Atom Decomposition with Adaptive Basis Selection Strategy for Matrix Completion

YAO HU*†, State Key Lab of CAD&CG, Zhejiang University
CHEN ZHAO*, State Key Lab of CAD&CG, Zhejiang University
DENG CAI, State Key Lab of CAD&CG, Zhejiang University
XIAOFEI HE, State Key Lab of CAD&CG, Zhejiang University
XUELONG LI, Center for OPTical IMagery Analysis and Learning (OPTIMAL), the School of Computer Science, Northwestern Polytechnical University

Estimating missing entries in matrices has attracted many attention due to its wide range of applications like image inpainting and video denoising, which are usually considered as low rank matrix completion problems theoretically. It is common to consider nuclear norm as a surrogate of the rank operator since it is the tightest convex lower bound of the rank operator under certain condition. However, most nuclear norm minimization based approaches involve numbers of singular value decomposition (SVD) operations. Given a matrix $X \in \mathbb{R}^{m \times n}$, the time complexity of SVD operation is $O(mn^2)$, which brings prohibitive computational burden on large scale matrices, limiting the further usage of these methods in real applications. Motivated by this observation, a series of atom decomposition based matrix completion methods have been studied. The key of these methods is to reconstruct the target matrix by pursuit methods in a greedy way, which only involves the computation of top SVD and has great advantages on efficiency compared with the SVD based matrix completion methods. However, due to gradually serious accumulation error, atom decomposition based methods usually result in unsatisfactory reconstruction accuracy. In this article, we propose a new efficient and scalable atom decomposition algorithm for matrix completion called *Adaptive Basis Selection Strategy* (**ABSS**). Different from traditional greedy atom decomposition methods, a two-phase strategy is conducted to generate the basis separately via different strategies according to their different nature. At first, we globally prune the basis space to eliminate the unimportant basis as much as possible and locate the probable subspace containing the most informative basis. Then another group of basis are learned to improve the recovery accuracy based on local information. In this way, our proposed algorithm breaks through the accuracy bottleneck of traditional atom decomposition based matrix completion methods, meanwhile it reserves the innate efficiency advantages over SVD based matrix completion methods. We empirically evaluate the proposed algorithm **ABSS** on real visual image data and large-scale recommendation datasets. Results have shown that **ABSS** has much better reconstruction accuracy with comparable cost to atom decomposition based methods. At the same time, it outperforms the state-of-art SVD based matrix completion algorithms by similar or better reconstruction accuracy with enormous advantages on efficiency.

Categories and Subject Descriptors: H.2.8 [**Database Management**]: Database Applications - Data Mining

General Terms: Algorithms, Experimentation

Additional Key Words and Phrases: Matrix Completion, Atom Decomposition, Basis Selection

1. INTRODUCTION

Matrix completion has arisen many applications in machine learning, computer vision and pattern recognition, such as recommender system [Koren et al. 2009; Chen et al. 2012], social network [Zhuang et al. 2015], video denoising [Ji et al. 2010], multi-task and multi-class learning [Argyriou et al. 2008; Negahban et al. 2011]. Without any prior knowledge, it is obvious that a perfect full

---

* Equal contribution
† Corresponding Author
Author's addresses: Y. Hu, C. Zhao, D. Cai, X. He, State Key Lab of CAD&CG, Zhejiang University, 388 Yu Hang Tang Road, Hangzhou, Zhejiang 310058, China; email:{huyao001, chenzhaozju, dengcai, xiaofeihe}@gmail.com. X. Li, Center for OPTical IMagery Analysis and Learning (OPTIMAL), the School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, Shaanxi, China; email: xuelong_li@nwpu.edu.cn.

DOI: 0000001.0000001

Table I. The Comparison of Running time between Full SVD and Top SVD

Size	100×100	1000×1000	5000×1000	5000×5000	5000×10000	10000×10000
Full SVD	1.70e-03(s)	0.14(s)	1.29(s)	22.40(s)	56.22(s)	265.03(s)
Top SVD	4.15e-04(s)	0.01(s)	0.09(s)	0.53(s)	1.12(s)	2.29(s)

recovery of matrix is virtually impossible, e.g., Fig. 1 (a). A natural assumption is that the underlying matrix comes from a restricted class, in which the matrix has low rank or approximately low rank structure. Mathematically, given an incomplete matrix Y$\in \mathbb{R}^{m\times n}$, we wish to recover the missing information by solving the rank minimization problem:

$$\begin{aligned} \min_{X\in\mathbb{R}^{m\times n}} \quad & rank(X) \\ s.t. \quad & X_{ij} = Y_{ij}, (i,j) \in \Omega, \end{aligned} \tag{1}$$

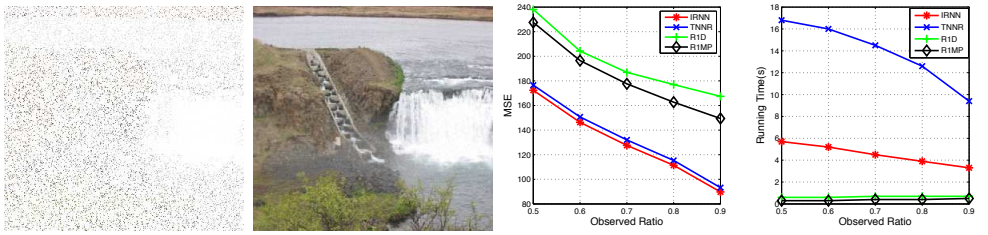
where $X \in \mathbb{R}^{m\times n}$, and Ω is the location of the observed entries.

However, the above rank minimization problem is NP-hard due to the non-convex and non-continuous nature of the rank operator. To effectively solve problem (1), variants of the surrogates for $rank(\cdot)$ have been explored [Srebro and Shraibman 2005; Wang et al. 2013]. Motivated by the progress of compressed sensing, nuclear norm is adopted as the most popular one, which can be defined as $\|X\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i(X)$ and $\sigma_1(X) \geq \sigma_2(X) \geq \cdots \geq \sigma_{\min(m,n)}(X) \geq 0$ are the singular values of X. This is because nuclear norm is the convex envelop of the rank operator over the domain $\|X\|_2 \leq 1$, where $\|\cdot\|_2$ stands for the spectral norm of the matrix. The relationship between $rank(\cdot)$ and nuclear norm is analogous to the relationship of ℓ_0 and ℓ_1 norm. Recent theoretical breakthrough has shown that solving nuclear norm minimization problem can perfectly recover the matrix from sufficient observed entries under some general assumptions.[Candès and Plan 2010; Candès and Tao 2010; Candès and Recht 2012; Recht 2011].

To solve nuclear norm minimization problem, existing approaches usually iteratively generate an approximate solution sequence $\{X_k\}$ that gradually converges to the optimal matrix X^*. According to the fundamental operations of these approaches, they can be categorized into two classes: singular value decomposition (SVD) based approaches and atom decomposition based methods.

A number of SVD based approaches have been proposed to solve nuclear norm minimization problem, including singular value thresholding (SVT) [Cai et al. 2010], Truncated Nuclear Norm Regularization (TNNR) [Hu et al. 2013] and Iteratively Reweighted Nuclear Norm (IRNN) [Lu et al. 2014]. Although these sophisticated algorithms can globally recover the missing information with strong theoretical guarantee, they all involve hundreds of expensive SVD operations, which brings prohibitive computation burden on running time and memory for large scale matrices.

To avoid numbers of SVD operations, researchers study to extend the successful atom decomposition techniques for vectors into matrix space, where the atom is substituted by a rank-one matrix as one basis, and one most suitable basis is greedily added in each iteration. Suppose in the k-th iteration, the matrix $X_k = X_{k-1} + \theta_k \mathbf{M}_k$, where $\mathbf{M}_k$ is the current most suitable rank-one matrix and θ_k is the related coefficient. Inspired by this motivation, a series of efficient algorithms have been proposed for large scale matrix completion problems, such as Rank-one Descent (R1D) [Dudik et al. 2012], Matrix Norm Boosting Approach [Zhang et al. 2012b], Rank-one Matrix Pursuit (R1MP) [Wang et al. 2014]. The above greedy atom decomposition based methods only depend on the computation of top singular vectors (top SVD) in each iteration. Table I shows the comparison between the efficiency of computing top SVD and full SVD. When the size of matrix increases, the computational cost for top SVD is far less than that of full SVD. Therefore, in spite of more iterations are needed for greedy atom decomposition methods to reach convergence, their total computational costs are much less compared with SVD based matrix completion methods. However, atom decomposition based greedy methods generally suffer from the unsatisfactory recovery performance. This is mainly because the inappropriate basis selection strategy in each iteration, which is the main bottleneck of these methods, leading to serious accumulation error with iterations and finally converging to the suboptimal solution.



(a) Damaged image (b) Recovered image (c) MSE vs observed ratio (d) Time vs observed ratio

Fig. 1. The comparison of different matrix completion methods on accuracy and efficiency in image recovery problem, where we can see that SVD based methods have better recovery performance but suffer from the expensive computational cost while atom decomposition methods are just the opposite.

Fig. 1 shows the mean square error (MSE) and running time for both SVD based methods and atom decomposition algorithms in image inpainting application. Given an incomplete image as in Fig. 1 (a), Fig. 1 (b) is its recovered image. By comparing the results of these methods, we can see that SVD based methods obtain much better recovery performance in Fig. 1 (c) while atom decomposition based algorithms have great advantages on efficiency in Fig. 1 (d). Therefore, it is interesting and necessary to explore whether we can combine their respective advantages together into a scalable matrix completion algorithm.

In this article, we propose a new scalable matrix completion algorithm based on atom decomposition called *Adaptive Basis Selection Strategy* (**ABSS**) for matrix completion. Considering the bottleneck of the basis selection strategy by traditional greedy atom decomposition methods, a two-phase method is proposed. In each phase, a different strategy is designed to select the basis separately according to their different nature. In the first phase, each basis $\mathbf{M}_k$ contains the most essential information and the corresponding approximate solution sequence X_k is relatively sparse. So we globally eliminate the unnecessary basis and prune the basis space into a smaller subspace to ensure the selected several basis as informative as possible. In the second phase, the basis needed to be learned contain the rest less but still essential information for better recovery. And here the approximate solution sequence becomes much denser around the optimal solution X^*. In this situation, we propose an *Adaptive Rank One Descent* algorithm to greedily generate the basis and the corresponding coefficients based on previously obtained local information. By cycling these two phases, our proposed method can obtain satisfactory recovery performance and is scalable to large scale matrices. Experimental results on both visual applications and recommender system have demonstrated that **ABSS** has similar or better recovery performance compared with SVD based matrix completion methods while retains the outperforming efficiency advantages of greedy atom decomposition methods.

