

# AN IMPROVED INITIALIZATION FOR LOW-RANK MATRIX COMPLETION BASED ON RANK-1 UPDATES

Ahmed Douik and Babak Hassibi

Department of Electrical Engineering, California Institute of Technology, Pasadena, CA, USA

## ABSTRACT

Given a data matrix with partially observed entries, the low-rank matrix completion problem is one of finding a matrix with the lowest rank that perfectly fits the given observations. While there exist convex relaxations for the low-rank completion problem, the underlying problem is inherently non-convex, and most algorithms (alternating projection, Riemannian optimization, etc.) heavily depend on the initialization. This paper proposes an improved initialization that relies on successive rank-1 updates. Further, the paper proposes theoretical guarantees under which the proposed initialization is closer to the unknown optimal solution than the all zeros initialization in the Frobenius norm. To cope with the problem of local minima, the paper introduces and uses random norms to change the position of the local minima while preserving the global one. Using a Riemannian optimization routine, simulation results reveal that the proposed solution succeeds in completing Gaussian partially observed matrices with a random set of revealed entries close to the information-theoretical limits, thereby significantly improving on prior methods.

**Index Terms**— Matrix completion, rank minimization, Riemannian optimization, rank-1 update, random matrices.

## 1. INTRODUCTION

Real-world data sets come, often times, naturally in a matrix form with missing entries. From such partially revealed model, one wishes to infer the remaining entries. For example, in the *collaborative filtering* problem [1], companies are interested in predicting their consumers' preferences in order to provide them with recommendations.

Recovering the missing data is generally an ill-posed problem due to the existence of an infinite number of solutions. Indeed, filling the missing entries with any values completes the matrix. However, the data is highly redundant which, often, means that the matrix is low-rank. Indeed, due to the fact that few features contribute to a person's rating, the NETFLIX problem [2], a famous instance of *collaborative filtering* problems, has a low-rank solution.

Rank minimization of a partially observed matrix is NP-hard [3], even for a matrix of rank 4 [4]. The influential paper

[5] demonstrates that the problem can be well-approximated and efficiently solved by semi-definite programming (SDP) [6, 7] using the nuclear norm relaxation. Besides the nuclear norm relaxation, multiple non-convex methods have been proposed in the literature. For example, the authors in [8] suggest an alternating minimization approach and the authors in [9] propose an adaptive sampling approach to solving the more general problem of tensor completion. Similarly, taking advantage of the geometry of the low-rank constraint, the authors in [10] recast the problem as an unconstrained optimization on the Grassmann manifold. Reference [11] suggests a regularized version of the Riemannian optimization and shows its convergence.

As mentioned earlier, the low-rank matrix completion problem is intrinsically non-convex. Many of the local optimization algorithms start with an initialization, and therefore, the performance heavily depends on that initialization. The paper's main contribution is to propose a rank-1 update based approach to discover an improved initialization and gives theoretical guarantees under which the found initial point is closer to the unknown optimal solution than the all zeros initialization in the Frobenius norm sense.

The philosophy of the approach is a greedy determination of the best rank-1 update such that the distance to the partially observed data is minimized. Such rank-1 updates have been used in different contexts in the literature, e.g., [12, 13, 14, 15, 16]. However, the aforementioned works used the approach to directly solve the problem whereas this paper aims to discover an improved initialization to solve the original non-convex problem. The work is closely related to the orthogonal rank-1 pursuit in [17]. However, instead of considering the usual Frobenius norm for which the rank-1 update is no longer orthogonal, this paper builds an algebraic structure for which the update is indeed orthogonal in the appropriate weighted Frobenius norm. To the best of the authors' knowledge, the theoretical guarantees and the proofs technique and tools are novel. To mitigate the effect of the local minima, the paper introduces a random norm that randomizes the locations of the local minima while preserving the global one. Therefore, by starting near the global solution and shuffling the positions of the local minima, the proposed algorithm has a high probability of success which is confirmed by the simulation results.

