2012 IEEE 12th International Conference on Data Mining

Robust Matrix Completion via Joint Schatten p-Norm and ℓ_p-Norm Minimization

Feiping Nie*, Hua Wang†, Xiao Cai*, Heng Huang* and Chris Ding*
*Department of Computer Science and Engineering
University of Texas, Arlington, USA
Email: feipingnie@gmail.com, xiao.cai@mavs.uta.edu, {heng,chqding}@uta.edu
†Department of Electrical Engineering & Computer Science
Colorado School of Mines, Golden, Colorado 80401
Email: huawangcs@gmail.com

***Abstract*—The low-rank matrix completion problem is a fundamental machine learning 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 and social network link prediction. All empirical results show our new method outperforms the standard matrix completion methods.**

***Keywords*-recommendation system; matrix completion; low-rank matrix recovery; optimization;**

I. Introduction

The prediction of the incomplete observations of an evolving matrix is a challenge of interest in many machine learning applications [1], [2], [3], such as friendship prediction in social network, rating value estimation in recommendation system and collaborative filtering, link prediction in protein-protein interaction network. All these problem 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 [4], [5], [6], [7], [8], [2], 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 [9], [10], [11], [12], [7]. 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.

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 non-convex 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 our new robust matrix completion method outperforms the standard missing value prediction approaches. In summary, we highlight the main contributions of this paper as follows:

1, We propose a new and reasonable objective function for the robust matrix completion task.

2, Optimizing the objective function is a non-trivial problem, we derive an optimization algorithm to solve this problem.

3, We derive the optimal solution to the problem (20), which generalizes a famous soft thresholding result in [8], and can be used in many other Schatten p-norm minimization problems.

1550-4786/12 $26.00 © 2012 IEEE
DOI 10.1109/ICDM.2012.160

II. A New Robust Matrix Completion

A. 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 $\|v\|_p = (\sum_i^n |v_i|^p)^{\frac{1}{p}}$, 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 $\|v\|_p^p = \sum_i^n |v_i|^p$.

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. If we define $0^0 = 0$, then when $p = 0$, Eq. (2) is the rank of X.

B. Robust Matrix Completion Objective

We denote $X_\Omega = \{X_{ij} \,|(i,j) \in \Omega\}$, and $\|X_\Omega\|_p^p = \sum_{(i,j)\in\Omega} |X_{ij}|^p$. 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 solves the following problem:

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

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

$$\min_X \|X_\Omega - D_\Omega\|_2^2 + \gamma \|X\|_{S_1}. \tag{4}$$

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 [13]. In this paper, we replace the $\|X\|_{S_1}$ 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 ($0 < p \leq 1$) as the error function to improve the robustness to outliers in given data [14], and propose to solve the following robust matrix completion problem (we use the same p for two norms to avoid one more parameter):

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

[1]When $p \geq 1$, $\|\mathbf{v}\|_p = \left(\sum_{i=1}^n |v_i|^p\right)^{\frac{1}{p}}$ 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 $\|\mathbf{x}+\mathbf{y}\|_p \leq K\left(\|\mathbf{x}\|_p + \|\mathbf{y}\|_p\right)$ 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.

III. Proposed Algorithm

Solving the problem in Eq. (5) is challenge since both of the terms in Eq. (5) are non-smooth and the Schatten p-norm is somewhat intractable. We use the Augmented Lagrangian Method [15], [16], [17] to solve this problem, and focus on the solutions to the related subproblems.

A. Brief Description of Augmented Lagrangian Method

Consider the constrained optimization problem:

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

The algorithm using the Augmented Lagrangian Method (ALM) to solve the problem (6) is described in Algorithm 1

It has been proved that under some rather general conditions, the Algorithm 1 converges Q-linearly to the optimal solution [17]. This property makes ALM very attractive.

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

Algorithm 1: Algorithm to solve the problem (6).

