ACCELERATING ITERATIVE HARD THRESHOLDING FOR LOW-RANK MATRIX COMPLETION VIA ADAPTIVE RESTART

Trung Vu and Raviv Raich

School of EECS, Oregon State University, Corvallis, OR 97331-5501, USA

{vutru, raich}@oregonstate.edu

ABSTRACT

This paper introduces the use of adaptive restart to accelerate iterative hard thresholding (IHT) for low-rank matrix completion. First, we analyze the local convergence of accelerated IHT in the non-convex setting of matrix completion problem (MCP). We prove the linear convergence rate of the accelerated algorithm inside the region near the solution. Our analysis poses a major challenge to parameter selection for accelerated IHT when no prior knowledge of the "local Hessian condition number" is given. To address this issue, we propose a simple adaptive restart algorithm for MCP to recover the optimal rate of convergence at the solution, as motivated in [1]. Our numerical result verifies the theoretical analysis as well as demonstrates the outstanding performance of the proposed algorithm.

Index Terms— Matrix completion, SVD, Iterative hard thresholding, Nesterov's Accelerated Gradient.

1. INTRODUCTION

Low-rank matrix completion is a fundamental problem that arises in many areas of signal processing and machine learning such as collaborative filtering [2], system identification [3] and dimension reduction [4]. The problem can be explained as follows. Let $M \in \mathbb{R}^{m\times n}$ be the underlying matrix with low rank r and a subset its entries $\mathcal{S} = \{(i,j) \mid M_{ij} \text{ is observed}\}$. We aim to recover the unknown entries of M, belonging to the complement set $\mathcal{S}^c$. Alternatively, one would solve the following optimization problem:

$$\min_{X\in\mathbb{R}^{m\times n}} \operatorname{rank}(X) \quad \text{s.t. } X_{ij} = M_{ij}, \quad \forall (i,j) \in \mathcal{S}. \tag{1}$$

In one of the pioneer works, Candès and Recht [5] introduced a convex relaxation to the original non-convex matrix completion problem and presented conditions under which the solutions of the two problems coincide. Moreover, they provided an expression for the number of known entries required to recover the original matrix. This breakthrough leads to the class of proximal-type algorithms for nuclear norm minimization [6–9] with rigorous mathematical guarantees and extensions of classic acceleration techniques. The disadvantage of convex-relaxed methods, nonetheless, is either high computational complexity (for interior-point methods) or slow convergence rate (often sublinear for proximal-type methods). To address those issues, iterative hard thresholding has been proposed to directly solve the non-convex rank minimization problem [10–12]. Each IHT iteration takes one step in the direction of the gradient and one step projecting onto the set of rank-r matrices. Since the process is akin to hard-thresholding singular values, we refer to the methods using it as iterative hard thresholding algorithms, as opposed to their aforementioned soft thresholding counterparts. When the solution is low-rank, IHT is extremely efficient in both computational complexity and empirical convergence (linear rate). Notwithstanding, mathematical guarantees of non-convex IHT algorithms for MCP are generally restricted to local convergence [13, 14].

Despite the similarity between IHT and projected gradient methods, there have been but a few efforts in accelerating IHT and characterizing the performance thereby. In a very recent work, Khanna and Kyrillidis [15] introduces the use of acceleration to plain IHT yet in the context of rank minimization with affine constraints (ARMP). The authors provided convergence guarantees based on restricted strong convexity and smoothness properties of the loss function. However, as pointed out in [5], the results and techniques for ARMP cannot apply to MCP for which the restricted properties does not hold. Additionally, they left an open question on the optimal momentum step sizes that guarantee better performance over plain IHT. While determining an optimal tuning is NP-hard [16], our experiment indeed shows that a careless choice of step sizes might worsen the performance of plain IHT in a matrix completion setting. Thus, we believe answering this question is the key to the practicality of accelerated IHT in both ARMP and MCP.

In this paper, we consider IHT for solving low-rank ma-

This work is partially supported by the National Science Foundation grants CCF-1254218 and DBI-1356792.

trix completion and connect the classic theory of accelerated gradient methods with recent analyses of the local convergence of plain IHT in [13]. The contribution of our work is threefold: (i) we propose a variant of Nesterov's Accelerated Gradient in a MCP-IHT setting and analyze the local convergence thereof, (ii) we identify the choice of momentum step sizes that guarantees the optimal acceleration, (iii) we propose an adaptive restart algorithm that can asymptotically recover the local rate in practice. The numerical experiment verifies our theoretical analysis and demonstrates the superior performance of the proposed algorithm compared to common existing methods for low-rank matrix completion.