The rest of the article is organized as follows: In the next section, we provide a brief description of the related work. In section 3, we present our motivation. And then in section 4 we introduce the proposed ABSS algorithm in detail for nuclear norm minimization problem and further extend it to truncated nuclear norm naturally in section 5. In section 6, evaluation results on both visual applications and recommender system are presented to demonstrate the effectiveness of our proposed method. Finally, we provide some concluding remarks in section 7.

2. RELATED WORK

Before introducing our approach, we firstly give a brief review of the related work in matrix completion, which can be roughly classified into two categories: SVD based methods and atom decomposition based methods.

2.1. SVD based Matrix Completion Methods

Considering the nonconvexity and discontinuous nature of the rank function in problem (1) for direct optimization, recent theoretical studies have proved that nuclear norm can be treated as a convex surrogate since it is the tightest convex lower bound of the rank operator for matrices in unit spectral norm ball. Furthermore, when the target matrix satisfies incoherence condition, we can exactly recover the missing information by solving the nuclear norm minimization problem:

$$\begin{aligned} &\min_{X\in\mathbb{R}^{m\times n}} \quad \|X\|_* \\ &s.t. \quad P_\Omega(X) = P_\Omega(Y), \end{aligned} \tag{2}$$

where $P_\Omega(\cdot)$ stands for the orthogonal projection operator onto the span of matrices vanishing outside, i.e., $(P_\Omega(Y))_{ij}$ equals to Y_{ij} if $(i,j) \in \Omega$ and 0 otherwise. Although the above optimization problem is convex, it is still difficult to solve it efficiently. [Fazel 2002] firstly proves that problem (2) is equivalent to a Semi-Definite programming (SDP) problem. However, existing solvers such as SDPT3 [Tütüncü et al. 2001] and SeDuMi [Sturm 1999] fail to solve large scale matrices efficiently. This limitation restricts its usage to real applications.

To better solve problem (2), a series of approaches are proposed to solve its relaxation problem and obtain approximate solution. By incorporating an extra Frobenius regularization to obtain a new objective function with strong convexity property, [Cai et al. 2010] propose a new method called *Sigular Value Threshoding* (SVT) to solve the relaxed optimization problem. However, the optimization of SVT is actually based on gradient descent method and has an unsatisfactory global convergence rate $O(\frac{1}{N})$. To accelerate SVT, [Hu et al. 2012] adopt the Nestrov's method with adaptive line search scheme to solve the convex and smooth dual problem of SVT and then directly obtain the primal solution based on the relationship of the primal-dual optimal solution. In this way, they can accelerate the convergence rate of SVT to be $O(\frac{1}{N^2})$.

On the other hand, different from solving the hard-constrained problem as in (2), [Toh and Yun 2010] apply the accelerated proximal gradient (APG) optimization technique to solve a soft-constrained problem instead as follows:

$$\min_{X\in\mathbb{R}^{m\times n}} \frac{1}{2}\|P_\Omega(X) - P_\Omega(Y)\|_F^2 + \mu\|X\|_*. \tag{3}$$

They also give a rigid theoretical analysis and show that the error between the recovered matrix and ground truth is smaller than ϵ after $O(\frac{1}{\sqrt{\epsilon}})$ iterations. Recently proposed method Active Subspace [Hsieh and Olsen 2014] deals with the soft-constrained nuclear norm regularization problem from a different view, which firstly finds the most possible subspace to produce the best approximation with the fixed rank constraint and gradually enlarges the subspace to make sure it includes the optimal solution.

Some researchers also point out solving matrix completion problem in a non-convex way. [Hu et al. 2013] propose to minimize the sum of the smallest $\min(m,n) - r$ singular values rather than all the singular values. By this way, they claim that they can exactly recover the low rank structure as long as it exists. Moreover, due to the rigorous uniform sampling condition of nuclear norm based methods to obtain a guaranteed recovery, researchers also consider using weighted nuclear norm to enhance the low rank structure extraction with better convergence rate under weaker assumption [Toh and Yun 2010; Mohan and Fazel 2010; Lu et al. 2014].

Overall, SVD based matrix completion methods generally have satisfactory recovery performance with hundreds of expensive SVD operations, which brings prohibitive computational burden for matrix completion on large scale matrices.

2.2. Atom Decomposition based Matrix Completion Methods

2.2.1. Atom Decomposition in Vector Space. In vector space, atom decomposition is a common technique widely used in signal processing [Wright et al. 2009; Tropp and Gilbert 2007]. Specifi-

cally, any signal can be decomposed by a collection of atoms as follows

$$\mathbf{s} = \sum_{\gamma \in \Gamma} \alpha_\gamma \phi_\gamma, \tag{4}$$

where Γ is an index set, ϕ_γ stands for a unit vector and α_γ is the related coefficient. Researchers wish to construct a proper atom set to adapt different purposes (i.e., smallest error) as much as possible.

To find the suitable representations, [Chen et al. 2001] propose the basis pursuit (BP) method to construct the atom set, where linear programming is used for the optimization of ℓ_1 regularization. [Mallat and Zhang 1993] adopt the matching pursuit (MP) method for building up the atom set by adding the most suitable atom to reduce the error between the current approximation and the signal in each iteration. To accelerate MP, [Pati et al. 1993] propose the orthogonal matching pursuit (OMP) as a refinement of all atoms to obtain faster convergency. In addition, some forward-backward methods are also proposed in vector space [Zhang 2009; Liu et al. 2014] to generate a better atom decomposition, where backward step can correct the error generated by forward algorithm.

2.2.2. Atom decomposition in matrix space. Motivated by the efficiency of the atom decomposition techniques in vector space, researchers try to generalize it to matrix space. Here each atom (basis) is a rank one matrix in the form of $\mathbf{u}\mathbf{v}^T$, where $\|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1$, $\mathbf{u} \in \mathbb{R}^m$ and $\mathbf{v} \in \mathbb{R}^n$. However, the set of such rank one matrix is overcomplete and uncountable, which makes the atom decomposition not unique. Therefore, the main challenge arising here is to select a proper finite basis set for the matrix space. Recently, [Dudik et al. 2012] prove the relationship between ℓ_1 norm in vector space and nuclear norm in matrix space. Then they reframe the challenging non-smooth nuclear norm minimization problem into a surrogate infinite-dimensional optimization problem with a regular ℓ_1 regularization penalty in vector space, which can be solved by coordinate descent method and only needs to compute the top singular vector pairs using lanczos iterations [Chen 2005] or other techniques. The greedy efficient component optimization (GECO) [Shalev-Shwartz and Tewari 2011] also extends the well-known orthogonal matching pursuit from the vector case to the matrix case. However, It empirically reduces the number of iterations without theoretical guarantees. [Zhang et al. 2012b] use the similar approach, and give rigorous theoretical analysis to guarantee that only $O(\frac{1}{\epsilon})$ steps is needed to obtain an ϵ-accuracy solution.

Although several methods have been proposed to greedily generate the basis by computing the top singular vector pair in each iteration, they determine the related weight in a different way. [Dudik et al. 2012] set the weight as constant, [Shalev-Shwartz and Tewari 2011] explore to optimize the weight via a forward greedy selection procedure, [Zhang et al. 2012b] apply the boosting method to learn the weight and [Wang et al. 2014] update the weight by solving a regression problem. Comparing with SVD based matrix completion methods, atom decomposition based methods usually fail to obtain comparable recovery performance even if they have great advantages on efficiency.

3. MOTIVATION

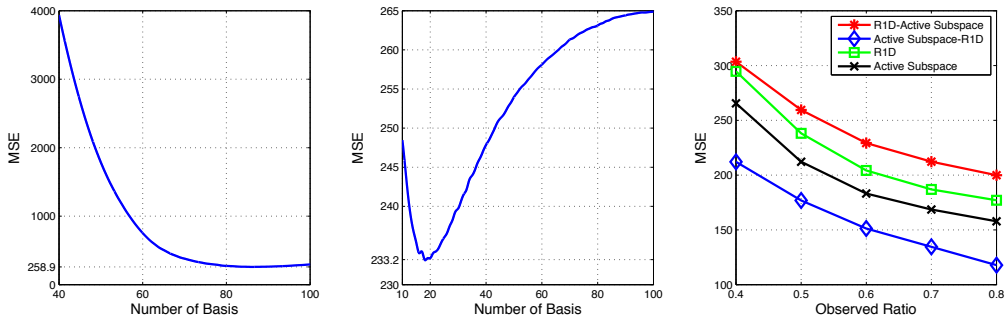
In this article, we focus on the general nuclear norm minimization problem:

$$\min_{X \in \mathbb{R}^{m \times n}} \quad F(X) = \mu \|X\|_* + L(X), \tag{5}$$

where $L(X)$ is a twice differential loss function and parameter μ is a trade-off parameter to control the degree of regularization. For simplicity of the following description, we firstly set $L(X) = ||P_\Omega(X) - P_\Omega(Y)||_F^2$, which enforces the consistency between the recovered matrix X and incomplete matrix Y on the observed entries. Inspired by the atom decomposition methods in vector space, we represent any matrix $X \in \mathbb{R}^{m \times n}$ as follows:

$$X = \sum_{i \in \mathcal{I}} \theta_i \mathbf{M}_i. \tag{6}$$

Note that $\mathcal{M} = \{\mathbf{M}_i \,|i \in \mathcal{I}\}$ is the set of all rank-one matrices with unit Frobenius norm and $\mathcal{I}$ is the index set. Similar with the case in vector space, each rank-one matrix $\mathbf{M}_i$ can be seen as one basis in matrix space. Mathematically, each rank-one matrix with unit Frobenius norm can be written as the product of two unit vectors. Namely, we can represent each basis as $\mathbf{M}_i = \mathbf{u}_i\mathbf{v}_i^T$, where $\mathbf{u}_i \in \mathbb{R}^m$ and $\mathbf{v}_i \in \mathbb{R}^n$ with $\|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1$. However, the basis set $\mathcal{M}$ is overcomplete and uncountable. Obviously, it is impossible for us to compute all basis. Therefore, the existing atom decomposition based matrix completion methods consider choosing finite basis for better approximation as soon as possible from different perspectives such as [Dudik et al. 2012; Zhang et al. 2012b; Wang et al. 2014]. These algorithms iteratively add a new basis and the related weight to the learned basis and obtain the approximate solution sequence $\{X_k \,|X_k = \sum_{i=1}^{k} \theta_i \mathbf{M}_i = \sum_{i=1}^{k} \theta_i \mathbf{u}_i\mathbf{v}_i^T\}$ in a greedy manner. With the increase of the number of learned basis, the approximate solution sequence converges to a local minimal X^*. Evaluation results have shown that the performance of these greedy atom decomposition methods varies and none of them have better or comparable recovery accuracy compared with the SVD based matrix completion methods such as TNNR [Hu et al. 2013] and IRNN [Lu et al. 2014]. This frustrating circumstance is mainly caused by the improper basis selection strategies, which is also the critical problem we need to study in the following section.