B. Solving Problem (5) Using ALM

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) [18] to solve this problem. Specifically, we optimize the problem with respect to one variable when fix

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

Algorithm 2: Algorithm to solve the problem (5).

the other two variables, which result in the following three subproblem.

When fix 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

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

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

The detailed algorithm to solve the problem in Eq. (5) is described in Algorithm 2.

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

C. 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.,

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

We can see that the gradient of $h(x)$ is

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

Note that $h(x)$ is a quadratic equation in one variable, its convexity can be easily analyzed. Denote a constant v as

$$\nu = \left(\frac{\lambda p(1-p)}{2}\right)^{\frac{1}{2-p}}. \tag{16}$$

The optimal solution to problem (13) can be obtained by

$$\begin{cases} g(\nu) \geq 0, g(-\nu) \leq 0 & x^* = 0 \\ g(\nu) < 0, g(-\nu) \leq 0 & x^* = \arg\min_{x\in\{0,x_1\}} h(x) \\ g(\nu) \geq 0, g(-\nu) > 0 & x^* = \arg\min_{x\in\{0,x_2\}} h(x) \\ g(\nu) < 0, g(-\nu) > 0 & x^* = \arg\min_{x\in\{0,x_1,x_2\}} h(x) \end{cases} \tag{17}$$

where $x_1 = 0$ and $x_2 = 0$ the root of $g(x) = 0$, which can be easily obtained with Newton method initialized at 2ν and -2ν, respectively.

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

$$\min_{x\geq 0} \frac{1}{2}(x-a)^2 + \lambda|x|^p, \tag{18}$$

the optimal solution to problem (18) can be obtained by

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

D. Solving the Subproblem (12)

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

$$\min_X \frac{1}{2}\|X - A\|_F^2 + \lambda\|X\|_{S_p}^p \tag{20}$$

Denote the SVD of X by $X = U\Delta V^T$. We prove in the Appendix that for the optimal solution X, U and V are the left and right singular vector of A respectively, and the i-th singular value δ_i is the optimal solution of the following problem:

$$\min_{\delta_i \geq 0} \frac{1}{2}(\delta_i - a_i)^2 + \lambda\delta_i^p \tag{21}$$

where a_i is the i-th singular value of A. The optimal solution to the problem (21) can be obtained according to Eq. (19). **It is interesting to see that when $p = 1$, the derived solution is exactly the same as in [8], and our result extend the result in [8] to the case of $0 < p < 1$.**

IV. EXPERIMENTS

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

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

A. 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:

$$\frac{\left\|X^{(k)} - X^{(k-1)}\right\|_F}{\max\left(\|X^k\|_F, 1\right)} \leq \text{Tol} \ , \tag{22}$$

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

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

$$\text{RE} := \|X_{\text{sol}} - M\|_F / \|M\|_F \ , \tag{23}$$

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

Study of parameter p. Because p in Eq. (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 Figure 1, 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. Upon this preliminary result, unless otherwise specified, we will set $p = 0.1$ in all subsequent experiments.

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 setup their parameters using 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 I. The average number of iterations (denoted as iter) is also reported in Table I, as well as the ratio (denoted by q/d_r) between the number of known entries and the degree 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 I, we can see that the proposed method achieves more accurate 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.

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 $\sigma = nf \frac{\|M\|_F}{\|N\|_F}$ 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 II.

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 [19] and further confirm the correctness of our method.

B. 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 [20]. In this section, we evaluate the proposed method in the task of collaborative filtering.

Data sets. We perform our experiments using the following data sets.

The MovieLens data contains 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

Table I
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	5666	3	439	1.08e-3	78	1.59e-4	26	7.47e-5
200/10	15665	4	496	4.66e-4	74	1.19e-4	25	6.17e-5
500/10	49471	5	491	5.92e-4	76	9.86e-5	27	5.34e-5

Table II
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	5666	3	442	2.45e-2	81	2.36e-3	24	6.39e-4
200/10	15665	4	486	6.61e-3	77	3.21e-3	23	5.15e-4
500/10	49471	5	488	8.81e-3	73	2.21e-3	21	4.92e-4

Table III
PERFORMANCE OF THE COMPARED METHODS MEASURED BY NMAE IN COLLABORATIVE FILTERING. TOP: 20% RATINGS ARE KNOWN AS TRAINING SAMPLES; BOTTOM: 50% RATINGS ARE KNOWN AS TRAINING SAMPLES.