2. PRELIMINARIES

We begin with a review of some preliminaries on iterative hard thresholding methods for low-rank matrix completion.

Definition 1. *Let $M \in \mathbb{R}^{m\times n}(m \geq n)$ be a rank-r matrix and $M = U\Sigma V^T$ be its singular value decomposition (SVD), where Σ is a diagonal $m \times n$ matrix with diagonal entries*

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_n = 0$$

and U, V are $m \times m$ and $n \times n$ unitary matrices, respectively. We partition U, Σ, V as follows:

$$U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}, V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}$$

where $\Sigma_1 = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$, $\Sigma_2 = \mathbf{0}$; U_1, V_1 and U_2, V_2 are semi-unitary matrices corresponding to the partition of Σ.

Definition 2. *A row selection matrix $S \in \mathbb{R}^{s\times m}(s \leq m)$ is a semi-unitary matrix obtained by a subset of s rows from the identity matrix I_m. Left-multiplying a matrix $X \in \mathbb{R}^{m\times n}$ by S returns an $s \times n$ matrix corresponding to set of rows in X.*

Definition 3. *Sampling operator $X_{\mathcal{S}}$ maps the matrix entries not in $\mathcal{S}$ to 0:*

$$[X_{\mathcal{S}}]_{ij} = \begin{cases} X_{ij} & \text{if } (i,j) \in \mathcal{S}, \\ 0 & \text{if } (i,j) \in \mathcal{S}^c. \end{cases}$$

Definition 4. *Let $X \in \mathbb{R}^{m\times n}$ be a matrix with arbitrary rank. Define the rank-r projection of X as:*

$$\mathcal{P}_r(X) \in \underset{Y\in\mathbb{R}^{m\times n}}{\mathrm{argmin}} \|Y - X\|_F \text{ s.t. } \mathrm{rank}(Y) \leq r.$$

The solution of this minimization is obtained by computing the top r singular values and vectors of X [17]. Moreover, this projection is unique if either $\sigma_r(X) > \sigma_{r+1}(X)$ or $\sigma_r(X) = 0$, where $\sigma_r(.)$ denotes the r-th largest singular value. In the rest of this paper, we implicitly refer the solution of problem (1) and its SVD to the notations in Definition 1. This also implies our assumption that $\mathrm{rank}(M) = r$. Furthermore, we denote the cardinality of $\mathcal{S}$ by s and the row selection matrix corresponding to $\mathcal{S}^c$ by $S_c \in \mathbb{R}^{(mn-s)\times mn}$.

Algorithm 1 Iterative Hard Thresholding

1: $X^{(0)} = M_{\mathcal{S}}$
2: **for** $k = 1, 2, \ldots$ **do**
3: $\quad X^{(k)} = \mathcal{P}_r\big(X^{(k-1)} - \alpha_k[X^{(k-1)} - M]_{\mathcal{S}}\big)$

3. BACKGROUND

3.1. ARMP-IHT versus MCP-IHT

Iterative hard thresholding for low-rank matrices was first proposed in the context of ARMP. In [10], Jain et. al. considered the following robust formulation of ARMP:

$$\min_{X\in\mathbb{R}^{m\times n}} \frac{1}{2}\|\mathcal{A}(X) - b\|_2^2 \text{ s.t. } \mathrm{rank}(X) \leq r \qquad (2)$$

where $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^s$ is an affine transformation and $b \in \mathbb{R}^s$ is the set of indirect observations. Adapting the idea of projected gradient descent, the authors proposed the Singular Value Projection (SVP) algorithm with the basic update

$$X^{(k)} = \mathcal{P}_r(X^{(k-1)} - \eta_k\mathcal{A}^T(\mathcal{A}(X^{(k-1)}) - b)).$$

Under assumptions on Restricted Isometry Property (RIP) of the affine operator $\mathcal{A}$, the authors showed that their algorithm converges to the solution at a linear rate. In an independent work, Goldfarb and Ma [11] proved the geometric convergence for a special case of unit step size. Later on, there have been efforts in improving the performance of ARMP-IHT, namely Normalized IHT [18] and accelerated IHT [15]. All of these works use the standard RIP assumptions in order to prove the global convergence.