## 2. DEFINITIONS AND PROBLEM FORMULATION

### 2.1. Notation and Definitions

Throughout this paper, vectors and matrices are denoted by bold lower and upper case characters, respectively. For a vector  $\mathbf{x}$ , the notation  $x_i$  refers to the  $i$ -th entry of the vector. Likewise, for a matrix  $\mathbf{X}$ , the notations  $\mathbf{x}_i$  and  $\mathbf{X}_{ij}$  represent the  $i$ -th columns and the entry at the  $i$ -th row and the  $j$ -th column, respectively. The  $\ell_2$  and the Frobenius norms of a vector  $\mathbf{x}$  and a matrix  $\mathbf{X}$  are denoted by  $\|\mathbf{x}\|_2$  and  $\|\mathbf{X}\|_F$ , respectively. The weighted Frobenius norm of a matrix  $\mathbf{X}$ , wherein the weighing matrix  $\Psi$  consists of positive entries, is defined by  $\|\mathbf{X}\|_\Psi = \|\sqrt{\Psi} \odot \mathbf{X}\|_F$ , where  $\odot$  represents the usual Hadamard product of two matrices  $\mathbf{X}$  and  $\mathbf{Y}$ , i.e.,  $(\mathbf{X} \odot \mathbf{Y})_{ij} = \mathbf{X}_{ij} \mathbf{Y}_{ij}$ , and  $\sqrt{\Psi}$  is the entry-wise square root of the matrix  $\Psi$ .

The set of strictly positive integers less than  $n$ , i.e.,  $\{1, \dots, n\}$ , is denoted by  $[n]$ . The set of strictly positive pair of integers  $(i, j)$  with  $1 \leq i \leq n_1$  and  $1 \leq j \leq n_2$  is denoted by  $[n_1] \times [n_2]$ .

Let  $\mathbf{A}$  be an  $n_1 \times n_2$  partially observed matrix and let  $\Omega \subseteq [n_1] \times [n_2]$  be the set of observed indices. The paper uses the standard assumption that entries are revealed identically and independently from a uniform distribution with probability  $p$ . For ease of notations, we use the symbol  $\Omega$  to denote the binary matrix such that  $\Omega_{ij} = 1$  if and only if the entry is revealed. The complement of  $\Omega$ , i.e., the set of unknown entries, is denoted by  $\Omega^c = [n_1] \times [n_2] \setminus \Omega$ .

**Definition 1.** A matrix  $\mathbf{B} \in \mathbb{R}^{n_1 \times n_2}$  is said to fit the partially observed matrix  $\mathbf{A}$  of size  $n_1 \times n_2$  and observed set  $\Omega$  if and only if  $\mathbf{A}_{ij} = \mathbf{B}_{ij}$ ,  $\forall (i, j) \in \Omega$ . For such matrices, we use the notation  $\mathbf{A} \simeq_\Omega \mathbf{B}$ . The index  $\Omega$  in  $\simeq_\Omega$  is omitted in the rest of the paper as it is clear from the context.

### 2.2. Problem Formulation

Given a partially observed data matrix  $\mathbf{A}$  of size  $n_1 \times n_2$  with observed set  $\Omega$ , the problem of low-rank matrix completion is the problem of finding a low-rank matrix that fits the data. In other words, the goal is to find matrix  $\mathbf{A}^*$  such that:

$$\mathbf{A}^* = \arg \min_{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}} \text{Rank}(\mathbf{X}) \quad (1a)$$

$$\text{s.t.} \quad \mathbf{X} \simeq \mathbf{A}. \quad (1b)$$

The paper uses the standard assumption that the matrix  $\mathbf{A}^*$  is unique for a sufficient number of revealed entries. It also uses the standard assumption that the entries of  $\mathbf{A}^*$  are drawn from the same distribution, e.g.,  $\mathbf{A}^* = \mathbf{U}\mathbf{V}^T$ , with  $\mathbf{U}$  and  $\mathbf{V}$  being two independent standard normal matrices. Define  $r^* = \text{Rank}(\mathbf{A}^*)$  as the minimum achievable rank. Assuming that the completion rank  $r^*$  is given a priori, a traditional method to efficiently solving the above problem with a moderate computation complexity is to reinterpret the problem as

---

### Algorithm 1 Rank-1 Update Matrix Completion Algorithm

---

**Require:**  $\mathbf{A}$ ,  $\Omega$ , and  $r^*$ .

- 1: Initialize  $\mathbf{X}_0^* = \mathbf{0}$ .
  - 2: **for**  $n = 1$  to  $r^*$  **do**
  - 3:   Solve  $(\mathbf{P}'_n)$  to obtain  $\mathbf{X}_n^*$ .
  - 4: **end for**
  - 5: Output  $\mathbf{X}_{r^*}^*$ .
- 

an optimization on the Grassmann manifold. While such optimization is efficient, it is intrinsically non-convex, and thus, its performance highly depends on the used initialization. The rest of this paper suggests an improved initialization based on a rank-1 update approach. Strong theoretical guarantees are given to attest the efficiency of the initialization.