Data	FPC	APG	PMF	WNMF	Our ($p = 1$)	Our ($p = 0.1$)
movie-100K	2.49e-01	1.94e-01	2.26e-01	2.31e-01	1.81e-01	**8.92e-2**
movie-1M	2.53e-01	1.96e-01	2.32e-01	2.42e-01	1.89e-01	**9.04e-2**
movie-10M	2.38e-01	1.89e-01	2.21e-01	2.36e-01	1.78e-01	**8.09e-2**
Epinion	3.15e-01	2.37e-01	2.75e-01	3.07e-01	2.23e-01	**1.75e-1**
movie-100K	2.09e-01	1.74e-01	2.16e-01	2.04e-01	1.66e-01	**8.14e-2**
movie-1M	2.13e-01	1.84e-01	2.22e-01	2.02e-01	1.71e-01	**8.36e-2**
movie-10M	1.98e-01	1.77e-01	2.11e-01	1.95e-01	1.53e-01	**7.94e-2**
Epinion	2.45e-01	2.13e-01	2.55e-01	2.48e-01	2.03e-01	**1.84e-1**

has been filtered and refined by GroupLens lab[2] as three data sets with the following characteristics (1) **movie-100K**: 100,000 ratings for 1682 movies by 943 users; (2) **movie-1M**: 1 million ratings for 3900 movies by 6040 users; (3) **movie-10M**: 10 million ratings for 10681 movies by 71567 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 2671 users and 1375 items with 75308 ratings.

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 Section IV-A. Instead, we compute the Normalized Mean Absolute Error (NMAE) as in [12], [21]:

$$\text{NMAE} = \frac{\sum_{(i,j)\in\Gamma} |M_{ij} - X_{ij}|}{|\Gamma|\,(r_{\max} - r_{\min})}\ , \tag{24}$$

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

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

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

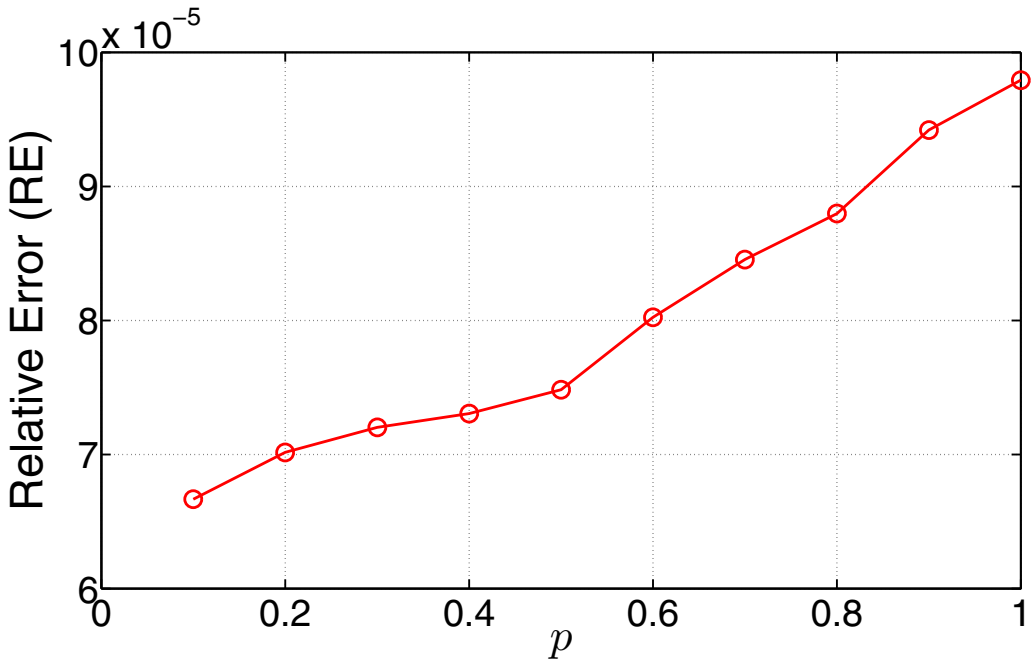

Figure 1. Matrix recovery performance of the proposed method with different values of p.