The matrix completion problem is a special case of ARMP where the affine operator $\mathcal{A}$ is a sampling operator:

$$\min_{X\in\mathbb{R}^{m\times n}} \|X_{\mathcal{S}} - M_{\mathcal{S}}\|_F^2 \text{ s.t. } \mathrm{rank}(X) \leq r. \qquad (3)$$

Unfortunately, this operator does not satisfy RIP in general, shattering the global convergence guarantees established in ARMP. Still, Jain et. al. suggested to apply SVP for solving MCP (see Algorithm 1) and made a conjecture that SVP converges linearly to the solution matrix M with high probability, provided M is incohenrent [10]. It took some time before the first theoretical guarantee is obtained in [13], considering a special case of SVP, called IHTSVD algorithm. When the step size α_k equals 1, one can simplify the gradient update in Algorithm 1 as $X^{(k-1)} - \alpha_k[X^{(k-1)} - M]_{\mathcal{S}} = [X^{(k-1)}]_{\mathcal{S}^c} + M_{\mathcal{S}}$. For convenience, we call this operator the *observation projection*, denoted by $\mathcal{P}_{M,\mathcal{S}}$. It simply sets entries of $X^{(k)}$ that are in $\mathcal{S}$ to those corresponding values of M. The IHT iterates now serve as alternating projections between $\mathcal{P}_r$ and $\mathcal{P}_{M,\mathcal{S}}$. More importantly, the authors provided a quantitative analysis on the local convergence of IHTSVD, based on the approximation of rank-r projection operator near

the solution. Let us restate their results in Theorem 1 and Theorem 2. We use our own notations for the purpose of consistency.

Theorem 1. *(Rephrased from [13]) Given the matrix M in Definition 1. Denote $\epsilon = \min_{\sigma_i > \sigma_{i+1}} \{\sigma_i - \sigma_{i+1}\}$. Let $\Delta \in \mathbb{R}^{m\times n}$ be a perturbation matrix such that $\|\Delta\|_F < \frac{\epsilon}{2}$. Then the rank-$r$ projection of $M + \Delta$ is given by*

$$\mathcal{P}_r(M+\Delta) = M + \Delta - U_2 U_2^T \Delta V_2 V_2^T + Q(\Delta)$$

where $Q : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ satisfies $\|Q(\Delta)\|_F = O(\|\Delta\|_F^2)$.

Theorem 2. *(Rephrased from [13]) If the matrix $S_c(V_2 \otimes U_2)$ has full rank, then Algorithm 1 with a unit step size converges to M locally at a linear rate $1-\sigma^2$, where $\sigma = \sigma_{\min}\big(S_c(V_2 \otimes U_2)\big)$. In other words, there exists a neighborhood $\mathcal{E}(M)$ of M and a constant C such that if $X^{(0)} \in \mathcal{E}(M)$, then*

$$\left\|X^{(k)} - M\right\|_F \le C\big(1-\sigma^2\big)^k \left\|X^{(0)} - M\right\|_F.$$

3.2. Nesterov's Accelerated Gradient for ARMP-IHT

We consider the plain IHT as a first-order gradient method and apply momentum techniques to accelerate it. In [19], Nesterov demonstrated a simple modification to gradient descent that provably improves the convergence rate dramatically. The method, known as Nesterov's Accelerated Gradient (NAG), can be described as follows

$$x^{(k)} = y^{(k-1)} - \alpha_k \nabla f(y^{(k-1)})$$
$$y^{(k)} = x^{(k)} + \beta_k (x^{(k)} - x^{(k-1)})$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuous differentiable, smooth convex function to be minimized. For optimizing an μ-strongly convex, L-smooth function, it is well-known that NAG obtains a linear convergence rate at $1 - \sqrt{\mu/L}$ by setting [19]

$$\alpha_k = \frac{1}{L}, \qquad \beta_k = \frac{1-\sqrt{\mu/L}}{1+\sqrt{\mu/L}}. \tag{4}$$

This scheme is often called optimal since it achieves the lower complexity bound for first-order methods on minimizing a strongly convex, smooth function derived by Nemirovski and Yudin [20].

