Knowl Inf Syst
DOI 10.1007/s10115-013-0713-z

REGULAR PAPER

Joint Schatten p-norm and ℓ_p-norm robust matrix completion for missing value recovery

Feiping Nie · Hua Wang · Heng Huang · Chris Ding

Received: 21 March 2013 / Revised: 29 June 2013 / Accepted: 29 November 2013

Abstract The low-rank matrix completion problem is a fundamental machine learning and data mining problem with many important applications. The standard low-rank matrix completion methods relax the rank minimization problem by the trace norm minimization. However, this relaxation may make the solution seriously deviate from the original solution. Meanwhile, most completion methods minimize the squared prediction errors on the observed entries, which is sensitive to outliers. In this paper, we propose a new robust matrix completion method to address these two problems. The joint Schatten p-norm and ℓ_p-norm are used to better approximate the rank minimization problem and enhance the robustness to outliers. The extensive experiments are performed on both synthetic data and real-world applications in collaborative filtering prediction and social network link recovery. All empirical results show that our new method outperforms the standard matrix completion methods.

Keywords Matrix completion · Schatten p-norm · ℓ_p-norm · Recommendation system · Social network

1 Introduction

The prediction of the incomplete observations of an evolving matrix is a challenge of interest in many machine learning, data mining, and information retrieval applications [1–3], such as

F. Nie · H. Huang (✉) · C. Ding
Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX, USA
e-mail: heng@uta.edu

F. Nie
e-mail: feipingnie@gmail.com

C. Ding
e-mail: chqding@uta.edu

H. Wang
Department of Electrical Engineering and Computer Science,
Colorado School of Mines, Golden, CO, USA
e-mail: huawangcs@gmail.com

Published online: 20 December 2013

friendship prediction in social network, rating value estimation in recommendation system and collaborative filtering, and link prediction in protein–protein interaction network. All these problems can be seen as a special case of matrix completion where the goal is to impute the missing entries of the data matrix. As one emerging technique of compressive sensing, the problem of matrix completion has been extensively studied on both theory and algorithms [2,4–8] and also became popular after the recent concluded million-dollar Netflix competition.

The matrix completion methods assume that the values in the data matrix (graph) are correlated and the rank of the data matrix is low. The missing entries can be recovered using the observed entries by minimizing the rank of the data matrix, which is an NP hard problem. Instead of solving such an NP hard problem, the researchers minimize the trace norm (the sum of the singular values of the data matrix) as the convex relaxation of the rank function. Many recent research has been focusing on solving such trace norm minimization problem or extended formulations [7,9–20]. Meanwhile, instead of strictly keeping the values of the observed entries, the recent research work relaxed it to minimize the prediction errors (using squared error function) on the observed entries [7].

Although the trace norm minimization-based matrix completion objective is a convex problem with global solution, the relaxation may make the solution seriously deviate from the original solution. It is desired to solve a better approximation of the rank minimization problem without introducing much computational cost. In this paper, we reformulate the matrix completion problem using the Schatten p-norm. When $p \to 0$, our new objective can approximate the rank minimization better than the trace norm. Moreover, to improve the robustness of matrix completion method, we introduce the ℓ_p-norm $(0 < p \leq 1)$ error function for the prediction errors on the observed entries. Thus, our new objective minimizes the joint Schatten p-norm and ℓ_p-norm $(0 < p \leq 1)$. When $p \to 0$, our objective is more robust and effective than the standard matrix completion methods, which is a special case of our objective when $p = 1$. Although our objective function is not a convex problem (when $p < 1$), we derive an efficient algorithm based on the alternating direction method. With extensive experiments, we observe that under a large number of random initializations, our new nonconvex objective can always find a better convergency result for the matrix completion without introducing much extra computational cost.

We evaluate our new method using both synthetic and real-world data sets. Six benchmark data sets from collaborative filtering and social network link prediction applications are utilized in our validations. All empirical results show that our new robust matrix completion method outperforms the standard missing value prediction approaches. To evaluate the robustness of joint Schatten p-norm and ℓ_p-norm, we also add Gaussian noise into benchmark data sets and then perform all compared methods for matrix completion. In all experimental results, our method still has the best performance on all data sets and also is affected least by noise in all methods.

In summary, we highlight the main contributions of this paper as follows:

- We propose a novel objective function for the robust matrix completion task by utilizing joint Schatten p-norm and ℓ_p-norm.
- Optimizing the proposed objective function is a nontrivial problem; thus, we derive an optimization algorithm to solve this problem.
- We derive the optimal solution to the problem (24), which generalizes a famous soft thresholding result in [8] and can be used in many other Schatten p-norm minimization problems.

2 A new robust matrix completion method

2.1 Definitions of ℓ_p-norm and Schatten p-norm

The ℓ_p-norm[1] ($0 < p < \infty$) of a vector $v \in \mathbb{R}^{n\times 1}$ is defined as $\|v\|_p = \left(\sum_i^n |v_i|^p\right)^{\frac{1}{p}}$, where v_i is the i-th element of v. Thus, the p-norm of a vector $v \in \mathbb{R}^{n\times 1}$ to the power p is $\|v\|_p^p = \sum_i^n |v_i|^p$. Similarly, we can define the p-norm of a matrix $X \in \mathbb{R}^{n\times m}$ to the power p as $\|X\|_p^p = \sum_i^n \sum_j^m |x_{ij}|^p$.

The extended Schatten p-norm ($0 < p < \infty$) of a matrix $X \in \mathbb{R}^{n\times m}$ is defined as

$$\|X\|_{S_p} = \left(\sum_{i=1}^{\min\{n,m\}} \sigma_i^p\right)^{\frac{1}{p}}, \tag{1}$$

where σ_i is the i-th singular value of X. Thus, the Schatten p-norm of a matrix $X \in \mathbb{R}^{n\times m}$ to the power p is

$$\|X\|_{S_p}^p = \sum_{i=1}^{\min\{n,m\}} \sigma_i^p. \tag{2}$$

When $p = 1$, the Schatten 1-norm is the trace norm or nuclear norm, which is usually denoted as $\|X\|_*$. When $p \to 0$, from Fig. 1 we can see the extended Schatten p-norm of X approximates the rank of X. If we define $0^0 = 0$, then when $p = 0$, Eq. (2) is the rank of X.

2.2 Robust matrix completion objective

We denote $X_\Omega = \{X_{ij}\,|(i,j) \in \Omega\}$, and $\|X_\Omega\|_p^p = \sum_{(i,j)\in\Omega} \left|X_{ij}\right|^p$. Suppose we are given the observed values $D_\Omega = \{D_{ij}\,|(i,j) \in \Omega\}$ in a matrix D, the matrix completion task is to predict the unobserved values in the matrix D. The general rank minimization problem [21–28] solves the following problem:

$$\min_X \|X_\Omega - D_\Omega\|_2^2 + \gamma \operatorname{rank}(X), \tag{3}$$

This problem is NP hard due to the rank function in the objective. In practice, the rank is relaxed to the Schatten 1-norm (trace norm), and then we solve the following relaxed problem:

$$\min_X \|X_\Omega - D_\Omega\|_2^2 + \gamma \|X\|_*. \tag{4}$$

However, the relaxation may make the solution deviate seriously from the original solution. Meanwhile, the used squared error is sensitive to outliers.