(a) R1D (b) Active Subspace (c) Combination strategies

Fig. 2. The analysis of selected basis by different methods for image recovery problem in Fig. 1 (a). In (a)(b), we show the MSE vs. number of basis by R1D and active subspace separately, where the observed ratio is set to be 0.4; In (c), we evaluate the recovery performance of different combination strategies under different observed ratio. In each combination strategy, we choose one method to select the first 40 basis and compute the rest 60 basis based on the residual matrix by another method, from which we can see that different combination strategies vary very much.

Before introducing our motivations, we firstly analyze the influence of the learned basis by different atom decomposition methods for image recovery problem in Fig. 1 (a). Fig. 2(a) and 2(b) show the variations of the *Mean Square Error* (MSE) with number of learned basis of R1D [Dudik et al. 2012] and Active Subspace [Hsieh and Olsen 2014] respectively. Although the MSE of the recovered matrix by both R1D and Active Subspace methods decreases sharply with learning of the first several basis, R1D has much larger MSE value. For example, the lowest MSE of the recovered matrix by R1D is 258.9 while that of the recovered matrix by active subspace selection with just about 20 basis is 233.2. This comparison shows that the top basis generated by Active Subspace lead to a more accurate approximation for the original matrix. However, with the number of learned basis increasing, the performance of Active Subspace turns to be worse, which demonstrates that the selected tail basis have little or negative contribution for a better approximation. Geometrically speaking, the convergence procedure of approximate sequence $\{X_k\}$ can be divided into two phase. When k is small (i.e., few basis are learned), we wish to enforce X_k accurately to move towards the optimal solution X^*. In this phase, the key point is to eliminate the unimportant basis as much as possible and local information has little affects, which is also equivalent to prune the basis space.

While k is large (i.e., several basis are learned), $\{X_k\}$ becomes much denser around X^*. In this situation, local geometrical information becomes more critical. Therefore, Active Subspace is successful in the first phase while fails in the second phase. And R1D gradually deviates from the correct convergence routine due to the cumulative error caused by local based greedy basis learning at the beginning of optimization.

Motivated by the above observation, it is natural to design a combinational strategy to select the basis in different phases separately. In Fig. 2 (c), we conduct an evaluation of two simple combination strategies for the image recovery problem under different observed ratio. All the compared algorithms are set to select 100 basis. For each strategy, we select the first 40 basis by one method and generate the rest 60 basis with another method. From the comparison results, we can see that one strategy has excellent performance while others perform badly, which means that arbitrary combinational strategies cannot succeed. Therefore, how to choose a proper way to select the basis in each phase is very critical for an optimal strategy and we will study it in the following section.

4. ADAPTIVE BASIS SELECTION STRATEGY

As mentioned in the previous section, we propose to separately generate the basis in different phases. For the simplicity of description, we refer the learned basis from the first phase as main basis and the corresponding basis in the second phase as tail basis. Notice the facts that main basis contain most information of matrix reconstruction and have strong discriminative characteristics, which guarantee the correct solution path along the convergence routine. Therefore, the global geometrical issues are more preferred in this phase. Meanwhile, tail basis have relatively less but also essential clues, which mainly depend on local geometrical information. Based on different characteristics of main basis and tail basis, we adopt different strategies to generate them as accurate as possible.

4.1. Global Pruning Phase: Select the Main Basis

In this subsection, we globally prune the infinite basis space to seek the subspace containing the most information. And then the related coefficients for each basis is computed for better reconstruction in the chosen subspace. In this way, the learned basis effectively ensure the approximate solution sequence $\{X_k\}$ consistently step towards the optimal recovery X^* .

4.1.1. Generate the Most Informative Subspace. Recall the fact that the set of all rank-one matrices with unit Frobenius norm $\mathcal{M}$ is infinite. To choose the finite most informative basis set, we should firstly measure the amount of information for each basis. Considering the sub-differential of the objective function $F(X)$ in (5) with respect to each coefficient θ_i, it falls into the interval $[\langle \mathbf{M}_i, \nabla L(X)\rangle - \mu, \langle \mathbf{M}_i, \nabla L(X)\rangle + \mu]$. Mathematically, basis $\mathbf{M}_i$ is useless for the minimization of $F(X)$ if and only if $\langle \mathbf{M}_i, \nabla L(X)\rangle = 0$ and $|\langle \mathbf{M}_i, \nabla L(X)\rangle| \leq \mu$. This is because $\theta_i = 0$ is one possible optimal solution in this situation and the corresponding basis actually makes no contribution for reconstruction. Following the similar idea in [Hsieh and Olsen 2014], the most informative basis can be found in the complementary subspace $\mathcal{A}$ defined as follows:

$$\mathcal{A} = \{\mathbf{M}_i \,|\, \langle \mathbf{M}_i, \nabla L(X)\rangle \neq 0 \text{ or } |\langle \mathbf{M}_i, \nabla L(X)\rangle| > \mu, \forall \mathbf{M}_i \in \mathcal{M}\}. \tag{7}$$

Under this definition, all basis in the complementary subspace $\mathcal{A}$ contribute to the update for current solution X_k, leading the correct direction to the optimal recovery X^*. The following theorem provides a simple way to accurately generate $\mathcal{A}$.

THEOREM 4.1. *[Hsieh and Olsen 2014] Assume $U\Sigma V^T$ is the singular value decomposition of $X \in \mathbb{R}^{m\times n}$, where $\Sigma = diag(\{\sigma_i\}_{1\leq i\leq \min(m,n)})$. Then the shrinkage operator $\mathcal{D}_\mu(\cdot)$ is defined as $\mathcal{D}_\mu(X) = U\mathcal{D}_\mu(\Sigma)V^T, \mathcal{D}_\mu(\Sigma) = diag(\max\{\sigma_i - \mu, 0\})$. Suppose U_G and V_G are the singular vectors obtained by $\mathcal{D}_\mu(X - \nabla L(X))$. Let $\mathbf{u}_A$ be an orthogonal basis of span(U, U_G), $\mathbf{v}_A$ be an orthogonal basis of span(V, V_G), then we can find that*

$$\mathbf{M}_A = \mathbf{u}_A\mathbf{v}_A^T \in \mathcal{A}. \tag{8}$$

Based on the above theorem, we can firstly compute the top singular value vectors $\{(\mathbf{u}_1, \mathbf{v}_1), (\mathbf{u}_2, \mathbf{v}_2), \cdots, (\mathbf{u}_s, \mathbf{v}_s)\}$ through a QR decomposition. And then we can generate the main basis as $\{\mathbf{M}_1 = \mathbf{u}_1\mathbf{v}_1^T, \mathbf{M}_2 = \mathbf{u}_2\mathbf{v}_2^T, \cdots, \mathbf{M}_s = \mathbf{u}_s\mathbf{v}_s^T\}$.

4.1.2. Compute the coefficients for Better Reconstruction. After selecting a collection of basis, we need to compute the corresponding coefficients θ_i for each basis to obtain the best recovery under the learned basis. With the learned basis, the original problem (5) degenerates the optimization for the coefficients. For the ease of description, we denote U_A and U_A to be the matrices where the i-th column is previously obtained $\mathbf{u}_i$ and $\mathbf{u}_i$ respectively. Based on these symbols, we can reformulate the subproblem in matrix form as follows:

$$\min_{\Sigma_A \in \mathbb{R}^{k\times k}} \quad \mu \left\|U_A \Sigma_A V_A^T\right\|_* + L(U_A \Sigma_A V_A^T). \tag{9}$$

To solve the above subproblem, we follow the same idea from [Zaid et al. 2012] to reframe the challenging non-smooth optimization problem with nuclear norm penalty (9) into a surrogate infinite-dimensional optimization problem with a regular ℓ_1-regularization penalty based on the following theorem.

THEOREM 4.2. *[Zaid et al. 2012] Given the basis set $\mathcal{M}$, the nuclear norm of X can be expressed as*

$$\|X\|_* = \min\{\|\theta\|_1 \,|\, \exists\, \mathbf{M}_i \in \mathcal{M}, s.t., X = \sum_{i=1}^{s} \theta_i \mathbf{M}_i\}. \tag{10}$$

Therefore, we can perform a second-order subspace minimization on the weights obtained so far. Furthermore, theoretical analysis has demonstrated that the optimization problem (9) in matrix space is equivalent to the following problem in vector space:

$$\begin{array}{ll} \min\limits_{\theta_1,\ldots,\theta_k} & \mu \sum\limits_{j=1}^{s} \theta_j + L(\sum\limits_{j=1}^{s} \theta_j \mathbf{u}_j \mathbf{v}_j^T) \\ s.t. & \theta_j \geq 0, j = 1, 2, \ldots, s. \end{array} \tag{11}$$

This is a convex and smooth problem with simple box constraints, which can be solved by Quasi-Newton algorithm with box constraints such as L-BFGS-B [Zhu et al. 1997]. This algorithm only requires objective and gradient evaluations with respect to the finite-dimensional $\theta = [\theta_1, \theta_2, \cdots, \theta_s]^T \in \mathbb{R}^s$.