## 3. PROPOSED IMPROVED INITIALIZATION

### 3.1. Problem Reformulation

As shown in the previous section, solving the matrix completion problem requires an exponential number of floating point operations. This section proposes an iterative approach to the problem by introducing the “distance” operator  $\|\mathbf{A} - \mathbf{X}\|_\Omega^2 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (\mathbf{A}_{ij} - \mathbf{X}_{ij})^2 \Omega_{ij}$ . This section suggests replacing the matrix completion problem (1) by the following set of fixed rank optimization problems indexed by  $n$ :

$$(\mathbf{P}_n) \quad \min_{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}} \|\mathbf{A} - \mathbf{X}\|_\Omega^2$$

$$\text{s.t.} \quad \text{Rank}(\mathbf{X}) = n. \quad (2)$$

Let  $r^*$  be the smallest index  $n$  for which the program  $(\mathbf{P}_n)$  in (2) produces a zero solution, then  $r^*$  is the rank of the optimal solution to the matrix completion optimization problem in (1). Furthermore, we have  $\mathbf{A}^* \simeq \mathbf{X}_{r^*}$  wherein  $\mathbf{A}^*$  and  $\mathbf{X}_{r^*}$  are the optimal arguments of (1) and (2).

The set of  $(\mathbf{P}_n)$  suffers from the rank constraint that makes it difficult to solve. Indeed, the search space, i.e., matrices of a given rank, is neither linear nor convex. Hence, unless a good initialization is provided, the problem suffers from local optima. Furthermore, solving  $(\mathbf{P}_n)$  iteratively until reaching the rank  $r^*$  is not only a computation bottleneck but also a waste of computational resources as each iteration does not use the result of the previous one. The rest of the paper circumvents these drawbacks by adding extra constraints to the successive problem to produce the improved initialization.

### 3.2. Proposed Initialization

As discussed above, in the initial step, the closest rank-1 matrix to the solution is discovered. In other words, the solution  $\mathbf{X}_1^*$  is initialized by solving  $(\mathbf{P}_1)$ . For  $n > 1$ , the problems  $(\mathbf{P}_n)$  are difficult to solve and do not use the previous results. This section suggests updating the existing solution

with a rank-1 matrix to find the solution to the next problem. In particular, it solves, at each iteration  $n$ , the distance minimization problem (2) by imposing a certain structure to the solution. In this context, the proposed modified fixed rank optimization problem is formulated as follows:

$$\begin{aligned}
(\mathbf{P}'_n) \quad & \mathbf{X}_n^* \in \arg \min_{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}} \|\mathbf{A} - \mathbf{X}\|_{\Omega}^2 \\
& \text{s.t.} \quad \mathbf{X} = \mathbf{X}_{n-1}^* + \mathbf{xy}^T, \\
& \quad \mathbf{x} \in \mathbb{R}^{n_1}, \mathbf{y} \in \mathbb{R}^{n_2} \quad (3)
\end{aligned}$$

with the initialization  $\mathbf{X}_0^* = \mathbf{0}$ . The steps of the algorithm are summarized in Algorithm 1. All the optimization algorithms are solved in this paper by using a Riemannian optimization routine on the manifold of low-rank matrices [18].

#### 4. PERFORMANCE ANALYSIS AND IMPLEMENTATION

This section derives conditions under which the found initialization is closer to the global optimal solution to the matrix completion problem than the all zeros initialization in the Frobenius norm sense. To cope with the presence of local minima, we further present an implementable version that randomizes the local optima while preserving the global one.