When $p \to 0$, the Schatten p-norm $\|X\|_{S_p}^p$ will approximate the rank of X [29]. In this paper, we replace the $\|X\|_*$ by $\|X\|_{S_p}^p$; the value of p can be selected from (0, 1]. When p is set to a value smaller than 1, then the resulted problem will better approximate the original problem. We also use the ℓ_p-norm ($0 < p \le 1$) as the error function to improve the robustness

[1] When $p \ge 1$, $\|v\|_p = (\sum_{i=1}^n |v_i|^p)^{\frac{1}{p}}$ strictly defines a norm that satisfies the three norm conditions, while it defines a quasinorm when $0 < p < 1$. The quasinorm extends the standard norm in the sense that it replaces the triangle inequality by $\|\mathrm{x} + \mathrm{y}\|_p \le K(\|\mathrm{x}\|_p + \|\mathrm{y}\|_p)$ for some $K > 1$. Because the mathematical formulations and derivations in this paper equally apply to both norm and quasinorm, we do not differentiate these two concepts for notation brevity.

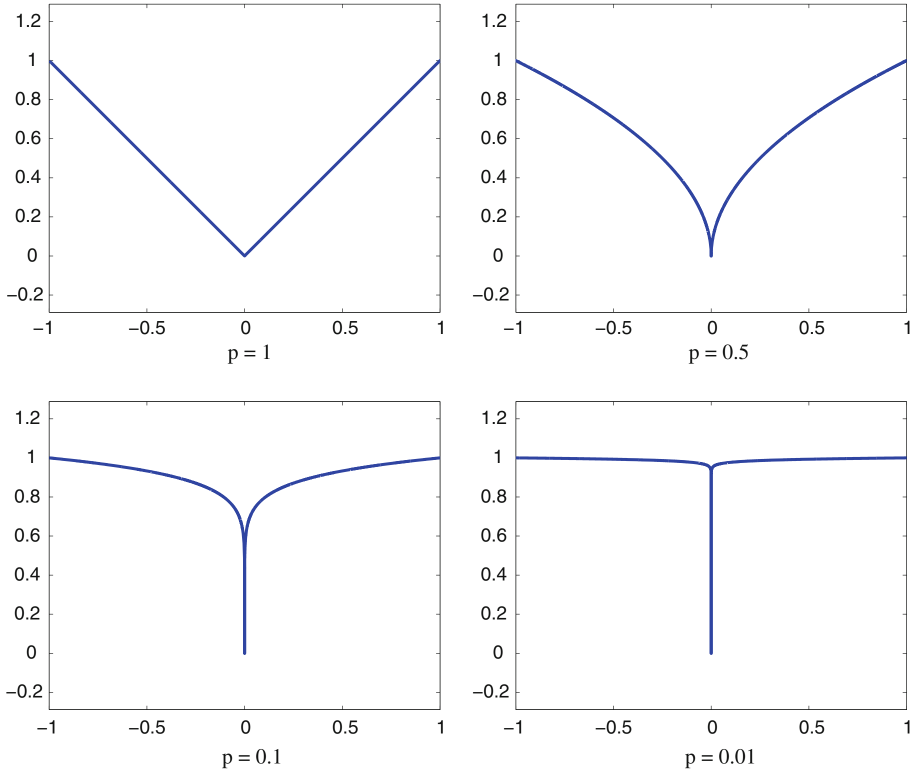


Fig. 1 The curves of function $|x|^p$ with different values of p

to outliers in given data [30] and propose to solve the following robust matrix completion problem (we could use the same p for the two norms to avoid one more parameter):

$$\min_X \|X_\Omega - D_\Omega\|_p^p + \gamma \|X\|_{S_p}^p . \tag{5}$$

3 Proposed algorithm

Solving the problem in Eq. (5) is challenge since both of the terms in Eq. (5) are nonsmooth and the Schatten p-norm is somewhat intractable. We use the augmented Lagrangian multiplier (ALM) method [31–33] to solve this problem and focus on the solutions to several related subproblems.

3.1 Brief description of augmented Lagrangian multiplier method

Consider the constrained optimization problem:

$$\min_{h(X)=0} f(X) \tag{6}$$

The algorithm using the augmented Lagrangian multiplier (ALM) method to solve the problem (6) is described in Algorithm 1.

It has been proved that under some rather general conditions, Algorithm 1 converges Q-linearly to the optimal solution [33]. This property makes the ALM method very attractive.

Algorithm 1 Algorithm to solve the problem (6).

Set $1 < \rho < 2$. Initialize $\mu > 0$, Λ
while not converge **do**
1. Update X by $\min_X f(X) + \frac{\mu}{2} \left\| h(X) + \frac{1}{\mu}\Lambda \right\|_F^2$
2. Update Λ by $\Lambda = \Lambda + \mu h(X)$
3. Update μ by $\mu = \rho\mu$

end while

3.2 Solving problem (5) using ALM method

We equivalently rewritten Problem (5) as:

$$\min_{X, E_\Omega = X_\Omega - D_\Omega, X = Z} \|E_\Omega\|_p^p + \gamma \|Z\|_{S_p}^p . \tag{7}$$

According to step 1 in Algorithm 1, we need to solve the following problem:

$$\begin{aligned} \min_{X, E_\Omega, Z} \; & \|E_\Omega\|_p^p + \gamma \|Z\|_{S_p}^p \\ & + \frac{\mu}{2} \left\| E_\Omega - (X_\Omega - D_\Omega) + \frac{1}{\mu}\Lambda_\Omega \right\|_F^2 \\ & + \frac{\mu}{2} \left\| X - Z + \frac{1}{\mu}\Sigma \right\|_F^2 . \end{aligned} \tag{8}$$

An accurate, joint minimization with respect to X, E_Ω, Z is difficult and costly; we can use the alternating direction method (ADM) [34] to solve this problem. Specifically, we optimize the problem with respect to one variable when fixing the other two variables, which result in the following three subproblems.

When fixing E_Ω, Z, the problem (8) is simplified to the following problem:

$$\min_X \|X_\Omega - M_\Omega\|_F^2 + \|X - N\|_F^2 , \tag{9}$$

where $M_\Omega = (E_\Omega + D_\Omega + \frac{1}{\mu}\Lambda_\Omega)$ and $N = (Z - \frac{1}{\mu}\Sigma)$. Denote $X_{\bar{\Omega}} = \{X_{ij} | (i, j) \notin \Omega\}$, the optimal solution to problem (9) can be easily obtained by

$$X_\Omega = \frac{M_\Omega + N_\Omega}{2}, \quad X_{\bar{\Omega}} = N_{\bar{\Omega}} \tag{10}$$

When fixing X, Z, the problem (8) is simplified to the following problem:

$$\min_{E_\Omega} \frac{1}{2} \|E_\Omega - H_\Omega\|_F^2 + \frac{1}{\mu} \|E_\Omega\|_p^p , \tag{11}$$

where $H_\Omega = X_\Omega - D_\Omega - \frac{1}{\mu}\Lambda_\Omega$.

When fixing X, E_Ω, the problem (8) is simplified to the following problem:

$$\min_Z \frac{1}{2} \|Z - G\|_F^2 + \frac{\gamma}{\mu} \|Z\|_{S_p}^p , \tag{12}$$

where $G = X + \frac{1}{\mu}\Sigma$

Based on the ALM method in Algorithm 1, the detailed algorithm to solve the problem in Eq. (5) is described in Algorithm 2.