4.2. Local Pursuit Phase: Adaptive Tail Basis Generation

When several basis have been learned in the previous phase, approximate solution sequence $\{X_k\}$ turns to be dense around the optimal solution X^*. Therefore, for better reconstruction, in this phase we need to select t new basis $\mathbf{M}_{s+j}, j = 1, \ldots, t$ to fit the residual matrix $R_0 = P_\Omega(Y) - P_\Omega(\sum_{i=1}^{s} \theta_i \mathbf{M}_i)$. And a fine-tuning is more preferred under this circumstance, which means that the key of tail basis selection is to effectively utilize the local information of obtained approximate solution so far. To do this, we propose a new greedy method called *Adaptive Rank-one Descent* (AR1D) to firstly select the basis and update the coefficients later.

Rank-one descent method [Dudik et al. 2012] is a greedy method to deal with the nuclear norm minimization problem (5) by reformulating the problem in the framework of ℓ_1 penalized boosting as

$$\min_{\tilde{\theta}} \; F(\tilde{\theta}) = \mu\|\tilde{\theta}\|_1 + L(\tilde{\theta}), \tag{12}$$

where $L(\tilde{\theta}) = \|(P_\Omega(\sum_{j=1}^{t} \tilde{\theta}_{s+j} \mathbf{M}_{s+j}) - P_\Omega(R_0)\|_F^2$, $\tilde{\theta} = [\tilde{\theta}_{s+1}, \tilde{\theta}_{s+2}, \cdots, \tilde{\theta}_{s+t}]^T \in \mathbb{R}^t$ and $\mathbf{M}_{s+j}, j = 1, 2, \ldots, t$, are the tail basis needed to learn. Coordinate descent method is used to solve (12), which picks the coordinate with steepest descent direction, i.e., smallest $\frac{\partial F(\tilde{\theta})}{\partial \tilde{\theta}_{s+j}}$. Therefore, we

ALGORITHM 1: Adaptive Rank-one Descent: AR1D(N,R_0,X_0,s)

Input: Number of total basis N, residual matrix R_0, initial recovery matrix X_0 and initial number of basis s.
Output: X_N.
for $k = 1, \ldots, N$ **do**
Compute the top singular vector pair $\{\mathbf{u}_{s+k}, \mathbf{v}_{s+k}\}$ of $-\nabla L(\tilde{\theta})$.
Update the coefficient $\tilde{\theta}_{s+k}$ by solving the following regression problem

$$\min_{\tilde{\theta}_{s+k}} \|R_k - \tilde{\theta}_{s+k}\mathbf{u}_{s+k}\mathbf{v}_{s+k}^T\|_F^2.$$

Update the recovery matrix: $X_k = X_{k-1} + \tilde{\theta}_{s+k}\mathbf{u}_{s+k}\mathbf{v}_{s+k}^T$.
Update the residual matrix: $R_{k+1} = R_k - \tilde{\theta}_{s+k}\mathbf{u}_{s+k}\mathbf{v}_{s+k}^T$.
end

can compute the corresponding tail basis as follows:

$$\arg\min_j \frac{\partial F(\tilde{\theta})}{\partial \tilde{\theta}_j} = \arg\min_j(\langle \mathbf{M}_{s+j}, \nabla\psi(\tilde{\theta})\rangle) = \arg\max_j(\mathbf{u}_{s+j}^T(-\nabla\psi(\tilde{\theta}))\mathbf{v}_{s+j}). \quad (13)$$

So we only need to compute the top singular-vector pair of $-\nabla\psi(\tilde{\theta})$ to generate the new basis $\mathbf{M}_{s+j} = \mathbf{u}_{s+j}\mathbf{v}_{s+j}^T$, which can be efficiently got by Lanczos iterations[Chen 2005]. Furthermore, empirical evaluations show that it is not necessary to compute the strict steepest direction in each iteration. A nearly steep (steepest up to $\frac{\epsilon}{2}, \epsilon > 0$) direction is good enough. Then, we need to decide the coefficients for the learned rank-one matrices, which is the step size for coordinate descent. Traditional rank-one descent adopts the line search strategy to choose proper step size. However, experimental results show that the step size is usually fixed, which means each rank-one matrix is treated equally. But there are two main shortcomings. Firstly, it is very difficult to tune the proper step size in advance. An overlarge step size will makes the algorithm diverge while a small step size leads to a slow convergence. Secondly, also the most important point, is that the approximation performance by this scheme is not satisfactory. Some evaluation results can be seen in the experiment section.

To overcome this bottleneck of rank one descent, after the generation of basis $\mathbf{M}_{s+j} = \mathbf{u}_{s+j}\mathbf{v}_{s+j}^T$, we follow the idea [Wang et al. 2014] to solve least square regression problem for greedily updating the coefficients. In the k-th iteration, the current residual $R_k = P_\Omega(Y) - P_\Omega(\sum_{i=1}^{s}\theta_i\mathbf{M}_i) - P_\Omega(\sum_{j=1}^{k-1}\tilde{\theta}_{s+j}\mathbf{M}_{s+j})$, and then we obtain $\tilde{\theta}_{s+k}$ by solving the following problem:

$$\min_{\tilde{\theta}_{s+k}} \|R_k - \tilde{\theta}_{s+k}\mathbf{M}_{s+k}\|_F^2. \quad (14)$$

By forcing the gradient of the above objective function to be zero, we can obtain the solution $\tilde{\theta}_{s+k}$ directly as

$$\tilde{\theta}_{s+k} = (\mathbf{M}_{s+k}^T\mathbf{M}_{s+k})^{-1}\mathbf{M}_{s+k}^T R_k. \quad (15)$$

To sum up, our proposed AR1D method has two steps: At first, based on previous learned basis, we adopt coordinate descent method to seek the most suitable basis; then we update the coefficients by solving a regression problem. The whole procedure of AR1D is summarized in Algorithm 1. By alternating these two steps, we can greedily construct much more accurate approximate solution. Empirical evaluation results have shown that our method can achieve better performance than traditional R1D and other atom decomposition based matrix completion methods.

ALGORITHM 2: The scheme of *Adaptive Basis Selection Strategy*

Input: Observed incomplete data matrix $P_\Omega(Y)$, regularization parameter μ, number of total basis N, which consists of s main basis and $t = N - s$ tail basis, number of generation schemes n.

Output: $\tilde{X}_T$.

Initialization: $X_0 = \text{AR1D}(N, P_\Omega(Y), 0, 0)$, $\tilde{X}_0 = X_0$.

for $T = 1, \ldots, n$ **do**

Phase1: Select the main basis in a global way

Select the most informative basis

$[U, \Sigma, V] = SVD(\tilde{X}_{T-1})$.

$[U_G, \Sigma_G, V_G] = S_\mu(\tilde{X}_{T-1} - \nabla L(\tilde{X}))$.

$[\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_s] = \mathsf{QR}([U_G(1:s), U])$, $[\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_s] = \mathsf{QR}([V_G(1:s), V])$.

Compute the related weights by solving

$$\begin{array}{cc} \min\limits_{\theta_1,\ldots,\theta_s} & \sum\limits_{j=1}^{s} \theta_j + L(\sum\limits_{j=1}^{s} \theta_j \mathbf{u}_j \mathbf{v}_j^T) \\ s.t. & \theta_j \geq 0, j = 1, 2, \ldots, s. \end{array},$$

Generate the approximate solution after Phase 1: $X_s = \sum\limits_{j=1}^{s} \theta_j \mathbf{u}_j \mathbf{v}_j^T$.

Phase2: Adaptive tail basis generation in a greedy way

Generate the matrix by tail basis: $X_{N-s} = \text{AR1D}(N - s, P_\Omega(Y) - P_\Omega(X_s), 0, s)$.

Update the recovery matrix: $\tilde{X}_T = X_s + X_{N-s}$.

end

4.3. Our Algorithm

As a summary, we propose a new combinatorial atom decomposition strategy called *Adaptive Basis Selection Strategy* (**ABSS**) for matrix completion, which consists of global pruning phase and local pursuit phase. In the global pruning phase, we use SVD based approach to learn basis ensuring the correct solution path towards the optimal solution; while in the local pursuit phase, an adaptive rank one descent method is designed to assist the better recovery. We define one selection scheme to be conducting global pruning phase and local pursuit phase once. By cycling several selection schemes, the proposed **ABSS** can obtain satisfactory recovery performance efficiently. The whole procedure has been presented in Algorithm 2. Note that we use adaptive rank one descent to generate an approximate solution as an initialization. Since the fast convergence rate of adaptive rank one descent, this initialization step only needs negligible computational cost while brings much better recovery performance. Furthermore, evaluation results in the experimental section show that usually only 2 selection schemes are enough for our proposed method to reach convergency.

5. EXTENSION TO TRUNCATED NUCLEAR NORM

5.1. Truncated Nuclear Norm

Although nuclear norm is the tightest convex surrogate of rank operation in the unit ball of spectral norm, its ability for capturing the low rank structure is limited in the real applications. A series of nonconvex variants have been studied from different perspectives, among which truncated nuclear norm [Hu et al. 2013] is the representative one. Different from traditional nuclear norm which treats each singular value by adding them together, truncated nuclear norm only minimizes the smallest $\min(m, n) - r$ singular values, since the rank of a matrix only corresponds to the first r nonzero singular values if the target matrix X indeed has the low rank structure with $rank(X) = r$. Mathematically, truncated nuclear norm can be defined as $\|X\|_r = \sum_{i=r+1}^{m} \sigma_i(X)$. Furthermore, the following theorem reveals the natural relationship between truncated nuclear norm and nuclear norm.