##### 4.1. Performance Analysis

In general, there is a discrepancy between the solutions produced by the optimization problems  $(\mathbf{P}_n)$  and  $(\mathbf{P}'_n)$ . Nevertheless, under some conditions, the solution reached by  $(\mathbf{P}'_n)$  presents an improved initialization to solving the fixed rank optimization. The section exhibits conditions under which the output of  $(\mathbf{P}'_{r^*})$  is a good initialization. Such *sufficient* conditions are obtained by introducing the extended singular value decomposition (E-SVD).

The E-SVD can be computed efficiently. However, due to space limitation, such construction is omitted herein and only the properties of the decomposition are given. The decomposition allows to write the partially observed matrix as the sum of rank one matrices that are orthogonal in the  $\Omega$ -norm sense. In other words, the extended singular value decomposition of a partially observed matrix is  $\mathbf{A} \simeq \sum_{i=1}^n \sigma_i \mathbf{u}_i \mathbf{v}_i^T$  with  $\langle \mathbf{u}_i \mathbf{v}_i^T | \mathbf{u}_j \mathbf{v}_j^T \rangle_{\Omega} = \delta_{ij}$ ,  $1 \leq i, j \leq n$ . Normalizing the right (only one side can be normalized) generalized eigenvectors, the decomposition can be written in the compact form  $\mathbf{A} \simeq \mathbf{U} \Sigma \mathbf{V}^T$  with  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n] \in \mathbb{R}^{n_1 \times n}$  and  $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n] \in \mathbb{R}^{n_2 \times n}$ ,  $n = \min(n_1, n_2)$  satisfying  $\mathbf{U}^T \mathbf{U} = \text{diag}(\gamma_1, \dots, \gamma_n)$  and  $\mathbf{V}^T \mathbf{V} = \mathbf{I}_n$ . Therefore, the  $\Omega$ -norm of  $\mathbf{A}$  is given by  $\|\mathbf{A}\|_{\Omega}^2 = \sum_{i=1}^n \sigma_i^2$ .

The solution  $\mathbf{X}_{r^*}$  reached by  $(\mathbf{P}'_{r^*})$  presents an improved initialization to solving the matrix completion problem if the criteria given in the following theorem is satisfied:

**Theorem 1.** *A sufficient condition for the output  $\mathbf{X}_{r^*}$  of Algorithm 1 after  $r^*$  iterations to serve as a good initialization*

---

#### Algorithm 2 Proposed Matrix Completion Algorithm

---

**Require:**  $\mathbf{A}$ ,  $\Omega$ ,  $r^*$  and  $T$ .

- 1: Estimate  $\hat{\mathbf{X}}$  using Algorithm 1.
  - 2: Initialize  $\mathbf{X}^* = \hat{\mathbf{X}}$ .
  - 3: **for**  $t = 1 : T$  **do**
  - 4:   Generate  $\Psi$  randomly from  $\Omega$ .
  - 5:   Initialize with  $\hat{\mathbf{X}}$  and solve  $(\mathbf{P}_{r^*})$  with the  $\Psi$  distance to obtain  $\mathbf{X}_t^*$ .
  - 6: **end for**
  - 7: Output  $\mathbf{X}^*$  the best  $\mathbf{X}_t^*$ ,  $1 \leq t \leq T$  in the  $\|\mathbf{A} - \mathbf{X}_t^*\|_{\Omega} / \|\mathbf{A}\|_{\Omega}$  sense.
- 

to solving the matrix completion problem  $(\mathbf{P}_{r^*})$  in (2), in the sense that it is closer in the Frobenius norm to the optimal solution than the all zeros matrix, is:

$$(1 - \alpha) \|\mathbf{A} - \mathbf{U}_{r^*} \Sigma_{r^*} \mathbf{V}_{r^*}^T\|_{\Omega} \leq \sqrt{1 + \alpha^2} \|\mathbf{A}\|_{\Omega}.$$