Subsequently, we derive the optimal solution to subproblems in Eqs. (11) and (12), respectively.

Algorithm 2 Algorithm to solve the problem (5).

Set $1 < \rho < 2$. Initialize $\mu > 0$, Λ_Ω, Σ, E_Ω, Z
while not converge **do**
1. Update X by Eq. (10)
2. Update E_Ω by the optimal solution to problem (11)
3. Update Z by the optimal solution to problem (12)
4. Update Λ_Ω by $\Lambda_\Omega = \Lambda_\Omega + \mu(E_\Omega - X_\Omega + D_\Omega)$, Update Σ by $\Sigma = \Sigma + \mu(X - Z)$
5. Update μ by $\mu = \rho\mu$

end while

3.3 Solving the subproblem (11)

Note that the elements $\{X_{ij} \,|(i, j) \in \Omega\}$ in subproblem (11) can be decoupled. For each element, we only need to solve the following problem:

$$\min_x \frac{1}{2}(x - a)^2 + \lambda\,|x|^p \tag{13}$$

Denote the objective function in the problem (13) by $h(x)$, i.e.,

$$h(x) = \frac{1}{2}(x - a)^2 + \lambda\,|x|^p\,. \tag{14}$$

Note that $h(x)$ is an equation with one variable, and its convexity can be easily analyzed. The shapes of the curve of $h(x)$ with different values of a are shown in Fig. 2; the following is the details of the analysis of $h(x)$.

The function $h(x)$ is not differentiable at $x = 0$. In the following analysis, we only consider to find the minimal solution to $h(x)$ in the case of $x \neq 0$, and then we compare it with $h(0)$ to obtain the final minimal solution to $h(x)$.

If $x \neq 0$, we can see that the fist derivative of $h(x)$ is

$$g(x) = h'(x) = x - a + \lambda p\,|x|^{p-1}\,\mathrm{sgn}(x), \tag{15}$$

where the sign function $\mathrm{sgn}(x)$ is defined as follows: $\mathrm{sgn}(x) = 1$ if $x > 0$ and $\mathrm{sgn}(x) = -1$ if $x < 0$. According to Eq. (15), we know $h'(x) < 0$ when x closes 0 but $x < 0$, $h'(x) > 0$ when x closes 0 but $x > 0$.

We can obtain the local minimum of $h(x)$ by finding the root of $g(x) = 0$. The first and second derivative of $g(x)$ are as follows:

$$g'(x) = h''(x) = 1 - \lambda p(1 - p)\,|x|^{p-2}\,, \tag{16}$$

$$g''(x) = h'''(x) = \lambda p(1 - p)(2 - p)\,|x|^{p-3}\,\mathrm{sgn}(x), \tag{17}$$

From Eq. (17), we can see $g(x)$ is concave when $x < 0$, and $g(x)$ is convex when $x > 0$. Denote a constant v as

$$v = (\lambda p(1 - p))^{\frac{1}{2-p}}\,. \tag{18}$$

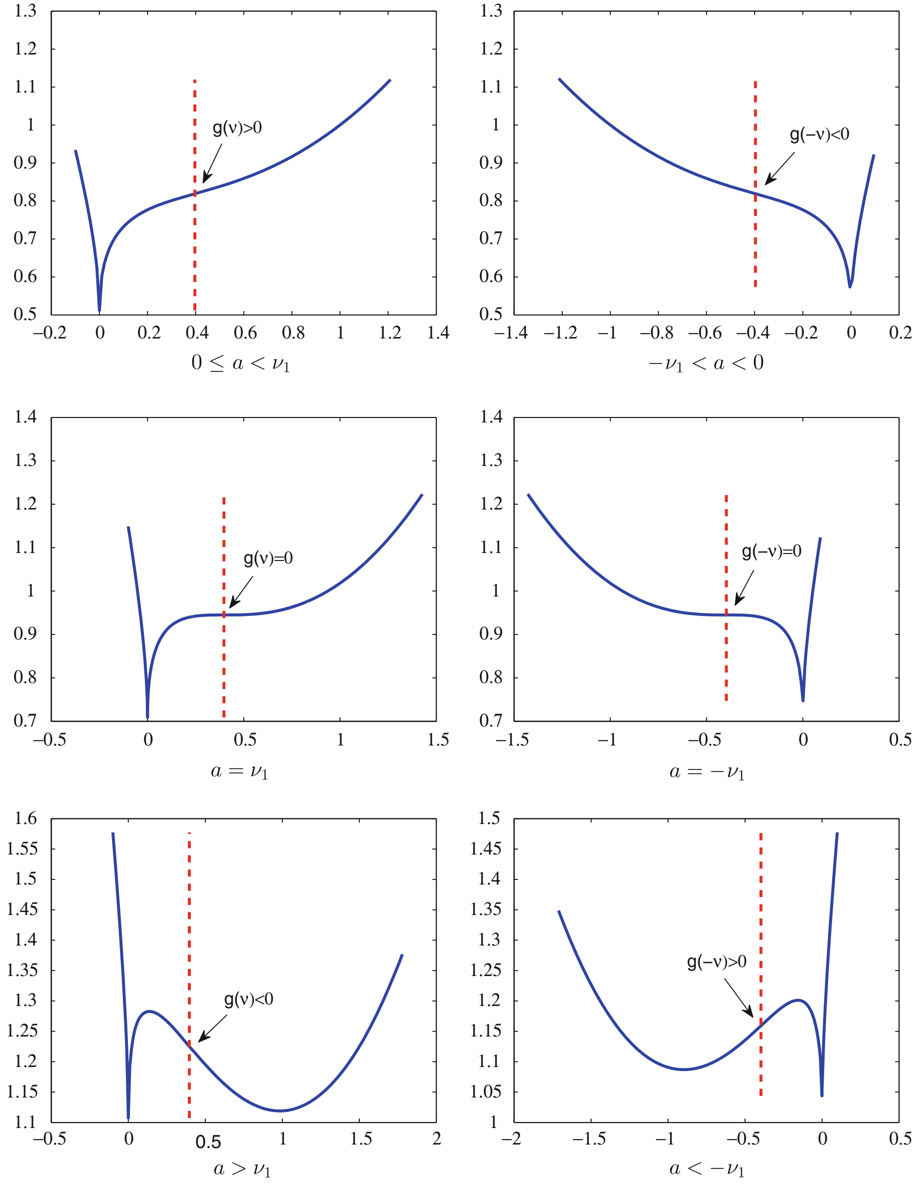


Fig. 2 The *shapes* of the curve of $h(x)$ in Eq. (14) with different values of a. The values ν and ν_1 are defined in Eqs. (18) and (20), respectively

Then we can see $g'(\nu) = 0$ and $g'(-\nu) = 0$. So $g(-\nu)$ is the maximum of $g(x)$ when $x < 0$, and $g(\nu)$ is the minimum of $g(x)$ when $x > 0$. We have the following three cases:

Case 1: $g(\nu) \geq 0$ and $g(-\nu) \leq 0$

In this case, $g(x)$ (or $h'(x)$) is always smaller than or equal to zero when $x < 0$, and $g(x)$ (or $h'(x)$) is always larger than or equal to zero when $x < 0$, which indicates that the minimal solution to $h(x)$ is 0.