THEOREM 5.1. *[Hu et al. 2013] Given $X \in \mathbb{R}^{m\times n}$ with rank $r < min(m,n)$, for any matrices $A \in \mathbb{R}^{r\times m}$, $B \in \mathbb{R}^{r\times n}$ such that $AA^T = I_r$, $BB^T = I_r$, we have*

$$\|X\|_r = \|X\|_* - \max_{AA^T=I_r, BB^T=I_r} Tr(AXB^T). \tag{16}$$

and the optimal solution for the second term in equation (16) *on the right hand side of the equality is $A = (\mathbf{u}_1, ..., \mathbf{u}_r)^T$, $B = (\mathbf{v}_1, ..., \mathbf{v}_r)^T$. Where $(\mathbf{u}_i, \mathbf{v}_i)$ is the i-th singular vector pair of X.*

5.2. Apply ABSS on Truncated Nuclear Norm Regularization based Matrix Completion

From the definition of truncated nuclear norm, we can see that by minimizing $\|X\|_r$, we always obtain the low rank structure as long as it exists. This motivates us to substitute the nuclear norm regularization in the objective function $F(X)$ in problem (5) to be truncated nuclear norm. In this way, we can better recover the missing information by minimizing the new objective function $G(X) = \|X\|_r + L(X)$. Following the same optimization scheme in [Hu et al. 2013], in the l-th iteration, we firstly fix X_l and obtain A_l and B_l by conducting the SVD operation on X. And then we fix A_l and B_l by solving a nuclear norm regularized problem

$$\min_{X\in\mathbb{R}^{m\times n}} \quad H(X) = \mu(\|X\|_* - Tr(A_l W B_l^T)) + L(X). \tag{17}$$

Similar with the case in section 4, we also apply the proposed **ABSS** to solve (17) with a sight modification. Recall the detail in section 4.1 for seeking the most informative subspace, the range of the subdifferential of $H(X)$ is $[\langle \mathbf{M}_i, \nabla L(X)\rangle - \mu(1 - \frac{\partial Tr(A_l X B_l^T)}{\partial \theta_i}), \langle \mathbf{M}_i, \nabla L(X)\rangle + \mu(1 - \frac{\partial Tr(A_l X B_l^T)}{\partial \theta_i})]$ and $0 \leq \frac{\partial Tr(A_l X B_l^T)}{\partial \theta_i} \leq \|\mathbf{u}_i \mathbf{v}_i^T\|_F \leq 1$. So here the most informative subspace is $\mathcal{B} = \{\mathbf{M}_i \,|\, \langle \mathbf{M}_i, \nabla L(X)\rangle \neq 0 \text{ or } |\langle \mathbf{M}_i, \nabla L(X)\rangle| > \mu(1 - \frac{\partial Tr(A_l X B_l^T)}{\partial \theta_i}),\ \forall \mathbf{M}_i \in \mathcal{M}\} \supseteq \mathcal{A}$. However, for the ease of computation, we can choose $\mathcal{A}$ instead of $\mathcal{B}$ to select the main basis. Empirical evaluations show that this relaxation works well. In addition to this step, all the rest procedures are similar with those of nuclear norm minimization problem. In Algorithm 3 we sum all the details of applying **ABSS** to solve truncated nuclear norm regularized problem. Note that Algorithm 3 has both outer and inner iterations, where each outer iteration needs one full SVD operation. However, experimental results reveal that usually only 2 outer iterations are enough for convergence. Therefore, the computational cost of applying **ABSS** for truncated nuclear norm regularized problem is slightly more than that of nuclear norm regularized problem, but still much less than the original SVD based approach [Hu et al. 2013].

6. EXPERIMENTS

In this section, we validate our proposed method **ABSS** for matrix completion by conducting several evaluations on real image data and recommendation datasets, where the scale of image matrices is relatively small and we focus more on the recovery accuracy while that of recommendation system is much larger and efficiency is more critical. To prove the effectiveness of our proposed method, we compare the following eight matrix completion algorithms:

— **OptSpace** [Keshavan et al. 2009] : which solves matrix completion problem from a matrix factorization view based on traditional SVD.
— **R1D** [Dudik et al. 2012]: Rank one descent, which transfers the nuclear norm minimization problem to a equivalent ℓ_1 norm minimization problem and then uses coordinate descent for optimization.
— **R1MP** [Wang et al. 2014]: Rank-One Matrix Pursuit, which generates the rank-one matrix by the top singular vector pair of the current residual and then updates the weight by solving a regression problem.
— **TNNR-ADMM** [Hu et al. 2013]: Truncated Nuclear Norm Regularization via Alternating Direction Method of Multipliers (ADMM), which recovers the missing entries by minimizing the

ALGORITHM 3: The framework of applying ABSS on Truncated Nuclear Norm Regularization

Input: Rank parameter r , regularization parameter μ and tolerance ϵ.
Initialize: $X_1 \in \mathbb{R}^{m\times n}$
REPEAT
STEP 1. Given X_l

$$[U_l, \Sigma_l, V_l] = svd(X_l),$$

where

$$U_l = (\mathbf{u}_1, \cdots, \mathbf{u}_m) \in \mathbb{R}^{m\times m},$$

$$V_l = (\mathbf{v}_1, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n\times n}.$$

Compute A_l and B_l as

$$A_l = (\mathbf{u}_1, \cdots, \mathbf{u}_r)^T, B_l = (\mathbf{v}_1, \cdots, \mathbf{v}_r)^T.$$

STEP 2. Update X_{l+1} by applying **ABSS** to solve the following subproblem:

$$\min_{X\in\mathbb{R}^{m\times n}} F(X) = \mu(\|X\|_* - Tr(A_l W B_l^T)) + L(X). \tag{18}$$

until $\|X_{l+1} - X_l\|_F \leq \epsilon$.

truncated nuclear norm instead of traditional nuclear norm and utilizes ADMM method to solve it.

— **IRNN** [Lu et al. 2014]: Iteratively Reweighted Nuclear Norm, which iteratively solves a weighted singular value thresholding problem.

— **Active Subspace** [Hsieh and Olsen 2014]: which selects a small active subspace at first, then gradually enlarges the subspace until it contains the final solution.

— **ABSS**: Adaptive Basis Selection Strategy for Matrix Completion, which solves nuclear norm minimization problem by firstly pruning the basis space to get main basis, then locally refining the tail basis.

— **ABSS-TNN**: Adaptive Basis Selection Strategy for Matrix Completion on Truncated Nuclear Norm, which solves truncated nuclear norm minimization problem by applying the proposed basis selection strategy to accelerate the convergence.

Note that there are some other greedy atom decomposition algorithms for matrix completion such as GECO [Shalev-Shwartz et al. 2011], the matrix norm boosting approach [Zhang et al. 2012b]. But from the experimental results in [Wang et al. 2014], we find that R1MP outperforms these algorithms. Therefore we don't include them in our evaluation.

All experiments are conducted by using Matlab on a workstation with i7-3930 CPU and 64G memory.

6.1. Our method for nuclear norm minimization

At first, we compare the performance of **ABSS** with six state-of-art matrix completion algorithms (except **ABSS-TNN**) on image inpainting problem and recommender system.

6.1.1. Parameter Setting. For our proposed algorithm **ABSS**, we set the trade-off parameter μ in objective function (5) to be 100 and the number of selected basis N to be 100 for initialization. We only conduct the selection schemes twice to get the recovered matrix. For the consideration of efficiency, in the first selection scheme we reduce the number of total basis to be 40 and reset it to be 100 for the second scheme. In each selection scheme, we should firstly decide the number of main basis, then generate the rest tail basis. It is critical to choose the proper number of main basis for satisfactory results, which may have variations depending on the specific problem. Therefore, in

Table II. Image recovery comparison on random masked image in Fig. 3 by different methods.

(a) PSNR for different methods

Image Data	pic(a)	pic(b)	pic(c)
OptSpace	20.96	21.90	23.63
R1D	23.39	25.32	25.88
R1MP	24.12	25.81	26.88
Active SubSpace	24.11	25.72	25.70
TNNR-ADMM	24.93	26.88	26.67
IRNN	24.97	26.98	27.72
ABSS	**25.09**	**27.03**	**27.77**

(b) Running time for different methods

Image Data	pic(a)	pic(b)	pic(c)
OptSpace	15.5(s)	15.4(s)	15.9(s)
R1D	0.7(s)	0.6(s)	0.5(s)
R1MP	0.4(s)	0.4(s)	0.4(s)
Active SubSpace	8.2(s)	8.4(s)	6.2(s)
TNNR-ADMM	17.7(s)	23.4(s)	20.5(s)
IRNN	5.3(s)	4.8(s)	3.6(s)
ABSS	**1.1**(s)	**0.9**(s)	**0.9**(s)

the evaluations we try the possible main basis number from 10 to 25 and choose the one with best recovery performance to report.

All the code of the comparison algorithms are provided by their authors and for each baseline we try our best to tune the parameters to get the best recovered matrix as final result.

6.1.2. Image recovery with Random Mask. In this and next subsection, we test different matrix completion algorithms in image recovery problem. Specifically, natural image statistics have shown that images usually have low rank or approximately low rank structure, which means that the top several singular values dominate the most information of images [Zhang et al. 2012a]. In real applications, images are usually damaged by noises and texts. Mathematically, the missing information recovery from these incomplete images can be done via solving matrix completion problems.