5, we have $r_{\min} = 1$ and $r_{\max} = 5$.

Experimental results. 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 Section IV-A, 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 Factorization (WNMF) method. The former uses probabilistic model, while the latter is devised by extending nonnegative matrix factorization. Both of them have reported promising empirical results. We implement our method for two different settings of $p = 1$ and $p = 0.1$. For each data set, we run each compared method for 20 times and report the average results in Table III.

The results in Table III show that our method consistently outperforms the compared methods, sometimes very significantly, which provide one more concrete evidence to support the advantage of the proposed method. Moreover, as can be seen in Table III, the results of our method when $p = 0.1$ is much better than those when $p = 1$. This observation is in accordance with our theoretical analysis in that, the smaller the value of p is, the better the Schatten p-norm approximates the matrix rank; and the smaller the vale of p is, the more robust of our loss function is against outliers.

C. 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.

Data sets. We evaluate the performance of our method using the **Wikipedia 2** [22] and **Slashdot 3** [23] 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 the domination of the non-interacting user pairs. To alleviate the data skewness for fair comparison, we select top 2,000 highest degree users from each data set for experiments.

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%

Table IV
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
Our ($p = 1$)	0.135	0.301	0.101	0.224
Our ($p = 0.1$)	**0.142**	**0.322**	**0.112**	**0.245**

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 as the mean of the available entries as the threshold to convert the prediction into the binary matrix.

Experimental results. Besides comparing the matrix completion methods as in previous subsections, we also compare our method to two link prediction methods which are widely used in the studies of social networks, including common neighbors (CN) method [24] and SVD method [25]. 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 $\{100, 200, \ldots, 1,000\}$.

We evaluate the compared methods 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 Table IV. A first glance at Table IV shows that the proposed methods again is superior to other compared methods, which demonstrate their effectiveness in the task of link discovery on social networks. Moreover, when $p = 0.1$, our method achieved better results than those when $p = 1$, which once again validate the usage of small p for matrix completion.

V. 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 non-smooth 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.

APPENDIX

Denote $\sigma(A)$ as the ordered eigenvalue matrix of A, i.e. a diagonal matrix with the diagonal elements being the ordered eigenvalues of A. We have the following two results:

Theorem 1 (von Neumann): For any two matrices $A, B \in \mathbb{R}^{m\times n}$, $tr(A^T B) \leq tr(\sigma(A)^T \sigma(B))$.

Corollary 1: For any orthonormal matrices $Q, R \in \mathbb{R}^{n\times n}$, and any diagonal matrices $\Sigma, \Lambda \in \mathbb{R}^{n\times n}$, where the diagonal elements of Σ, Λ are ordered with the same order, we have $tr(\Sigma Q \Lambda R) \leq tr(\Sigma\Lambda)$.

In the following, we suppose all the eigenvalue matrices in compact SVD are the ordered eigenvalue matrix with the same order.