Case 2: $g(\nu) < 0$

In this case, it can be easy to see $g(-\nu) < 0$, so $g(x)$ (or $h'(x)$) is always smaller than zero when $x < 0$, and $g(x) = 0$ (or $h'(x) = 0$) has two roots at $x > 0$. We can also see that $h(x)$ is concave at $0 < x \le \nu$ and $h(x)$ is convex at $x > \nu$. So the minimal solution to $h(x)(x \ne 0)$ in this case is the root of $g(x) = 0$ (or $h'(x) = 0$) at $a < x < -\nu$.

Case 3: $g(-\nu) > 0$

In this case, it can be easy to see $g(\nu) > 0$, so $g(x)$ (or $h'(x)$) is always larger than zero when $x > 0$, and $g(x) = 0$ (or $h'(x) = 0$) has two roots at $x < 0$. We can also see that $h(x)$ is concave at $-\nu \le x < 0$ and $h(x)$ is convex at $x < -\nu$. So the minimal solution to $h(x)(x \ne 0)$ in this case is the root of $g(x) = 0$ (or $h'(x) = 0$) at $a < x < -\nu$.

In summary, the optimal solution to problem (13) can be obtained by

$$\begin{cases} g(\nu) \ge 0, g(-\nu) \le 0 & x^* = 0 \\ g(\nu) < 0 & x^* = \arg\min_{x \in \{0, x_1\}} h(x) \\ g(-\nu) > 0 & x^* = \arg\min_{x \in \{0, x_2\}} h(x) \end{cases} \tag{19}$$

where x_1 and x_2 are the roots of $g(x) = 0$ at $\nu < x < a$ and $a < x < -\nu$, respectively. The roots can be easily obtained with Newton method initialized at a.

Denote another constant ν_1 as

$$\nu_1 = \nu + \lambda p \, |\nu|^{p-1}, \tag{20}$$

Then it is easy to know that Eq. (19) is equivalent to

$$\begin{cases} -\nu_1 \le a \le \nu_1 & x^* = 0 \\ a > \nu_1 & x^* = \arg\min_{x \in \{0, x_1\}} h(x) \\ a < -\nu_1 & x^* = \arg\min_{x \in \{0, x_2\}} h(x) \end{cases} \tag{21}$$

Similarly, consider the following problem which will be used later:

$$\min_{x \ge 0} \frac{1}{2}(x-a)^2 + \lambda \, |x|^p, \tag{22}$$

the optimal solution to problem (22) can be obtained by

$$\begin{cases} a \le \nu_1 & x^* = 0 \\ a > \nu_1 & x^* = \arg\min_{x \in \{0, x_1\}} h(x) \end{cases} \tag{23}$$

where x_1 is the root of $g(x) = 0$ at $\nu < x < a$, which can be easily obtained with Newton method initialized at a.

3.4 Solving the subproblem (12)

We rewrite the subproblem (12) as follows to simplify the notation.

$$\min_X \frac{1}{2} \|X - A\|_F^2 + \lambda \|X\|_{S_p}^p \tag{24}$$

To analyze its solution, we first introduce the following result:

Theorem 1 (von Neumann) *For any two matrices $A, B \in \mathbb{R}^{m \times n}$, $tr(A^T B) \le tr(\sigma(A)^T \sigma(B))$, where $\sigma(A)$ and $\sigma(B)$ are the ordered singular value matrices of A and B with the same order, i.e., the singular values in $\sigma(A)$ and $\sigma(B)$ are ordered with the same order.*

The problem (24) can be solved based on the following theorem:

Theorem 2 *The optimal solution X to problem (24) is $Q\Delta R^T$, where Q and R are the left and right singular vector matrices of A, respectively, and the i-th diagonal element δ_i of the diagonal matrix Δ is given by the optimal solution to the following problem (σ_i is the i-th singular value of A:*

$$\min_{\delta_i \geq 0} \frac{1}{2}(\delta_i - \sigma_i)^2 + \lambda \delta_i^p \tag{25}$$

Proof Suppose the SVD of X and A are $X = U\Delta V^T$ and $A = Q\Sigma R^T$, respectively, where Δ, Σ are ordered singular value matrices with the same order. Then problem (24) can be written as

$$\min_X \frac{1}{2}\left\| U\Delta V^T - Q\Sigma R^T \right\|_F^2 + \lambda tr(\Delta^p) \tag{26}$$

According to Theorem 1, we know $\left\| U\Delta V^T - Q\Sigma R^T \right\|_F^2 = tr(\Delta^T\Delta) + tr(\Sigma^T\Sigma) - 2tr(X^T A) \geq tr(\Delta^T\Delta) + tr(\Sigma^T\Sigma) - 2tr(\Delta^T\Sigma) = \|\Delta - \Sigma\|_F^2$, the equality holds only when $U = Q$ and $V = R$. Therefore, problem (24) is minimized when $U = Q$ and $V = R$, and Δ is the optimal solution to the following problem:

$$\min_{\Delta \geq 0} \frac{1}{2}\|\Delta - \Sigma\|_F^2 + \lambda tr(\Delta^p) \tag{27}$$

which is equivalent to (δ_i and σ_i are the i-th singular value of X and A, respectively)

$$\min_{\delta_i \geq 0} \frac{1}{2}\sum_i(\delta_i - \sigma_i)^2 + \lambda \sum_i \delta_i^p \tag{28}$$

Problem (28) can be decoupled to solve problem (25) for each i. □

Problem (25) is the same as problem (22); thus, the optimal solution to the problem (25) can be easily obtained according to Eq. (23). It is interesting to see that when $p = 1$, the derived solution is exactly the same as in [8], and our result extends the result in [8] to the case of $0 < p < 1$.

4 Discussion of the selection of value p in the robust matrix completion method

Despite the proposed robust matrix completion method in Eq. (5) with general values of $p \in (0, 1]$, we are more interested in the special cases of selecting the specific values of p.

In Eq. (5), we use the $\ell_p(0 < p \leq 1)$-norm as the error function to improve the robustness to outliers in given data and use the Schatten $p(0 < p \leq 1)$-norm as the regularization function to achieve a low-rank matrix. When we use the ℓ_p-norm as the error function to improve the robustness, the value of p should be smaller than 2, but it does not indicate that smaller p is better performance when $p \leq 1$. Since the error function is convex when $p = 1$, we suggest using the ℓ_1-norm as the error function to improve the robustness to outliers. In contrast, when we use the Schatten p-norm as the regularization function to achieve a low-rank matrix, we prefer smaller p since Schatten p-norm approximates the rank function when $p \to 0$. However, when $0 \leq p < 1$, the Schatten p-norm is not convex. Hence, it is a trade-off between $0 \leq p < 1$ (more accurate to approximate rank but not convex) and $p = 1$ (convex but not accurate to approximate rank) when we use the Schatten p-norm as the regularization function to achieve a low-rank matrix. In practice, we are interested in the following two cases: Schatten 1-norm (trace norm) and Schatten 0-norm (rank).