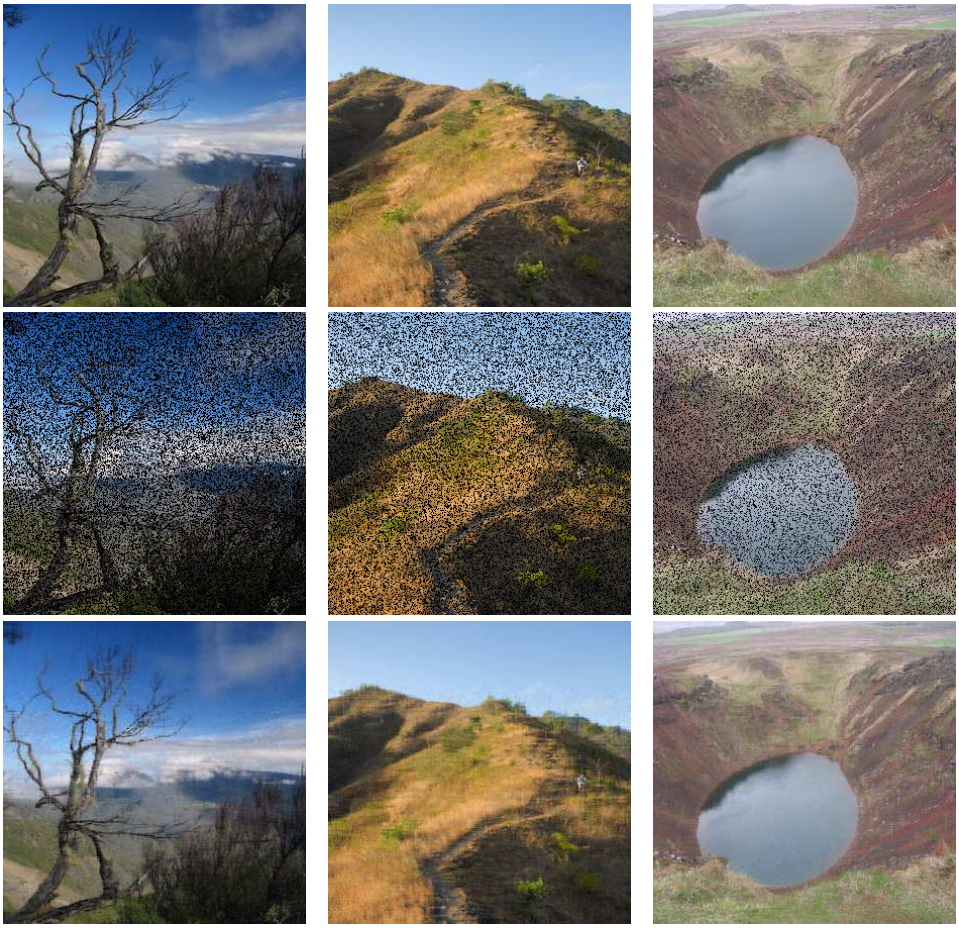
All of the test images are chosen from INRIA Holidays dataset [Jégou et al. 2008] and the Flickr 100k dataset [Philbin et al. 2007]. Since the test data is color image with three channels (red, green and blue), we deal with each channel separately and then combine the respective recovery matrices together to obtain the final result. For each test image, we firstly eliminate different ratio of pixels by masks and utilize the above seven matrix completion algorithms for inpainting. To evaluate the quality of the recovered image, we adopt the well known *Mean Square Error* (MSE) and *Peak Signal-to-Noise Ratio* (PSNR) as measures. Suppose the number of total missing pixels is P, MSE and PSNR can be defined as follows:

$$\text{MSE} = \frac{error_r^2 + error_g^2 + error_b^2}{3P}, \ \ \text{PSNR} = 10 \times \log_{10}(\frac{255^2}{MSE}). \tag{19}$$

In this subsection, we firstly conduct the image inpainting by solving matrix completion problem for a relatively easy task, where missing values are randomly distributed on the image. We randomly mask 20%, 30%, 40% pixels of one 300×300 image and then compare both PSNR values and running time for the recovered images by different matrix completion algorithms. The comparison results are presented in Fig. 3 and Table II. In Fig. 3, the top row is the original images, the middle row represents the masked images and the bottom row is the recovered images by our proposed **ABSS** with 40%, 30% and 20% missing values respectively. Table II shows the PSNR value and running time of all compared methods.

From Table II (a), we can see that **ABSS** has evident recovery accuracy advantages compared with greedy atom decomposition methods R1D and R1MP. Meanwhile, in comparison to SVD based methods such as OptSpace, Active Subspace, TNNR-ADMM and IRNN, **ABSS** also has better recovery performance. Note that although TNNR-ADMM and IRNN have the similar PSNR values to our proposed **ABSS**, we can find that **ABSS** has great advantages on efficiency from Table II(b). Specifically, **ABSS** only needs no more than 6.3% computational cost of TNNR-ADMM and 25% computational cost of IRNN. At the same time, the computational cost of **ABSS** is also comparable to greedy methods R1D and R1MP. Therefore, taking both accuracy and efficiency into consideration, our proposed **ABSS** has the best recovery performance.

6.1.3. Image Recovery with Text Mask. Different from image recovery problem with random mask in the above subsection, in real applications the positions of missing entries are usually not randomly distributed, such as the ads or texts on the images, which increases the difficulty for re-

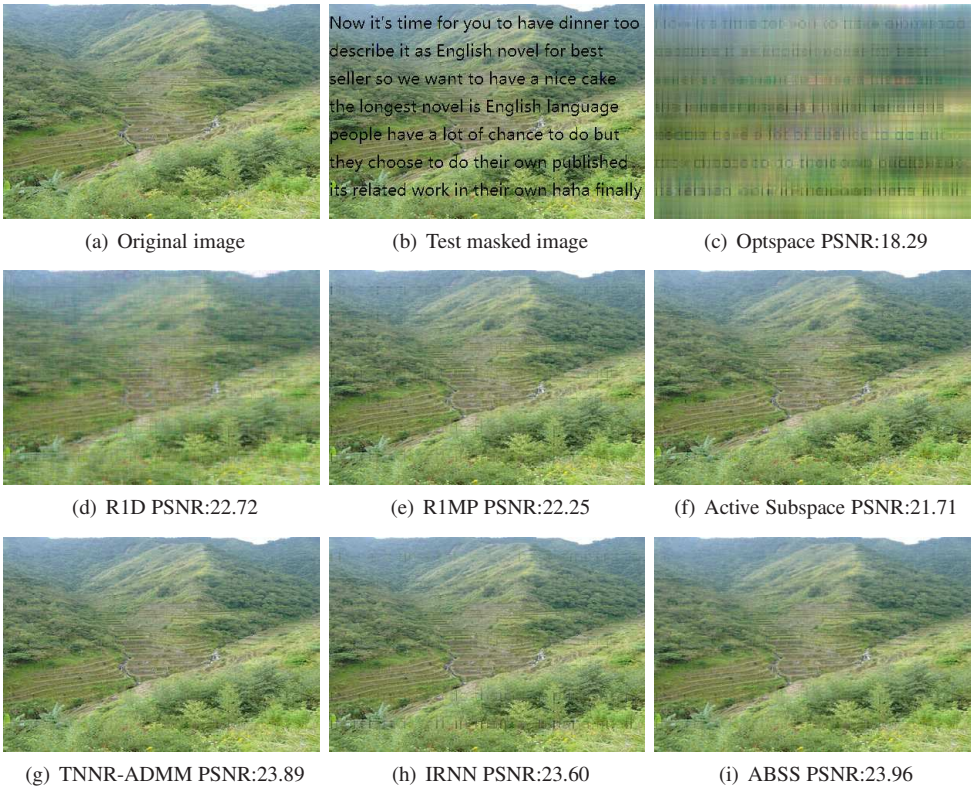


(a) 60% Observed PSNR:25.09 (b) 70% Observed PSNR:27.03 (c) 80% Observed PSNR:27.77

Fig. 3. Image recovery with random mask. We cover different percent of pixels of the original image to get a masked image and then **ABSS** is used to recover the missing values.

covery. To deal with the text removal problem, we firstly detect the positions of the missing entries. Then recovering the missing information turns into a matrix completion problem. Fig. 4 and Fig. 5 show the recovered images by different matrix completion algorithms, where we can see **ABSS** has the highest PSNR values. Table III shows the PSNR values and running time needed for recovery by different matrix completion algorithms. Similar with results in the above subsection, **ABSS** has the best recovery performance and comparable computational cost with atom decomposition based methods such as R1D and R1MP.

6.1.4. Recommendation System. In this subsection, we compare different matrix completion algorithms for large scale recommender system. Different from recovering the missing information in images, the matrices in recommender system have two striking characteristics: (1) the scale of these matrices is usually huge (larger than 100,000$\times$ 10,000), which brings unacceptable computational cost and memory requirement for matrix completion algorithms; (2) the ratio of the observed entries for these matrices is very small, e.g., we can only obtain 1% entries of the matrix from Netflix



(a) Original image (b) Test masked image (c) Optspace PSNR:18.29

(d) R1D PSNR:22.72 (e) R1MP PSNR:22.25 (f) Active Subspace PSNR:21.71

(g) TNNR-ADMM PSNR:23.89 (h) IRNN PSNR:23.60 (i) ABSS PSNR:23.96

Fig. 4. Comparison of different matrix completion algorithms for text removal problem. (a) Original image; (b) image with text; (c)-(i) are recovered images by OptSpace, R1D, R1MP, Active Subspace, TNNR-ADMM, IRNN and ABSS respectively. Experimental results show that our proposed ABSS algorithm outperforms the state of art methods.

Table III. Image recovery comparison on text masked images in Fig. 4 (b) and Fig. 5 (b) by different methods

Image Data	Image in Figure 4		Image in Figure 5	
	PSNR	Time	PSNR	Time
OptSpace	18.29	15.5(s)	18.91	14.5(s)
R1D	22.72	0.6(s)	26.24	0.6(s)
R1MP	22.25	0.4(s)	25.72	0.3(s)
Active SubSpace	21.71	16.7(s)	26.16	13.8(s)
TNNR-ADMM	23.89	16.7(s)	27.27	11.4(s)
IRNN	23.60	5.6(s)	27.25	5.2(s)
ABSS	**23.96**	**1.1**(s)	**27.43**	**1.1**(s)

dataset [Bennett and Lanning 2007]. In the following, we evaluate our proposed algorithm **ABSS** on MovieLens [Miller et al. 2003] dataset which is a widely used recommendation dataset [Zhang and Koren 2007; Toh and Yun 2010]. This dataset is collected by the GroupLens Research Project at the University of Minnesota containing rating information from users on movies. The statistics are given in table IV. The data matrix has been cleaned up such that users who have less than 20 ratings are removed. For MovieLens100K and MovieLens1M, there are 5 rating scores ranging from 1 (strongly unsatisfactory) to 5 (strongly satisfactory), and their data matrices are highly sparse (only about 6.30% entries are known for MovieLens100K and 4.46% entries are known for MovieLens1M). For MovieLens10M, there are 10 levels of scores with a step size 0.5 ranging from

(a) Original image (b) Text masked image (c) OptSpace PSNR:18.91

(d) R1D PSNR:26.24 (e) R1MP: PSNR:25.72 (f) Active SubSpace PSNR:26.16

(g) TNNR-ADMM PSNR:27.27 (h) IRNN PSNR:27.25 (i) ABSS PSNR:27.43

Fig. 5. Comparison of different matrix completion algorithms for text removal problem. (a) Original image; (b) image with text; (c)-(i) are recovered images by OptSpace, R1D, R1MP, Active Subspace, TNNR-ADMM, IRNN and ABSS respectively. Experimental results show that our proposed ABSS algorithm outperforms the competing methods.