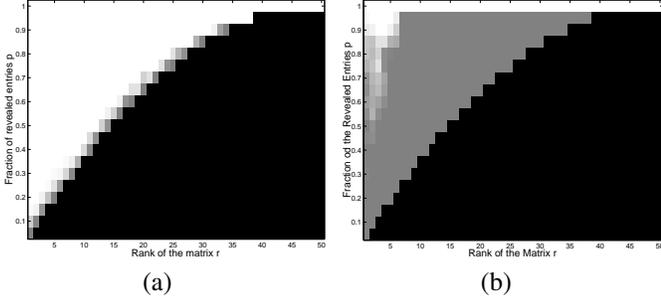
where  $\mathbf{U} \Sigma \mathbf{V}^T$  is the extended generalized singular value decomposition of  $\mathbf{A}$ ,  $\mathbf{U}_{r^*} \Sigma_{r^*} \mathbf{V}_{r^*}^T$  is the truncated E-SVD, and  $\alpha = \sqrt{\frac{1-p}{p}}$ , with  $p$  being the probability that an entry is revealed.

It is worth mentioning that the above theorem shows that the performance of the initialization can be assessed even before computing the actual point as it only requires computing the extended singular value which can be done efficiently in  $n^4$  operations. Furthermore, note that the sufficient condition given in Theorem 1 does not require knowledge of the completed matrix  $\mathbf{A}$ . Indeed, the knowledge of the partially observed matrix is sufficient to construct the extended singular value decomposition and to verify how good the initial guess is. However, since the condition is only sufficient, the actual initialization may be good even when the conditions of Theorem 1 are not met.

##### 4.2. Implementation

Due to the presence of local minima, initializing the algorithm close to the optimal does not guarantee convergence to the global minimum. In order to cope with the local minima, the paper introduces the following inner product  $\langle \mathbf{X} | \mathbf{Y} \rangle_{\Psi} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \mathbf{X}_{ij} \mathbf{Y}_{ij} \Psi_{ij}$ , wherein the matrix  $\Psi$  satisfies  $\Psi_{ij} > 0$  is  $(i, j) \in \Omega$  and 0 otherwise.

To mitigate the effects of the local minima, this section proposes using the previously defined  $\Psi$ -norms as an objective function when solving the optimization problem  $(\mathbf{P}_{r^*})$  in (2). Indeed, all  $\Psi$ -norms conserve the fitting property, i.e., if  $\mathbf{X} \simeq \mathbf{A}$ , then  $\|\mathbf{X} - \mathbf{A}\|_{\Psi}^2 = 0$  for all  $\Psi$ . However, the local minima of these norms do not necessary match since the result depends on the particular  $\Psi$ . Therefore, by exploiting the diversity of  $\Psi$ 's, a closer initialization point translates to a better convergence as shown in [19] in the context of a weighted low-rank approximation. The details of the algorithm are displayed in Algorithm 2.



**Fig. 1.** Perfect recovery of a randomly generated  $50 \times 50$  matrices of rank  $r$  and fraction of revealed entries  $p$  using the non-convex method in Algorithm 2 in (a) and the convex nuclear norm relaxation in (b).

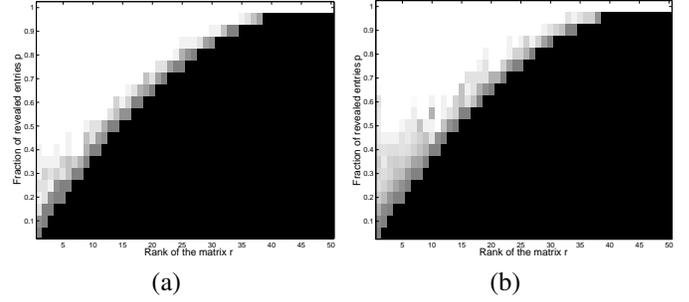
## 5. SIMULATION RESULTS

To attest the effectiveness of the proposed algorithm, this section shows simulation results of square  $n \times n$  matrices of rank  $r$ . The matrix  $\mathbf{A}$  is generated by multiplying two  $n \times r$  independent and identically distributed (i.i.d) Gaussian matrices  $\mathbf{U}$  and  $\mathbf{V}$  as  $\mathbf{A} = \mathbf{U}\mathbf{V}^T$ . The set of revealed entries is sampled independently from a Bernoulli random variable with probability  $p$ . The fixed rank optimization problem ( $\mathbf{P}_n$ ) is solved using the ManOpt toolbox for Riemannian optimization algorithms on manifolds provided in [18] and the trust-region algorithm [10]. These simulations use a maximum of  $T = 10$   $\Psi$ -norm. The recovery is declared perfect if  $\|\mathbf{X}^* - \mathbf{A}\|_F / \|\mathbf{A}\|_F < 10^{-3}$ . The color of each pixel reflects the average success with black being 100% failure and white being 100% success.