In the first case, we solve the following problem:

$$\min_X J = \|X_\Omega - D_\Omega\|_1 + \gamma \|X\|_* . \tag{29}$$

The problem (29) can also be solved using ALM method. In this case, the problem (13) becomes

$$\min_x \frac{1}{2}(x-a)^2 + \lambda |x| , \tag{30}$$

It is easy to see that the optimal solution to problem (30) is $x = \mathrm{sgn}(a)(|a| - \lambda)_+$, where $(x)_+$ is defined as follows: $(x)_+ = x$ if $x > 0$ and $(x)_+ = 0$ otherwise. Correspondingly, the problem (22) becomes

$$\min_{x \geq 0} \frac{1}{2}(x-a)^2 + \lambda |x| , \tag{31}$$

Note that $|x| = x$ under the constraint $x \geq 0$, so problem (31) can be written as

$$\min_{x \geq 0} \frac{1}{2}(x-(a-\lambda))^2. \tag{32}$$

The optimal solution can be easily obtained by $x = (a - \lambda)_+$.

The problem (29) is convex, so we can obtain the globally optimal solution to the problem (29). When all the elements in data D are observed, X_Ω and D_Ω become X and D, respectively, and thus, the problem is converted from robust matrix completion to robust principal component analysis (PCA) [35]. Therefore, robust PCA is a special case of our method.

In the second case, we solve the following problem:

$$\min_X J = \|X_\Omega - D_\Omega\|_1 + \gamma \mathrm{rank}(X). \tag{33}$$

If we define $0^0 = 0$, then the $\mathrm{rank}(X)$ is the Schatten 0-norm of X, and Theorem 2 also holds when $p = 0$. Thus, the problem (33) can also be solved using ALM method. In this case, the problem (22) becomes

$$\min_{x \geq 0} \frac{1}{2}(x-a)^2 + \lambda |x|^0 , \tag{34}$$

Only $x = 0$ and $x = a$ are possible to be the optimal solution to the problem (34), so the optimal solution to problem (34) can be easily obtained by

$$\begin{cases} a \leq \sqrt{2\lambda} & x^* = 0 \\ a > \sqrt{2\lambda} & x^* = a \end{cases} \tag{35}$$

5 Experiments

In this section, we empirically evaluate the proposed method in the matrix completion task on both synthetic data and two real-world applications of collaborative filtering and link discovery in social networks.

For simplicity, the regularization parameter γ in (5) is set to 1 in all our experiments.

5.1 Numerical results on synthetic data

To demonstrate the practical applicability of the proposed method for recovering low-rank matrices from their entries, we first perform the following numerical experiments. Following [4], for each (n, r, q) triplet, where n (we set $m = n$) is the matrix dimension, r is the predetermined rank, and q is the number of known entries, we experiment with the following procedures. We generate $M = M_L M_R^T$ as suggested in [4], where M_L and M_R are $n \times r$ matrices with i.i.d. standard Gaussian entries. We then select a subset Ω of q elements uniformly at random from $\{(i, j) : i = 1, \ldots, n, j = 1, \ldots, n\}$ as known entries, and our goal is to recover the rest entries of M given the incomplete input matrix.

The stopping criterion we use for our algorithm in all our experiments is as follows:

$$\frac{\left\| X^{(k)} - X^{(k-1)} \right\|_F}{\max(\left\| X^k \right\|_F, 1)} \leq \text{Tol}, \tag{36}$$

where Tol is a moderately small number. In our experiments, we set Tol $= 10^{-4}$.

We measure the accuracy of the computed solution X_{sol} of our algorithm by the relative error (RE) [9], which is a widely used metric in matrix completion and defined as follows:

$$\text{RE} := \frac{\|X_{\text{sol}} - M\|_F}{\|M\|_F}, \tag{37}$$

where M is the original matrix created in the above process.

5.1.1 Study of parameter p

Because p in (5) is the most important parameter of the proposed method, we first investigate its impact on our model. We vary the value of p in the range of $\{0.1, 0.2 \ldots, 1\}$ and perform incomplete matrix recovery as described above. For each value of p, we repeat the experiment for 50 times and report the average relative error in Fig. 3, from which we can see that the matrix recovery performance increases when the value of p decreases. This result clearly justifies the usefulness of the proposed method to introduce $p(< 1)$-norm in the proposed objective.

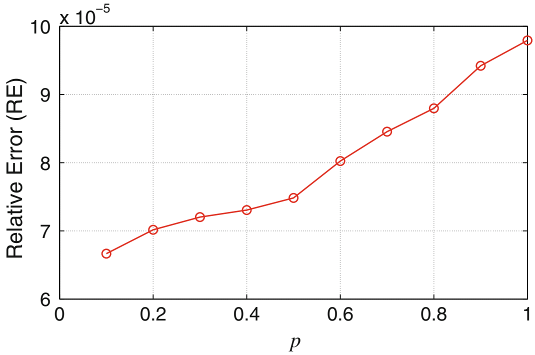


Fig. 3 Matrix recovery performance of the proposed method with different values of p

5.1.2 Comparison with other matrix completion methods on noiseless data

In order to demonstrate the effectiveness of the proposed method, we compare the performance of the proposed method against the following two matrix completion methods: fixed point continuation (FPC) method [12] and accelerated proximal gradient singular value thresholding (APG) method [9], which are the most recent methods and have demonstrated superior performances. We implement these two methods using the codes published by the respective authors and set up their parameters following the same settings as in [9]. In order for a fair comparison, we perform our experiments using the procedures described above with the same (n, r, q) triplet settings as in [9]. For each triplet setting, we repeat the experiment for 50 times and report the average performance in Table 1. The average number of iterations (denoted as iter) is also reported in Table 1, as well as the ratio (denoted by q/d_r) between the number of known entries and the degree of freedom of an $n \times n$ matrix of rank r. Following [8], the degree of an $n \times n$ matrix of rank r depends on $d_r = r(2n - r)$ degrees of freedom. As can be seen, q is selected to be 3, 4, and 5 times of the degrees of freedom of the corresponding input matrices.

From Table 1, we can see that the proposed method achieves more accurate results for matrix recovery than those delivered by the two compared methods. Moreover, our method uses substantially less iterations than the other two methods. These results clearly demonstrate the effectiveness of the proposed method in incomplete matrix recovery in terms of both quality and speed.

5.1.3 Comparison with other matrix completion methods on noisy data

Besides performing matrix completion on noiseless data, we also evaluate the proposed method on noisy data. Following [9], given a matrix M created by the aforementioned procedures, we corrupt it by a noise matrix N whose element are i.i.d. standard Gaussian variables. Then, we carry out the same procedures as before for matrix completion on $M + \sigma N$, where $\sigma = \mathrm{nf}\frac{\|M\|_F}{\|N\|_F}$ and nf is a given noise factor. We set nf = 0.1. Same as before, the experiment for each (n, r, q) triplet setting is repeated for 50 times, and the average results are reported in Table 2.