Suppose the compact SVD of X is $X = U\Delta V^T$, then $\|X\|_{S_p}^p = tr\Delta^p$. Denote the objective function in the problem (20) by $f(X)$, i.e.,

$$f(X) = f(U, \Delta, V) = \frac{1}{2}\left\|U\Delta V^T - A\right\|_F^2 + \lambda tr(\Delta^p) \quad (25)$$

As U and V are the left and right singular vectors of X respectively, they are both orthogonal matrices, i.e., $U^T U = I$ and $V^T V = I$. Denote the Lagrangian function of the problem (20) by

$$\begin{aligned}\mathcal{L}(U, \Delta, V, \eta, \omega) &= f(U, \Delta, V) - tr(\eta^T(U^T U - I)) \\ &\quad -tr(\omega^T(V^T V - I)). \end{aligned} \quad (26)$$

By setting the derivative of $\mathcal{L}(U, \Delta, V, \eta, \omega)$ with respect to U and V respectively, we have the following two equations:

$$AV\Delta - 2U\eta^T = 0 \quad (27)$$

$$A^T U\Delta - 2V\omega^T = 0 \tag{28}$$

According to Eq. (28), we have

$$\begin{aligned} & AA^T U\Delta - 2AV\omega^T = 0 \\ \Rightarrow\ & AA^T U\Delta - 4U\eta^T\Delta^{-1}\omega^T = 0 && (29)\\ \Rightarrow\ & AA^T U = 4U\eta^T\Delta^{-1}\omega^T\Delta^{-1} && (30)\\ \Rightarrow\ & U^T AA^T U = 4\eta^T\Delta^{-1}\omega^T\Delta^{-1} && (31)\end{aligned}$$

where Eq. (29) holds according to Eq. (27). From Eq. (31) we know $4\eta^T\Delta^{-1}\omega^T\Delta^{-1}$ must be symmetrical. Suppose the eigenvalue decomposition $4\eta^T\Delta^{-1}\omega^T\Delta^{-1} = Q\Pi Q^T$, then according to Eq. (30) we have

$$\begin{aligned} & AA^T U = UQ\Pi Q^T \\ \Rightarrow\ & AA^T UQ = UQ\Pi \end{aligned} \tag{32}$$

Therefore, UQ should be the left singular vectors of A.

On the other hand, According to Eq. (27), we have

$$\begin{aligned} & A^T AV\Delta - 2A^T U\eta^T = 0 \\ \Rightarrow\ & A^T AV\Delta - 4V\omega^T\Delta^{-1}\eta^T = 0 && (33)\\ \Rightarrow\ & A^T AV = 4V\omega^T\Delta^{-1}\eta^T\Delta^{-1} && (34)\\ \Rightarrow\ & V^T A^T AV = 4\omega^T\Delta^{-1}\eta^T\Delta^{-1} && (35)\end{aligned}$$

where Eq. (33) holds according to Eq. (28). From Eq. (35) we know $4\omega^T\Delta^{-1}\eta^T\Delta^{-1}$ must be symmetrical. Suppose the eigenvalue decomposition $4\omega^T\Delta^{-1}\eta^T\Delta^{-1} = R\Pi R^T$, then according to Eq. (34) we have

$$\begin{aligned} & A^T AV = VR\Pi R^T \\ \Rightarrow\ & A^T AVR = VR\Pi \end{aligned} \tag{36}$$

Therefore, VR should be the right singular vectors of A.

Denote the compact SVD of A by $A = UQ\Pi_1 R^T V^T$, then we have

$$\begin{aligned} & f(U,\Delta,V) = \tfrac{1}{2}\left\|U\Delta V^T - UQ\Pi_1 R^T V^T\right\|^2 + \lambda tr\Delta^p \\ & = \sum_i (\tfrac{1}{2}\delta_i^2 + \lambda\delta_i^p) - Tr(\Delta Q\Pi_1 R^T) \\ & \geq \sum_i (\tfrac{1}{2}\delta_i^2 + \lambda\delta_i^p) - Tr(\Delta\Pi_1) \\ & = \sum_i (\tfrac{1}{2}(\delta_i - a_i)^2 + \lambda\delta_i^p - \tfrac{1}{2}a_i^2) \end{aligned} \tag{37}$$

where $\delta_i \geq 0$ is the i-th diagonal element of Δ, a_i is the i-th diagonal element of Π_1, and the inequality holds according to Corollary 1. Therefore, minimizing $f(U,\Delta,V)$ is reduced to minimizing the following problem:

$$\min_{\delta_i\geq 0} \sum_i \frac{1}{2}(\delta_i - a_i)^2 + \lambda\delta_i^p, \tag{38}$$

which can be solved by solving the problem (21) for each i.