The idea of accelerating IHT has recently been studied in [15, 21] for ARMP. It is similar in spirit to the Accelerated Proximal Gradient algorithm for solving nuclear norm regularized linear least square problems [9]. While these algorithms enjoy the convexity of the norm operator and copious theoretical guarantees of proximal methods, the burden of non-convex projections over the rank constraint bears heavily on IHT methods. Moreover, as we mentioned, convergence guarantee for accelerated ARMP-IHT in [15] does not hold for MCP-IHT. To the best of our knowledge, there is no convergence analysis for accelerated IHT in a matrix completion setting to date.

Algorithm 2 NAG-IHT

1: $X^{(0)} = Y^{(0)} = M_{\mathcal{S}}$
2: **for** $k = 1, 2, \ldots$ **do**
3: $\quad X^{(k)} = \mathcal{P}_r\big(Y^{(k-1)}\big)$
4: $\quad Y^{(k)} = \mathcal{P}_{M,\mathcal{S}}\big(X^{(k)} + \beta_k(X^{(k)} - X^{(k-1)})\big)$

4. ACCELERATING MCP-IHT

In this section, we first describe an accelerated scheme for Algorithm 1 and provide some analysis of the local convergence of the algorithm. It remains a challenging problem on parameter selection that guarantees better performance of accelerated over plain IHT. To address this issue, we propose an adaptive restart technique that allows us to asymptotically recover the optimal rate of convergence in practice.

4.1. An NAG-variant of MCP-IHT

Motivated by the result in Theorem 2, we propose an NAG-variant of IHT in Algorithm 2. First, notice the specific choice of gradient step size ($\alpha_k = 1$) unveils the observation projection $\mathcal{P}_{M,\mathcal{S}}$. Interestingly, this choice of α_k matches the setting in (4), as the Lipschitz constant of the sampling operator is $L = 1$. Second, the order at each iteration guarantees the sequence $\{Y^{(k)}\}$ is consistent with the observation $\mathcal{S}$, i.e., $Y^{(k)}_{\mathcal{S}} = M_{\mathcal{S}}$. As a result, the error matrix depends only on the entries in $\mathcal{S}^c$, disentangling the subsequent analysis of convergence. Finally, the algorithm terminates when a stopping criteria is met, returning $Y^{(k)}$ as an estimate of the matrix. We state our main theoretical result for the convergence of NAG-IHT in Theorem 3.[1] Note that the convergence rate is described in a closed-form, which can be verified through experiments. By constrast, RIP constants in the standard analysis for ARMP are NP-Hard to compute [16]. Mainly, the optimal fixed step size for NAG-IHT is identified, guaranteeing the better performance of accelerated schemes over plain IHT in theory, i.e., $1 - \sigma$ versus $1 - \sigma^2$.

Theorem 3. *If the matrix $S_c(V_2 \otimes U_2)$ has full rank, then Algorithm 2 with momentum step size $\beta_k = \frac{1-\sigma}{1+\sigma}$ converges to M locally at a linear rate $1 - \sigma$, where $\sigma = \sigma_{\min}(S_c(V_2 \otimes U_2))$. In other words, there exists a neighborhood $\mathcal{E}(M)$ of M and a constant C such that if $Y^{(0)} \in \mathcal{E}(M)$, then*

$$\left\|Y^{(k)} - M\right\|_F \le C\big(1-\sigma\big)^k \left\|Y^{(0)} - M\right\|_F.$$

Further, this is the optimal rate for any fixed momentum step size in Algorithm 2.

[1] The proofs of Theorem 1 and Theorem 3 are given in at http://web.engr.oregonstate.edu/∼vutru/nag_appendix.pdf

Algorithm 3 ARNAG-IHT

$t = 1, X^{(0)} = Y^{(0)} = M_{\mathcal{S}}, f_0 = \left\| X_{\mathcal{S}}^{(0)} - M_{\mathcal{S}} \right\|_F^2$
for $k = 1, 2, \ldots$ **do**
  $X^{(k)} = \mathcal{P}_r\big(Y^{(k-1)}\big)$
  $Y^{(k)} = \mathcal{P}_{M,\mathcal{S}}\big(X^{(k)} + \frac{t-1}{t+2}(X^{(k)} - X^{(k-1)})\big)$
  $f_k = \left\| X_{\mathcal{S}}^{(k)} - M_{\mathcal{S}} \right\|_F^2$
  if $f_k > f_{k-1}$ **then** $t = 1$ **else** $t = t + 1$