Table 1 Matrix completion performances of the compared methods on noiseless data

Unknown M			FPC		APG		Our method	
n/r	q	q/d_r	Iter	Relative error	Iter	Relative error	Iter	Relative error
100/10	5,666	3	439	1.08e−3	78	1.59e−4	26	7.47e−5
200/10	15,665	4	496	4.66e−4	74	1.19e−4	25	6.17e−5
500/10	49,471	5	491	5.92e−4	76	9.86e−5	27	5.34e−5

Table 2 Matrix completion performances of the compared methods on noisy data

Unknown M			FPC		APG		Our method	
n/r	q	q/d_r	Iter	Relative error	Iter	Relative error	Iter	Relative error
100/10	5,666	3	442	2.45e−2	81	2.36e−3	24	6.39e−4
200/10	15,665	4	486	6.61e−3	77	3.21e−3	23	5.15e−4
500/10	49,471	5	488	8.81e−3	73	2.21e−3	21	4.92e−4

Again, the proposed method performs the best. Most importantly, the relative errors of our method are smaller than the noise level (nf = 0.1), which is consistent with (or even more accurate than) the theoretical results established in [36], and further confirm the correctness of our method.

5.2 Improved collaborative filtering by our method

Collaborative filtering is an important topic in data mining and has been widely used in recommendation system, which aims to predict unknown users' opinions to a set of items upon those known and is often formalized as a matrix completion problem [37]. In this section, we evaluate the proposed method in the task of collaborative filtering.

5.2.1 Data sets

We perform our experiments using the following data sets.

The MovieLens data contain 10,000,054 ratings and 95,580 tags applied to 10,681 movies by 71,567 users of the online movie recommender service MovieLens, which has been filtered and refined by GroupLens lab[2] as three data sets with the following characteristics

1. Movie-100K: 100,000 ratings for 1,682 movies by 943 users;
2. Movie-1M: 1 million ratings for 3,900 movies by 6,040 users;
3. Movie-10M: 10 million ratings for 10,681 movies by 71,567 users.

In addition, we also experiment with **Epinion** data.[3] In Epinion.com, users can assign products or reviewers integer ratings. These ratings and reviews will influence future users when they are deciding whether a product is worth buying or a movie is worth watching. The data set contains 2,671 users and 1,375 items with 75,308 ratings.

5.2.2 Evaluation metric

In collaborative filtering, some entries of the input matrix are missing; therefore, we cannot compute the relative error of the estimated output matrix as we did in Sect. 5.1. Instead, we compute the normalized mean absolute error (NMAE) as in [12,38]:

$$\mathrm{NMAE} = \frac{\sum_{(i,j)\in\Gamma} |M_{ij} - X_{ij}|}{|\Gamma|(r_{\max} - r_{\min})}, \tag{38}$$

where M_{ij} denotes the rating given by user i to item j, X_{ij} denotes the predicted rating given by user i to item j, and $r_{\max}$ and $r_{\min}$ are the upper and lower bounds of the ratings. Because the user ratings in all the data sets range from 1 to 5, we have $r_{\min} = 1$ and $r_{\max} = 5$.

5.2.3 Experimental results on noiseless data

We first experiment with noiseless data. For each data set, we randomly select 20 and 50 % ratings as known samples, and our task is to recovery the rest ratings from the incomplete input matrices. Besides comparing to the two matrix completion methods used in Sect. 5.1, we also compare the results of our method against two state-of-the-art collaborative filter methods: probabilistic matrix factorization (PMF) method and weighted nonnegative matrix

[2] http://www.grouplens.org/.

[3] http://www.trustlet.org/wiki/Downloaded_Epinions_dataset.

Table 3 Performance of the compared methods measured by NMAE in collaborative filtering

Method	Movie-100K	Movie-1M	Movie-10M	Epinion
FPC	2.49e−1	2.53e−1	2.38e−1	3.15e−1
APG	1.94e−1	1.96e−1	1.89e−1	2.37e−1
PMF	2.26e−1	2.32e−1	2.21e−1	2.75e−1
WNMF	2.31e−1	2.42e−1	2.36e−1	3.07e−1
Ours ($p = 1$)	1.81e−1	1.89e−1	1.78e−1	2.23e−1
Ours ($p = 0.1$)	8.92e−2	9.04e−2	8.09e−2	1.75e−1
Ours (Schatten 0-norm)	7.94e−1	7.02e−2	7.33e−2	1.79e−1
FPC	2.88e−1	2.96e−1	2.76e−1	3.68e−1
APG	2.25e−1	2.33e−1	2.21e−1	2.79e−1
PMF	2.59e−1	2.73e−1	2.58e−1	3.19e−1
WNMF	2.67e−1	2.78e−1	2.78e−1	3.58e−1
Ours ($p = 1$)	1.95e−1	2.01e−1	1.86e−1	2.36e−1
Ours ($p = 0.1$)	9.45e−2	9.49e−2	8.65e−2	1.86e−1
Ours (Schatten 0-norm)	8.17e−1	7.23e−2	7.54e−2	1.84e−1

20 % ratings are known as training samples
Top half: noiseless data; bottom half: noisy data

factorization (WNMF) method. The former uses a probabilistic model, while the latter is devised by extending nonnegative matrix factorization. Both of them have reported promising empirical results. We implement our method with three different settings: (1) We set $p = 1$ for both ℓ_p-norm and Schatten p-norm, which is denoted by "ours ($p = 1$)"; (2) we set $p = 0.1$ for both ℓ_p-norm and Schatten p-norm, which is denoted by "ours ($p = 0.1$)"; and (3) we set $p = 1$ for ℓ_p-norm and $p = 0$ for Schatten p-norm that solves (33), which is denoted by "Ours (Schatten 0-norm)." For each data set, we run each compared method for 20 times and report the average results in the top halves of Tables 3 and 4.

The results in the top halves of Tables 3 and 4 show that our method consistently outperforms the compared methods on the noiseless data, sometimes very significantly, which provide one more concrete evidence to support the advantage of the proposed method. Moreover, as can be seen in the top halves of Tables 3 and 4, the results of our method when $p = 0$ and $p = 0.1$ for the Schatten p-norm are much better than those when $p = 1$. This observation is in accordance with our earlier theoretical analysis in that the smaller the value of p is, the better the Schatten p-norm approximates the matrix rank, which is particularly true when $p = 0$ to approximate the rank minimization.

5.2.4 Experimental results on noisy data

Due to the usage of the ℓ_p-norm ($0 \leq p \leq 1$) in the error function, the proposed objective in (5) is robust against outliers. To evaluate this, we add Gaussian noise following the same way as in Sect. 5.1 where we set the noise factor nf = 0.1. We repeat the same experimental procedures as those on the noiseless data shortly before and report the results in the bottom halves of Tables 3 and 4. From the results, we can see that the proposed method with different parameter settings is still better than the competing methods. Moreover, compared with the results on the noiseless data, the performance degradations of our new method are much less than the competing methods, which, again, is consistent with our motivations and the

Table 4 Performance of the compared methods measured by NMAE in collaborative filtering

Method	Movie-100K	Movie-1M	Movie-10M	Epinion
FPC	2.09e−1	2.13e−1	1.98e−1	2.45e−1
APG	1.74e−1	1.84e−1	1.77e−1	2.13e−1
PMF	2.16e−1	2.22e−1	2.11e−1	2.55e−1
WNMF	2.04e−1	2.02e−1	1.95e−1	2.48e−1
Ours ($p = 1$)	1.66e−1	1.71e−1	1.53e−1	2.03e−1
Ours ($p = 0.1$)	8.14e−2	8.36e−2	7.94e−2	1.84e−1
Ours (Schatten 0-norm)	7.51e−2	7.99e−2	7.85e−2	1.31e−1
FPC	2.42e−1	2.55e−1	2.35e−1	2.84e−1
APG	2.02e−1	2.12e−1	2.05e−1	2.51e−1
PMF	2.57e−1	2.61e−1	2.51e−1	2.95e−1
WNMF	2.35e−1	2.38e−1	2.28e−1	2.87e−1
Ours ($p = 1$)	1.79e−1	1.81e−1	1.62e−1	2.13e−1
Ours ($p = 0.1$)	8.54e−2	8.86e−2	8.33e−2	1.95e−1
Ours (Schatten 0-norm)	7.73e−2	8.22e−2	8.08e−2	1.34e−1

50 % ratings are known as training samples
Top half: noiseless data; bottom half: noisy data

mathematical formulation of our new method in that the ℓ_p-norm loss function is more robust to noises and outliers when p is small.