It is worth to mentioning that this work and the above proof were finished one year ago, now we find a more concise proof as follows. 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 eigenvalue matrices with the same order. then $\|X - A\|_F^2 = tr(\Delta^T\Delta) + tr(\Sigma^T\Sigma) - 2tr(X^T A) \geq tr(\Delta^T\Delta) + tr(\Sigma^T\Sigma) - 2tr(\Delta^T\Sigma) = \|\Delta - \Sigma\|_F^2$, where the inequality holds based on Theorem 1. Therefore minimizing $f(U,\Delta,V)$ is reduced to minimizing Eq.(38).

ACKNOWLEDGMENT

This research was partially supported by NSF-CCF 0830780, NSF-DMS 0915228, NSF-CCF 0917274, NSF-IIS 1117965.

REFERENCES

[1] N. Srebro, J. Rennie, and T. Jaakkola, "Maximummargin matrix factorization," *NIPS*, vol. 17, pp. 1329–1336, 2004.

[2] J. Rennie and N. Srebro, "Fast maximum margin matrix factorization for collaborative prediction," *ICML*, 2005.

[3] J. Abernethy, F. Bach, T. Evgeniou, and J. P. Vert, "A new approach to collaborative filtering: Operator estimation with spectral regularization," *JMLR*, vol. 10, pp. 803–826, 2009.

[4] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[5] E. J. Candes and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2053–2080, 2009.

[6] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[7] R. Mazumder, T. Hastie, and R. Tibshirani, "Spectral regularization algorithms for learning large incomplete matrices," *submitted to JMLR*, 2009.

[8] J.-F. Cai, E. J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2008.

[9] K. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pacific Journal of Optimization*, vol. 6, pp. 615–640, 2010.

[10] S. Ji and Y. Ye, "An accelerated gradient method for trace norm minimization," *ICML*, 2009.

[11] Y.-J. Liu, D. Sun, and K.-C. Toh, "An implementable proximal point algorithmic framework for nuclear norm minimization," *Optimization Online*, 2009.

[12] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and bregman iterative methods for matrix rank minimization," *Mathematical Programming*, 2009.

[13] F. Nie, H. Huang, and C. H. Q. Ding, "Low-rank matrix recovery via efficient schatten p-norm minimization," in *AAAI*, 2012.

[14] F. Nie, H. Huang, X. Cai, and C. Ding, "Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization," in *NIPS*, 2010.

[15] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*. In R. Fletcher, editor, Optimization. Academic Press, London and New York, 1969.

[16] M. R. Hestenes, "Multiplier and gradient methods," *Journal of Optimization Theory and Applications*, vol. 4, pp. 303–320, 1969.

[17] D. P. Bertsekas, *Constrained optimization and lagrange multiplier methods*. Athena Scientific, 1996.

[18] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers & Mathematics with Applications*, vol. 2, no. 1, pp. 17–40, 1969.

[19] E. Candes and Y. Plan, "Matrix completion with noise," *Arxiv preprint arXiv:0903.3131*, 2009.

[20] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in *NIPS*, 2008.

[21] Q. Gu, J. Zhou, and C. Ding, "Collaborative filtering: Weighted nonnegative matrix factorization incorporating user and item graphs," in *SDM*, 2010.

[22] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Predicting positive and negative links in online social networks," in *WWW*, 2010.

[23] J. Leskovec, K. Lang, A. Dasgupta, and M. Mahoney, "Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters," *Internet Mathematics*, vol. 6, no. 1, pp. 29–123, 2009.

[24] M. Newman, "Clustering and preferential attachment in growing networks," *Physical Review E*, vol. 64, no. 2, p. 025102, 2001.

[25] D. Billsus and M. Pazzani, "Learning collaborative information filters," in *ICML*, 1998.