4.2. An adaptive restart scheme for NAG-IHT

Theorem 3 provides a theoretical guarantee for NAG-IHT but it implies that fixed-step-size strategy is impracticable when the value of σ is unknown. In this section, we propose a simple way to work around this issue. The idea stems from adaptive restart techniques for accelerated gradient schemes [1]: reset the momentum back to zero whenever we observe an increase in the function value. This facile heuristic was shown to asymptotically recover the local rate of convergence of NAG on minimizing a strongly convex smooth function and is generally used in sparse signal recovery. To the best of our knowledge, this work is the first to adopt adaptive restart heuristics to accelerate IHT. We describe our approach, named ARNAG-IHT, in Algorithm 3. It is important to highlight that the momentum need to grows from one iteration to the next in order to apply restart techniques. As a result, we use the incremental step size $\beta_k = \frac{t-1}{t+2}$ recommended in optimizing smooth convex functions [19]. The difference comes with conditional restarts (setting $t = 1$) whenever the square loss increases. Clearly, all three aforementioned algorithms share the same computational complexity per iteration.

5. EMPIRICAL RESULT

This section presents a numerical example to demonstrate our analysis for low-rank matrix completion. First, we generate a solution matrix $M \in \mathbb{R}^{m\times n}$ of rank r by taking the product of an $m \times r$ matrix and an $r \times n$ matrix, each having i.i.d. normally distributed entries. Next, we sample the observation set $\mathcal{S}$ uniformly at random. We compare ARNAG-IHT with the following methods: SVT [6], SVP-NewtonD [10], NIHT [18] and IHTSVD [13]. Although the analyses of SVP-NewtonD and NIHT only apply for ARMP, it is worth examining their empirical performance on MCP. In our own implementation of these algorithms, we use the set of parameters as suggested by the authors. For SVT, we set the step size $\delta = 1.2\frac{mn}{s}$ and the threshold $\tau = 5\sqrt{mn}$. For SVP-NewtonD, we set the step size $\eta_t = \frac{mn}{1.2s}$. NIHT, IHTSVD and ARNAG-IHT are parameter-free. Finally, we add NAG-IHT with two different fixed step sizes $\beta_k = \frac{1-\sigma}{1+\sigma}$ and $\beta_k = \frac{k-1}{k+2}$ for comparison.

Figure 1 illustrates the Frobenius norm of the error matrix as a function of the number of iterations. The dashed lines correspond to the theoretical convergence of IHTSVD (purple) at rate $1 - \sigma^2$ and NAG-IHT with step size $\beta_k = \frac{1-\sigma}{1+\sigma}$ (green) at rate $1 - \sigma$. As can be seen from the figure, both of the algorithms match the performance predicted in theory. SVT exhibits the slowest convergence due to the conservative nature of proximal-type algorithms. By contrast, all IHT algorithms enjoy a fast convergence at linear rates. Without acceleration, SVP-NewtonD and NIHT are clearly faster than IHTSVD. This can be explained by the fact that IHTSVD is a special case of SVP when the gradient step size is 1, whereas SVP-NewtonD and NIHT are improved versions of SVP with adaptive step sizes. Notwithstanding, ARNAG-IHT outperforms all other algorithms, asymptotically recovering the convergence rate at $1 - \sigma$. It approaches the "ideal" NAG-IHT with optimal step size in this experiment. Finally, we can observe the periodic behavior of momentum by setting the step size $\beta_k = \frac{k-1}{k+2}$, as experienced in the original version of NAG. However, it can be seen in Fig. 1 that this setting does not generally help improve the convergence of plain IHT.