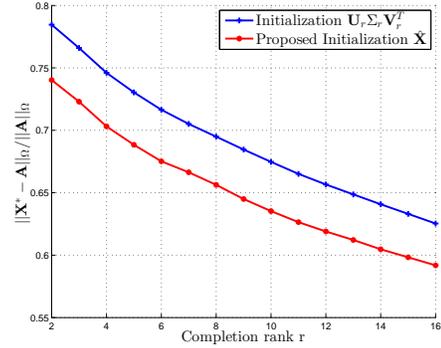
Figure 1 (a) plots the success rate of the proposed algorithm against both the rank of the partially observed matrix and the fraction of the revealed entries for a randomly generated matrix  $50 \times 50$  matrix for a maximum number of 500 iteration. Figure 1 (b) shows the performance of the nuclear norm approach. Unlike the proposed algorithm, the convex nuclear norm relaxation often times returns a matrix with a higher rank than wanted. For fair comparison, the solution  $\mathbf{X}^*$  obtained by the nuclear norm relaxation is projected by using the Eckart-Young-Mirsky low-rank approximation. By comparing both results from Figure 1 (a) and Figure 1 (b), it can be concluded that the proposed algorithm significantly outperforms the nuclear norm minimization in all configurations of rank and fraction of revealed entries.

The second part of the simulation compares the convergence speed for different initialization points using the same trust-region algorithm [10]. Since the initialization  $\mathbf{A} = \mathbf{U}_r \Sigma_r \mathbf{V}_r^T$ , i.e., the Eckart-Young-Mirsky low-rank approximation of  $\mathbf{A} \odot \Omega = \mathbf{U} \Sigma \mathbf{V}^T \odot \Omega$ , usually gives better results than the all zeros matrix, it is used in the simulations.

Figure 2 plots the success rate of correctly recovering a  $50 \times 50$  randomly generated matrix against the fraction of revealed entries and the rank of the partially observed matrix



**Fig. 2.** Perfect recovery of a randomly generated  $50 \times 50$  matrices of rank  $r$  and fraction of revealed entries  $p$  using Algorithm 2 for 50 iterations with the proposed initialization  $\hat{\mathbf{X}}$  in (a) and the initialization  $\mathbf{U}_r \Sigma_r \mathbf{V}_r^T$  in (b).



**Fig. 3.** Performance of the proposed initialization in finding the optimum of the fixed rank optimization ( $\mathbf{P}_r$ ) against the completion rank  $r$  for the Jester Datasets 1.

for a fixed number of 50 iteration of the trust-regions algorithm for the proposed initial guess  $\hat{\mathbf{X}}$  and the one found in literature  $\mathbf{U}_r \Sigma_r \mathbf{V}_r^T$ . For the same computational resources, the proposed algorithm is able to recover more matrices. An average comparison between the figures reveals that the proposed algorithm outperforms the initialization  $\mathbf{U}_r \Sigma_r \mathbf{V}_r^T$  in almost all configurations.

Finally, to attest the performance of the proposed initialization in real world data, the paper solves the non-convex optimization problems ( $\mathbf{P}_r$ ) for different completion rank  $r$  with the arbitrary initialization  $\mathbf{U}_r \Sigma_r \mathbf{V}_r^T$  and the proposed initialization  $\hat{\mathbf{X}}$ . The experiment is run on a data file containing anonymous ratings of 100 jokes from 24,983 users known as Jester Dataset [20]. The optimization problem ( $\mathbf{P}_r$ ) is solved for different  $r$  and the resulting objective, i.e.,  $\|\mathbf{X}^* - \mathbf{A}\|_\Omega / \|\mathbf{A}\|_\Omega$ , is plotted. Figure 3 shows the performance of the proposed initialization. One can note that the proposed scheme systematically outperforms the arbitrary initialization which confirms that it allows starting the optimization procedure near to the optimum.

## ACKNOWLEDGMENT

The authors would like to thank Dr. Philipp Walk for his helpful comments and suggestions.