5.3 Improved link discovery on social networks by our method

Link discovery on social graphs, which explores the relationships between users, plays a central role in understanding the structure of related social communities. Because most users on a social network only know a very small fraction of users and tag even fewer explicitly, the resulted social graphs are sparse and link discovery is necessary to mine more useful information to better understand a community. In this section, we evaluate the proposed matrix completion method by exploring link discovery problem on social networks.

5.3.1 Data sets

We evaluate the performance of our method using the **Wikipedia 2** [39] and **Slashdot 3** [40] data set. The former contains more than 7,000 users with 103,000 trust links, and the latter contains about 80,000 users with 900,000 trust links. The link coverage of these two graphs are as low as 0.21 and 0.01 %; therefore, they are very sparse and skewed due to the domination of the noninteracting user pairs. To alleviate the data skewness for fair comparison, we select top 2,000 highest degree users from each data set for experiments.

5.3.2 Experimental setups

The goal of our method is to infer the unobservable links in the network. However, due to the lack of ground truth, we have to hide existing links to simulate missing one. In this paper, we emulate to hide 90 % entries and do the imputation based on the remaining 10 % available information. The reason we hide large percentage of entries is to simulate the fact that most users from social Web sites such as Facebook and Linkedin, according to our observation, only explicitly express trust and distrust to a small fraction of peer users considering total number of users.

In order to make prediction, we need a threshold. Empirically, we select the mean of the available entries as the threshold to convert the prediction into the binary matrix.

5.3.3 Experimental results on noiseless data

Again we first experiment with the noiseless data. Besides comparing the matrix completion methods as in previous subsections, we also compare our method to two link prediction methods that are widely used in the studies of social networks, including common neighbors (CN) method [41] and SVD method [42]. The settings of the matrix completion methods including ours are same as before. For SVD method, we fine-tune the rank by searching the grid of $\{100, 200, \ldots, 1000\}$.

We evaluate the compared methods for two standard performance metrics broadly used in statistical learning, including precision and recall. The results of the compared methods on the two data sets are reported in the top half of Table 5. A first glance at the results shows that the proposed methods again are superior to other compared methods, which demonstrate their effectiveness in the task of link discovery on social networks. Moreover, when $p = 0$ and $p = 0.1$, our method achieved better results than those when $p = 1$, which once again validates the usage of small p for Schatten p-norm for matrix completion.

5.3.4 Experimental results on noisy data

Now, we experiment with noisy data to evaluate the robustness of the proposed method. We randomly split the entries of the input matrix into three parts: We hide the values of 85 % entries, assign the opposite values to 5 % entries to emulate noise, and keep the values of the remaining 10 % entries. We do the imputation and report the experimental results in the bottom half of Table 5. Again, we can see that our new method is superior to the other method, and the performance degradations due to outliers are much less than the competing methods, which clearly demonstrate the advantage of the proposed method in social link discovery.

6 Conclusions

In this paper, we proposed a new robust matrix completion method using joint Schatten p-norm and ℓ_p-norm ($0 < p \leq 1$). When $p \to 0$, the Schatten p-norm-based objective can approximate the rank minimization problem much better than the standard trace norm minimization to achieve better matrix completion results. The ℓ_p-norm-based error function enhances the robustness of the proposed objective. Both Schatten p-norm and ℓ_p-norm are nonsmooth terms. To solve this difficult optimization problem, we derive the algorithm based on the alternating direction method. Extensive experiments show that under arbitrarily random initializations, our new method can always get better matrix completion results without introducing much extra computational cost. The extensive experiments were performed on both synthetic and real-world applications (collaborative filtering and social network link prediction) data. All empirical results demonstrate the effectiveness of the proposed approach.

Table 5 Performance comparison for the task on link discovery on social networks

Method	Wikipedia 2		Slashdot 3	
	Precision	Recall	Precision	Recall
CN	0.071	0.205	0.058	0.149
SVD	0.088	0.211	0.064	0.166
FPC	0.107	0.244	0.084	0.189
APG	0.114	0.251	0.089	0.194
Ours ($p = 1$)	0.135	0.301	0.101	0.224
Ours ($p = 0.1$)	0.142	0.322	0.112	0.245
Ours (Schatten 0-norm)	0.151	0.337	0.134	0.268
CN	0.060	0.168	0.048	0.122
SVD	0.071	0.173	0.051	0.134
FPC	0.088	0.204	0.069	0.158
APG	0.091	0.203	0.073	0.159
Ours ($p = 1$)	0.126	0.285	0.095	0.208
Ours ($p = 0.1$)	0.134	0.302	0.106	0.230
Ours (Schatten 0-norm)	0.146	0.326	0.130	0.260

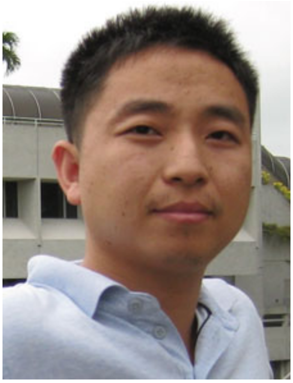
Top half: noiseless data; bottom half: noisy data

Acknowledgments This research was partially supported by NSF DMS-0915228, IIS-1117965, IIS-1302675, and IIS-1344152.

References

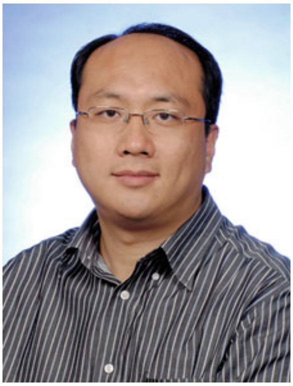
1. Srebro N, Rennie J, Jaakkola T (2004) Maximum margin matrix factorization. Conf Neural Inf Process Syst (NIPS) 17:1329–1336
2. Rennie J, Srebro N (2005) Fast maximum margin matrix factorization for collaborative prediction. In: International conference on machine learning (ICML)
3. Abernethy J, Bach F, Evgeniou T, Vert JP (2009) A new approach to collaborative filtering: operator estimation with spectral regularization. J Mach Learn Res (JMLR) 10:803–826
4. Candès E, Recht B (2009) Exact matrix completion via convex optimization. Found Comput Math 9(6):717–772
5. Candes EJ, Tao T (2009) The power of convex relaxation: near-optimal matrix completion. IEEE Trans Inform Theory 56(5):2053–2080
6. Recht B, Fazel M, Parrilo PA (2010) Guaranteed minimum rank solutions of linear matrix equations via nuclear norm minimization. SIAM Rev 52(3):471–501
7. Mazumder R, Hastie T, Tibshirani R (2010) Spectral regularization algorithms for learning large incomplete matrices. J Mach Learn Res (JMLR) 11:2287–2322
8. Cai J-F, Candes EJ, Shen Z (2008) A singular value thresholding algorithm for matrix completion. SIAM J Opt 20(4):1956–1982
9. Toh K, Yun S (2010) An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. Pac J Opt 6:615–640
10. Ji S, Ye Y (2009) An accelerated gradient method for trace norm minimization. In: International conference on machine learning (ICML)
11. Liu Y-J, Sun D, Toh K-C (2012) An implementable proximal point algorithmic framework for nuclear norm minimization. Math Program 133(1–2):399–436
12. Ma S, Goldfarb D, Chen L (2011) Fixed point and Bregman iterative methods for matrix rank minimization. Math Program 128(1–2):321–353
13. Recht B (2011) A simpler approach to matrix completion. J Mach Learn Res 12:3413–3430
14. Vishwanath S (2010) Information theoretic bounds for low-rank matrix completion. In: IEEE international symposium on information theory proceedings (ISIT), pp 1508–1512