0.5 (strongly unsatisfactory) to 5 (strongly satisfactory) and only 1.30% entries can be observed. In order for evaluation, we randomly choose 30%, 50%, 80% entries to be the training set respectively and leave the rest to be the test set.

Evaluation of recommendation algorithms has long been divided between accuracy metrics (e.g. precision/recall) and error metrics (notably, RMSE and MAE). The mathematical convenience and fitness for formal optimization methods have made error matrices like RMSE more popular [Cremonesi et al. 2010]. Suppose the existing rating locations of the test set are denoted as Ω. Just following the same way in many recent research papers [Hu et al. 2012], we also adopt the Root Mean Squared Error (RMSE) to measure the effectiveness of an algorithm for recommendation

Table IV. Statistics of the MovieLens dataset

Dataset	Row	Column	Rating	Observed Ratios
MovieLens100K	943	1682	10^5	6.30%
MovieLens1M	6040	3706	10^6	4.47%
MovieLens10M	71567	10677	10^7	1.30%

problems:

$$\text{RMSE} = \sqrt{\frac{\sum_{(i,j)\in\Omega}(X_{ij} - M_{ij})^2}{\sharp|\Omega|}}, \tag{20}$$

where X_{ij} is the recovered value and M_{ij} is the ground truth, and $\sharp|\Omega|$ is exactly the number of test ratings.

Table V shows the recovery accuracy of the unknown information in 3 datasets. We can see that **ABSS** has consistent much better recovery performance than atom decomposition based methods such as R1D, R1MP. Comparing with SVD based matrix completion methods, **ABSS** outperforms Active Subspace and IRNN in all cases. Except for the cases of Optspace on MovieLens100k with 30% training data and TNNR-ADMM on datasets such as MovieLens100k with 80% training data, **ABSS** also beats them in other evaluations. Furthermore, OptSpace, TNNR-ADMM and IRNN fail to converge in 10 hours on MoiveLens10M due to their prohibitive computational and memory cost. So we do not report the corresponding results here. Table VI shows the running time for different methods, we also find that the convergence of our proposed **ABSS** is much faster than SVD based matrix completion methods including Optspace, TNNR-ADMM and IRNN and comparable with atom decomposition methods R1D and R1MP. Note that in relatively sparse cases, the time needed for **ABSS** to reach convergence is sightly more than that for Active Subspace. This is mainly because the special sparse acceleration techniques of Active Subspace, which makes it more suitable to largely sparse case, while **ABSS** outperforms in the rest cases. Therefore, **ABSS** has the best performance from the overall consideration of recovery accuracy and efficiency for recommendation system.

The above results are consistent with our previous observation for image inpainting. The reason for SVD based matrix completion methods failing to deal with large scale matrices is that they involve multiple times of expensive full SVD operations, which limits the usage of these methods for large scale matrices in real applications. On the other hand, since atom decomposition based methods only depend on the top singular vector which is much cheaper for computing, R1D and R1MP have the least running time. However, due to the inaccurate basis selection strategies, they have worst recovery accuracy, i.e., highest RMSE value. Different from these two categories of matrix completion approaches, **ABSS** obtains the comparable recovery accuracy with SVD based methods as well as remains the efficiency advantages of atom decomposition based methods, which is the most suitable method to deal with large scale matrix completion problems.

Table V. RMSE comparisons for different methods on MovieLens Dataset

Dataset	MovieLens100k			MovieLens1M			MovieLens10M		
Training Ratio	30%	50%	80%	30%	50%	80%	30%	50%	80%
OptSpace	1.15	1.12	1.10	1.09	1.08	1.10	\	\	\
R1D	1.52	1.28	1.10	1.37	1.35	1.08	1.34	1.25	1.18
R1MP	1.75	1.61	1.41	1.57	1.32	1.20	1.42	1.23	1.15
Active SubSpace	1.34	1.15	1.10	1.19	1.03	0.95	1.24	1.09	0.93
TNNR-ADMM	1.29	1.23	1.04	1.03	0.96	0.92	\	\	\
IRNN	1.30	1.14	1.16	1.99	1.70	1.00	\	\	\
ABSS	**1.20**	**1.09**	**1.06**	**1.06**	**0.98**	**0.92**	**1.01**	**0.96**	**0.92**

Table VI. Running time comparisons for different methods on MovieLens Dataset

Dataset	MovieLens100k			MovieLensL1M			MovieLens10M		
Training Ratio	30%	50%	80%	30%	50%	80%	30%	50%	80%
OptSpace	9.38(s)	7.31(s)	8.12(s)	105.00(s)	118.35(s)	132.67(s)	\	\	\
R1D	0.93(s)	1.05(s)	1.34 (s)	11.57(s)	13.68(s)	15.67 (s)	320.18(s)	343.70(s)	356.05(s)
R1MP	0.44(s)	0.52(s)	0.73(s)	2.56 (s)	4.59(s)	7.76(s)	38.97(s)	63.19(s)	97.75(s)
Active SubSpace	1.38(s)	2.55(s)	6.19(s)	36.31(s)	57.55(s)	88.78(s)	541.37(s)	799.31(s)	1130.13(s)
TNNR-ADMM	518.76(s)	530.98(s)	523.35(s)	19588.76 (s)	19830.35(s)	25447.90(s)	\	\	\
IRNN	414.52(s)	407.07(s)	423.14(s)	17501.16 (s)	19424.88(s)	21471.58(s)	\	\	\
ABSS	**2.11(s)**	**2.45(s)**	**2.70(s)**	**25.00(s)**	**29.33(s)**	**34.51(s)**	**704.08(s)**	**745.15(s)**	**776.52(s)**

6.2. ABSS and ABSS-TNN

Aforementioned in Section 5, the newly proposed truncated nuclear norm [Hu et al. 2013] can better capture the intrinsic structure of the low rank matrices compared with traditional nuclear norm. In this subsection, we conduct a comprehensive comparison of applying our proposed atom basis selection scheme on nuclear norm and truncated nuclear norm respectively. During this evaluation, we keep the parameters of **ABSS** the same as the previous section.

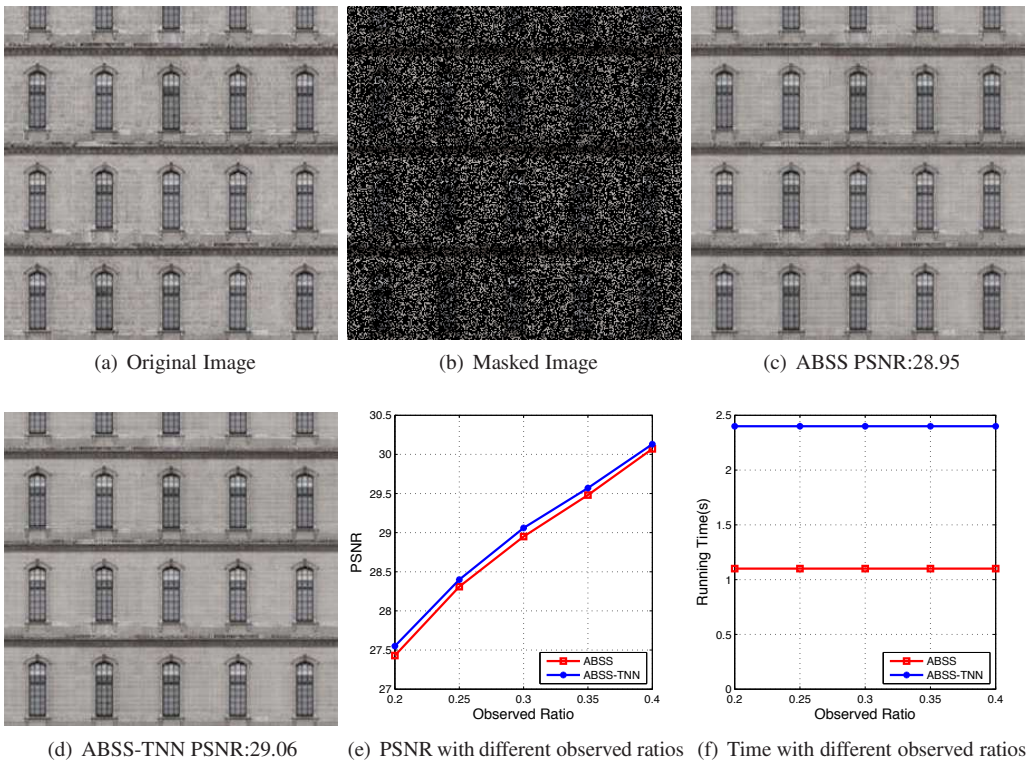
In Fig. 6 (a)-(d), we compare **ABSS** and **ABSS-TNN** in image recovery problem with random mask of 30% observed entries, where **ABSS-TNN** has better recovery accuracy. Fig. 6 (e) shows the PSNR values of these two algorithms varying observed ratio from 20% to 40%. We notice that **ABSS-TNN** has consistent better recovery performance over **ABSS**. Especially, the worse quality of the damaged image is, the advantage of **ABSS-TNN** is more evident. These facts also demonstrate that truncated nuclear norm better approximates the low rank structure. It is necessary to mention that the corresponding computational cost for **ABSS-TNN** is more expensive theoretically since it has both outer and inner loops. However, empirically **ABSS-TNN** only needs two outer loops (i.e., one full SVD operation for each outer loop) to reach convergence since each outer loop can effectively prune the solution space to escape the local minimum. Therefore, the total computation cost of **ABSS-TNN** is acceptable just as shown in Fig. 6 (f). For more difficult text removal problem in Figure 7, **ABSS-TNN** also has better PSNR values.

Overall, by the above experiments, we conclude that when the proposed adaptive basis selection scheme is applied to truncated nuclear norm, we can obtain better recovery performance with bearable extra computational cost, which is suitable for some applications with higher requirement for accuracy and fewer limitations on efficiency.

6.3. Sensitive Parameter Analysis

In this subsection, we analyze the sensitivity of the parameters in our algorithm. We aim to demonstrate whether our algorithm will fluctuate heavily with slight change of parameters and then give some guidelines on choosing proper parameters. In **ABSS**, we need to study how the final recovery performance varies with the number of the main basis s in each selection scheme, the number of selection scheme γ and the importance of initialization step. Furthermore, in **ABSS-TNN** the parameter μ to control the degree of the truncated nuclear norm regularization is also critical to tune. Therefore, we conduct the evaluation on images in Figure 7, where 60% pixels are randomly masked.