## 6. REFERENCES

- [1] David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry, "Using collaborative filtering to weave an information tapestry," *Communications of the ACM*, vol. 35, no. 12, pp. 61–70, Dec. 1992.
- [2] James Bennett, Charles Elkan, Bing Liu, Padhraic Smyth, and Domonkos Tikk, "The netflix prize," *Proc. of KDD Cup and Workshop (KDD' 2007)*, San Jose, California, USA, vol. 9, no. 2, pp. 51–52, Dec. 2007.
- [3] Alexander L. Chistov and Dima Grigoriev, "Complexity of quantifier elimination in the theory of algebraically closed fields," in *Proc. of the Mathematical Foundations of Computer Science 1984*, 1984, pp. 17–31.
- [4] Moritz Hardt, Raghu Meka, Prasad Raghavendra, and Benjamin Weitz, "Computational limits for matrix completion," in *Conference on Learning Theory*, 2014, pp. 703–725.
- [5] Emmanuel J Candès and Benjamin Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717–772, 2009.
- [6] Jos F. Sturm, "Using SeDuMi 1.02: A MATLAB toolbox for optimization over symmetric cones," 1998.
- [7] K. C. Toh, M.J. Todd, and R. H. Ttnc, "SDPT3 : A MATLAB software package for semidefinite programming," *Optimization Methods And Software*, vol. 11, pp. 545–581, 1999.
- [8] Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi, "Low-rank matrix completion using alternating minimization," in *Proc. of the 45-th annual ACM symposium on Theory of computing (STOC' 2013)*, Palo Alto, CA, USA. ACM, 2013, pp. 665–674.
- [9] Akshay Krishnamurthy and Aarti Singh, "Low-rank matrix and tensor completion via adaptive sampling," in *Advances in Neural Information Processing Systems*, 2013, pp. 836–844.
- [10] Nicolas Boumal and Pierre-antoine Absil, "RTRMC: A Riemannian trust-region method for low-rank matrix completion," in *Advances in neural information processing systems*, 2011, pp. 406–414.
- [11] Bart Vandereycken, "Low-rank matrix completion by riemannian optimization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1214–1236, 2013.
- [12] Zebang Shen, Hui Qian, Tengfei Zhou, and Song Wang, "Simple atom selection strategy for greedy matrix completion," in *Proceedings of the 24th International Conference on Artificial Intelligence*. 2015, IJCAI'15, pp. 1799–1805, AAAI Press.
- [13] Miroslav Dudik, Zaid Harchaoui, and Jérôme Malick, "Lifted coordinate descent for learning with trace-norm regularization," in *Artificial Intelligence and Statistics*, 2012, pp. 327–336.
- [14] Martin Jaggi, Marek Sulovsk, et al., "A simple algorithm for nuclear norm regularized problems," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 471–478.
- [15] Shai Shalev-Shwartz, Alon Gonen, and Ohad Shamir, "Large-scale convex minimization with a low-rank constraint," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, Lise Getoor and Tobias Scheffer, Eds., New York, NY, USA, June 2011, ICML '11, pp. 329–336, ACM.
- [16] Xinhua Zhang, Dale Schuurmans, and Yao-liang Yu, "Accelerated training for matrix-norm regularization: A boosting approach," in *Advances in Neural Information Processing Systems*, 2012, pp. 2906–2914.
- [17] Zheng Wang, Ming-Jun Lai, Zhaosong Lu, Wei Fan, Hasan Davulcu, and Jieping Ye, "Orthogonal rank-one matrix pursuit for low rank matrix completion," *SIAM Journal on Scientific Computing*, vol. 37, no. 1, pp. A488–A514, 2015.
- [18] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a Matlab toolbox for optimization on manifolds," *Journal of Machine Learning Research*, vol. 15, pp. 1455–1459, 2014.
- [19] Nathan Srebro, Tommi Jaakkola, et al., "Weighted low-rank approximations," in *Icml*, 2003, vol. 3, pp. 720–727.
- [20] Ken Goldberg, Theresa Roeder, Dhruv Gupta, and Chris Perkins, "Eigentaste: A constant time collaborative filtering algorithm," *Inf. Retr.*, vol. 4, no. 2, pp. 133–151, July 2001.