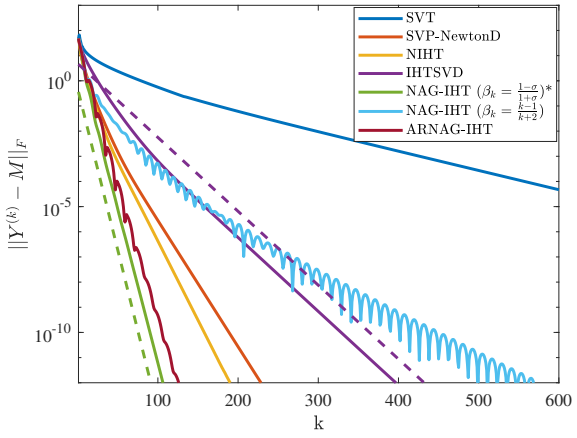

Fig. 1. The distance to the solution (in log-scale) as a function of the number of iterations for different algorithms (solid) and their corresponding theoretical bounds up to a constant (dashed). The parameters are set to $m = 50, n = 40, r = 3$, and $s = 1000$. Asterisks indicate algorithms using theoretical step sizes that are not available in practice.

6. CONCLUSION AND FUTURE WORK

We proposed the use of NAG to boost the performance of IHT for low-rank matrix completion. We analyzed the local convergence of NAG-IHT and established the optimal step size to guarantee faster convergence over plain IHT. We further introduced an adaptive restart algorithm that helps recover the optimal linear rate of convergence in practice. Our numerical evaluation showed evidence that the proposed scheme dramatically improves the performance of IHT for matrix completion problem. Still, understanding when and how our approach works in case the input matrix is noisy and not close to being low-rank is left as an open question for future work.

7. REFERENCES

[1] B. O'Donoghue and E. Candès, "Adaptive restart for accelerated gradient schemes," *Foundations of Computational Mathematics*, vol. 15, no. 3, pp. 715–732, 2015.

[2] J. Rennie and N. Srebro, "Fast maximum margin matrix factorization for collaborative prediction," in *Proceedings of the 22nd international conference on Machine learning*. ACM, 2005, pp. 713–719.

[3] Z. Liu and L. Vandenberghe, "Interior-point method for nuclear norm approximation with application to system identification," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 3, pp. 1235–1256, 2009.

[4] E. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 11, 2011.

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

[6] J.-F. Cai, E. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[7] D. Goldfarb, S. Ma, and L. Chen, "Fixed point and Bregman iterative methods for matrix rank minimization," *Mathematical Programming*, vol. 128, no. 1-2, pp. 321–353, 2011.

[8] S. Ji and J. Ye, "An accelerated gradient method for trace norm minimization," in *Proceedings of the 26th Annual International Conference on Machine Learning*, 2009, pp. 457–464.

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

[10] P. Jain, R. Meka, and I. Dhillon, "Guaranteed rank minimization via singular value projection," in *Advances in Neural Information Processing Systems (NIPS)*, 2010, pp. 937–945.

[11] D. Goldfarb and S. Ma, "Convergence of fixed-point continuation algorithms for matrix rank minimization," *Foundations of Computational Mathematics*, vol. 11, no. 2, pp. 183–210, 2011.

[12] K. Lee and Y. Bresler, "ADMiRA: Atomic decomposition for minimum rank approximation," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4402–4416, 2010.

[13] E. Chunikhina, R. Raich, and T. Nguyen, "Performance analysis for matrix completion via iterative hard-thresholded SVD," in *2014 IEEE Workshop on Statistical Signal Processing (SSP)*, 2014, pp. 392–395.

[14] M.-J. Lai and A. Varghese, "On convergence of the alternating projection method for matrix completion and sparse recovery problems," *eprint arXiv:1711.02151*, 2017.

[15] R. Khanna and A. Kyrillidis, "IHT dies hard: Provable accelerated iterative hard thresholding," in *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics*, 2018, vol. 84, pp. 188–198.

[16] A. Tillmann and M. Pfetsch, "The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 1248–1259, 2014.

[17] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.

[18] J. Tanner and K. Wei, "Normalized iterative hard thresholding for matrix completion," *SIAM Journal on Scientific Computing*, vol. 35, no. 5, 2013.

[19] Y. Nesterov, *Introductory lectures on convex optimization: a basic course*, Kluwer Academic Publishers, 2004.

[20] A. Nemirovski and D. Yudin, *Problem complexity and method efficiency in optimization*, John Wiley & Sons, 1983.

[21] A. Kyrillidis and V. Cevher, "Matrix recipes for hard thresholding methods," *Journal of mathematical imaging and vision*, vol. 48, no. 2, pp. 235–265, 2014.