, and Z. Shen, *A singular value thresholding algorithm for matrix completion*, SIAM J. Optim., 20(2010), pp. 1956–1982.
- E. J. Candès, M. B. Wakin, and S. P. Boyd, *Enhancing sparsity by reweighted l_1 minimization*, J. Fourier Anal. and Appl., 14 (2008), pp. 877–905.
- R. Chartrand and V. Staneva, *Restricted isometryproperties and nonconvex compressive sensing*, Inverse Problems, 24 (2008), pp. 1–14.
- R. Chartrand and W. Yin, *Iteratively reweighted algorithms for compressive sensing*, in *International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 3869–3872.

- X. J. Chen, *Smoothing methods for nonsmooth, nonconvex minimization*, Math. Program., 134 (2012), pp. 71–99.
- X. J. Chen, F. Xu, and Y. Y. Ye, *Lower bound theory of nonzero entries in solutions of l_2-l_p minimization*, SIAM J. Sci. Comput., 32 (2011), pp. 2832–2852.
- X. J. Chen, D. D. Ge, Z. Z. Wang and Y. Y. Ye, *Complexity of unconstrained l_2-l_p minimization*, Math. Program., 143 (2014), pp. 371–383.
- I. Daubechies, R. DeVore, M. Fornasier, and C. S. Gntrk, *Iteratively reweighted least squares minimization for sparse recovery*, Commun. Pur. Appl. Math., 63 (2010), pp. 1–38.

M. Fazel, *Matrix rank minimization with applications*, Ph.D. thesis, Stanford University, 2002.

M. Fazel, H. Hindi, and S. Boyd, *A rank minimization heuristic with application to minimum order system approximation*, in *Proceedings of the American Control Conference*, IEEE, 2001, pp. 4734–4739.

D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Comput. Math. Appl., 2 (1976), pp. 17–40.

Y. Gao and D. F. Sun, *A majorized penalty approach for calibrating rank constrained correlation matrix problems*, Technical report, Department of Mathematics, National University of Singapore, Singapore, 2010.

- K. Goldberg, T. Roeder, D. Gupta and C. Perkins, *Eigentaste–A constant time collaborative filtering algorithm*, Inf. Retr., 4 (2001), pp. 133–151.
- R. H. Keshavan and S. Oh, *A gradient descent algorithm on the grassman manifold for matrix completion*, DOI: 10.1016/j.trc.2012.12.007, 2009.
- R. M. Larsen, *PROPACK–Software for large and sparse SVD calculations*, software available at http://sun.stanford.edu/∼rmunk/PROPACK/.
- A. S. Lewis, *Derivatives of spectral functions*, Math. Oper. Res., 21 (1996), pp. 576–588.

- A. S. Lewis and H. S. Sendov, *Twice differentiable spectral functions*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 368–386.
- N. Linial, E. London, and Y. Rabinovich, *The geometry of graphs and some of its algorithmic applications*, Combinatorica, 15 (1995), pp. 215–245.
- S. Ma, D. Goldfarb, and L. Chen, *Fixed point and Bregman iterative methods for matrix rank minimization*, Math. Program., 128 (2011), pp. 321–353.
- R. Meka, P. Jain, and I. S. Dhillon, *Guaranteed rank minimization via singular value projection*, In *Proceedings of Neural Information Processing Systems* (NIPS), 2010.

K. Mohan and M. Fazel, *Iterative reweighted least squares for matrix rank minimization*, J. Mach. Learn. Res., 13(2012), pp. 3441–3473.

J. M. Otega and W. C. Rheinboldt, *Iterative solutions of nonlinear equations in several variables*, Academic Press, New York, 1970.

C. Tomasi and T. Kanade, *Shape and motion from image streams under orthography: a factorization method*, Int. J. Comput. Vision, 9 (1992), pp. 137–154.

Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, Springer, 2004.

THANK YOU !