15. Koltchinskii V, Lounici K, Tsybakov A (2011) Nuclear norm penalization and optimal rates for noisy low rank matrix completion. Ann Stat 39:2302–2329
16. Salakhutdinov R, Srebro N (2010) Collaborative filtering in a non-uniform world: learning with the weighted trace norm. Adv Neural Inf Process Syst (NIPS) 23:1–8
17. Pong TK, Tseng P, Ji S, Ye J (2010) Trace norm regularization: reformulations, algorithms, and multi-task learning. SIAM J Opt 20(6):3465–3489
18. Nie F, Wang H, Cai X, Huang H, Ding C (2012) Robust matrix completion via joint schatten p-norm and l_p-norm minimization. In: IEEE international conference on data mining (ICDM), pp 566–574
19. Huang J, Nie F, Huang H, Lei Y, Ding C (2013) Social trust prediction using rank-k matrix recovery. In: 23rd international joint conference on artificial intelligence
20. Huang J, Nie F, Huang H (2013) Robust discrete matrix completion. In: Twenty-seventh AAAI conference on artificial intelligence (AAAI-13), pp 424–430
21. Tan VY, Balzano L, Draper SC (2011) Rank minimization over finite fields. In: IEEE international symposium on information theory proceedings (ISIT), pp 1195–1199
22. Meka R, Jain P, Dhillon IS (2010) Guaranteed rank minimization via singular value projection. In: Conference on neural information processing systems (NIPS)
23. Gabidulin EM (1985) Theory of codes with maximum rank distance. Problemy Peredachi Informatsii 21(1):3–16
24. Loidreau P (2008) Properties of codes in rank metric. In: Eleventh international workshop on algebraic and combinatorial coding theory, pp 192–198
25. Fazel M, Hindi H, Boyd SP (2001) A rank minimization heuristic with application to minimum order system approximation. IEEE Am Control Conf 6:4734–4739
26. Fazel M, Hindi H, Boyd SP (2003) Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices. IEEE Am Control Conf 3:2156–2162
27. Blomer J, Karp R, Welzl E (1997) The rank of sparse random matrices over finite fields. Random Struct Algorithms 10(4):407–420
28. Ma S, Goldfarb D, Chen L (2011) Fixed point and Bregman iterative methods for matrix rank minimization. Math Program 128(1–2):321–353
29. Nie F, Huang H, Ding CHQ (2012) Low-rank matrix recovery via efficient schatten p-norm minimization. In: AAAI conference on artificial intelligence
30. Nie F, Huang H, Cai X, Ding C (2010) Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization. In: Conference on neural information processing systems (NIPS)
31. Powell MJD (1969) A method for nonlinear constraints in minimization problems. In: Fletcher R (ed) Optimization. Academic Press, London
32. Hestenes MR (1969) Multiplier and gradient methods. J Opt Theory Appl 4:303–320
33. Bertsekas DP (1996) Constrained optimization and lagrange multiplier methods. Athena Scientific, Belmont
34. Gabay D, Mercier B (1969) A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput Math Appl 2(1):17–40
35. Wright J, Ganesh A, Rao S, Ma Y (2009) Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization. In: The proceedings of the conference on neural information processing systems. pp 1–9
36. Candes E, Plan Y (2010) Matrix completion with noise. Proc IEEE 98(6):925–936
37. Salakhutdinov R, Mnih A (2008) Probabilistic matrix factorization. Adv Neural Inf Process Syst (NIPS) 20:1257–1264
38. Gu Q, Zhou J, Ding C (2010) Collaborative filtering: weighted nonnegative matrix factorization incorporating user and item graphs. In: Siam data mining conference
39. Leskovec J, Huttenlocher D, Kleinberg J (2010) Predicting positive and negative links in online social networks. In: International world wide web conference (WWW). ACM, pp 641–650
40. Leskovec J, Lang K, Dasgupta A, Mahoney M (2009) Community structure in large networks: natural cluster sizes and the absence of large well-defined clusters. Int Math 6(1):29–123
41. Newman M (2001) Clustering and preferential attachment in growing networks. Phys Rev E 64(2):025102
42. Billsus D, Pazzani M (1998) Learning collaborative information filters. In: International conference on machine learning (ICML), pp 46–54



Feiping Nie received his B.S. degree in Computer Science from North China University of Water Conservancy and Electric Power, China in 2000, received his M.S. degree in Computer Science from Lanzhou University, China in 2003, and received his Ph.D. degree in Computer Science from Tsinghua University, China in 2009. Currently, he is a research assistant professor at the University of Texas, Arlington, USA. His research interests include machine learning and its application fields, such as pattern recognition, data mining, computer vision, image processing and information retrieval.

Hua Wang is an Assistant Professor at the Department of Electrical Engineering and Computer Science of Colorado School of Mines. He received the Ph.D. degree in Computer Science from the University of Texas at Arlington in 2012. Before that, he received the Bachelor's degree from Tsinghua University, China and the Master's degree from Nanyang Technological University, Singapore in 1999 and 2003, respectively. His research interests include machine learning and data mining, as well as their applications in bioinformatics, health informatics, medical image analysis and computer vision.



Heng Huang received the B.S. degree in Automation and the M.S. degree in Information Measurement Technology and Instruments from Shanghai Jiao Tong University, Shanghai, China, in 1997 and 2001, respectively. He received the Ph.D. degree in Computer Science from Dartmouth College in 2006. Since 2012, he has been an associate professor in computer science and engineering department at University of Texas at Arlington. From 2007 to 2012, he was an assistant professor at the same department. His research interests include machine learning, data mining, bioinformatics, health informatics, medical image analysis, and computer vision.

Chris Ding received his Ph.D. degree in Physics from Columbia University in 1987. He has been postdoc researcher in California Institute of Technology (1987-1993) and Jet Propulsion Laboratory (1993-1996). He was a scientist in Lawrence Berkeley National Laboratory from 1996 to 2007. He joined University of Texas at Arlington as a professor in 2007. His main research areas are machine learning, data mining, and bioinformatics. He also works on information retrieval, web link analysis, and high performance computing.