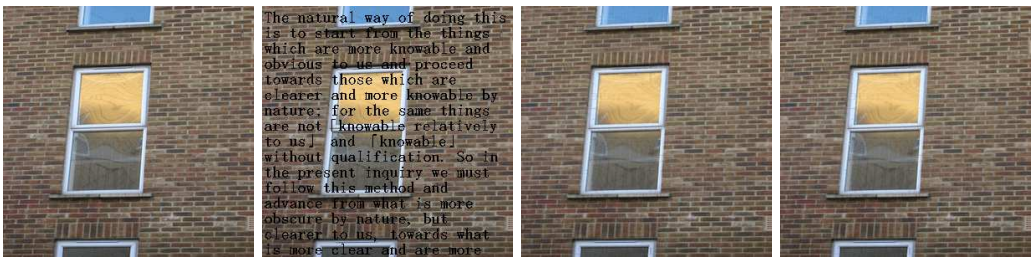
Fig. 8 shows different parameters' influence on the final recovery. Recall that the selected finite basis set consists of main basis and tail basis. Specifically, the variation of the PSNR values with different number of main basis m in each scheme is presented in Fig. 8 (a). When m increases, The PSNR firstly increases to the peak and then decreases. This is mainly because the number of main basis is fixed according to the intrinsic structure of matrices. On the other hand, it also demonstrates that main basis category and tail basis category have different structure, which lead us to use different strategies to generate the main basis and tail basis separately. Fig. 8 (b) shows the PSNR values varying with the number of the selection scheme. We can see that the PSNR values increase in succession. However, considering the efficiency issue, usually no more than 3 selection schemes are enough to obtain a satisfactory result. In Algorithm 2, we firstly adopt the



(a) Original Image (b) Masked Image (c) ABSS PSNR:28.95

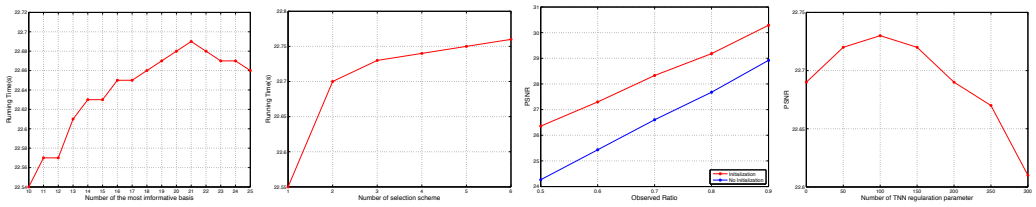
(d) ABSS-TNN PSNR:29.06 (e) PSNR with different observed ratios (f) Time with different observed ratios

Fig. 6. Comparison between **ABSS** and **ABSS-TNN** in image recovery with random mask. (a) The original image; (b) 70% masked image; (c) recovered image by **ABSS**; (d) recovered image by **ABSS-TNN**; (e) PSNR values with different observed ratios; (f) running time vs different observed ratios. These results show that **ABSS-TNN** has better performance, especially for heavily damaged images.



(a) Original Image (b) masked by text (c) ABSS:24.41 (d) ABSS-TNN:24.54

Fig. 7. Comparison between **ABSS** and **ABSS-TNN** for image recovery with text mask. (a) Original Image; (b) image masked by text; (c)recovered image by **ABSS**, PSNR=24.41; (d) recovered image by **ABSS-TNN**, PSNR=24.54.



(a) (b) (c) (d)

Fig. 8. PSNR vs different parameters of the proposed algorithm on the image of Figure 7. (a) PSNR vs number of selected basis; (b) PSNR vs number of selection scheme; (c) PSNR vs observed ratio with and without initialization;(d) PSNR vs TNN regularization parameter.

proposed *Adaptive Rank one Descent* to generate a coarse approximated matrix for initialization. To validate its necessity, we evaluate ABSS with/without this initialization step. Fig. 8 (c) shows that this initialization step can effectively improve the recovery performance. In Fig. 8 (d), we can see that PSNR value increases consistently with trade-off parameter μ in objective function (17) changing from 0 to 100 and then descends sharply. This means that too large regularization trade-off parameter μ causes high bias, it is more reasonable to choose a relatively small one.

7. CONCLUSION

In this paper, we propose an efficient and scalable atom decomposition method called *Adaptive Basis Selection Strategy* (**ABSS**) for matrix completion to estimate the missing entries accurately and efficiently. The key idea is to adopt different strategies for selecting the basis according to their corresponding characteristics in different phases respectively. By firstly pruning the basis space to seek several most informative basis in a global way, we guarantee the correct solution path along the convergence routine. And then we propose a local based strategy named *Adaptive Rank one Descent* to greedily generate the basis and the coefficients for better fitting the residual matrix. Comparing with the existing atom decomposition methods, the set of basis selected by our approach is more consistent with its innate nature. Thus the recovered matrix is more accurate. At the same time, **ABSS** inherits the great efficiency advantages of atom decomposition based methods, which is more suitable for large scale matrix completion problems. Extensive experimental results on real image data and recommendation datasets prove the effectiveness of our proposed method on both recovery accuracy and efficiency.

For future research, we will focus on the huge scale matrix completion problem, where the total data matrix even cannot be stored in the memory. It is interesting to explore how to apply our method in a distributed way to solve such huge scale matrix completion problems.

REFERENCES

Andreas Argyriou, Theodoros Evgeniou, and Massimiliano Pontil. 2008. Convex Multi-task Feature Learning. *Machine Learning* 73, 3 (2008), 243–272.

James Bennett and Stan Lanning. 2007. The netflix prize. In *Proceedings of ACM SIGKDD Cup and Workshop*.

Jian-Feng Cai, Emmanuel J. Candès, and Zuowei Shen. 2010. A Singular Value Thresholding Algorithm for Matrix Completion. *SIAM Journal on Optimization* 20, 4 (2010), 1956–1982.

Emmanuel J. Candès and Yaniv Plan. 2010. Matrix Completion With Noise. *Proc. IEEE* 98, 6 (2010), 925–936.

Emmanuel J. Candès and Benjamin Recht. 2012. Exact Matrix Completion via Convex Optimization. *Foundations of Computational mathematics* 55, 6 (2012), 111–119.

Emmanuel J. Candès and Terence Tao. 2010. The Power of Convex Relaxation: Near-optimal Matrix Completion. *IEEE Transactions on Information Theory* 56, 5 (2010), 2053–2080.

Bisheng Chen, Jingdong Wang, Qinghua Huang, and Tao Mei. 2012. Personalized video recommendation through tripartite graph propagation. In *Proceedings of ACM International Conference on Multimedia*.

Ke Chen. 2005. *Matrix preconditioning techniques and applications*. Cambridge University Press.

Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. 2001. Atomic Decomposition by Basis Pursuit. *SIAM Journal on scientific computing* 43, 1 (2001), 129–159.

Paolo Cremonesi, Yehuda Koren, and Roberto Turrin. 2010. Performance of recommender algorithms on top-n recommendation tasks. In *Proceedings of ACM Conference on Recommender systems*.

MIroslav Dudik, Zaid Harchaoui, and Jérome Malick. 2012. Lifted Coordinate Descent for Learning with Trace Norm Regularization. In *Proceedings of International Conference on Artificial Intelligence and Statistics*.

Maryam Fazel. 2002. *Matrix rank minimization with applications*. Ph.D. Dissertation. PhD thesis, Stanford University.

Cho-Jui Hsieh and Peder Olsen. 2014. Nuclear norm minimization via active subspace selection. In *Proceedings of International Conference on Machine Learning*.

Yao Hu, Debing Zhang, Jun Liu, Jieping Ye, and Xiaofei He. 2012. Accelerated singular value thresholding for matrix completion. In *Proceedings of ACM SIGKDD International Conference on Knowledge discovery and data mining*.

Yao Hu, Debing Zhang, Jieping Ye, Xuelong Li, and Xiaofei He. 2013. Fast and Accurate Matrix Completion via Truncated Nuclear Norm Regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 9 (2013), 2117–2130.

Hervé Jégou, Matthijs Douze, and Cordelia Schmid. 2008. Hamming embedding and weak geometric consistency for large scale image search. In *Proceedings of European Conference on Computer Vision*.

Hui Ji, Chaoqiang Liu, Zuowei Shen, and Yuhong Xu. 2010. Robust video denoising using low rank matrix completion. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*.

Raghunandan H. Keshavan, Sewoong Oh, and Andrea Montanari. 2009. Matrix completion from a few entries. In *IEEE International Symposium on Information Theory*. 324–328.

Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix Factorization Techniques for Recommender Systems. *Computer* 42, 8 (2009), 30–37.

Ji Liu, Ryohei Fujimaki, and Jieping Ye. 2014. Forward-backward greedy algorithms for general convex smooth functions over a cardinality constraint. In *Proceedings of International Conference on Machine Learning*.

Canyi Lu, Jinhui Tang, Shuicheng Yan, and Zhouchen Lin. 2014. Generalized nonconvex nonsmooth low-rank minimization. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*.

Stéphane G. Mallat and Zhifeng Zhang. 1993. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing* 41, 12 (1993), 3397–3415.

Bradley N. Miller, Istvan Albert, Shyong K. Lam, Joseph A. Konstan, and John Riedl. 2003. MovieLens unplugged: experiences with an occasionally connected recommender system. In *Proceedings of International Conference on Intelligent user interfaces*.

Karthik Mohan and Maryam Fazel. 2010. Iterative reweighted least squares for matrix rank minimization. In *Annual Allerton Conference on Communication, Control, and Computing (Allerton)*.

Sahand Negahban, Martin J Wainwright, and others. 2011. Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annals of Statistics* 39, 2 (2011), 1069–1097.

Yagyensh Chandra Pati, Ramin Rezaiifar, and PS Krishnaprasad. 1993. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proceedings of Asilomar Conference on Signals, Systems and Computers*.

James Philbin, Ondrej Chum, Michael Isard, Josef Sivic, and Andrew Zisserman. 2007. Object retrieval with large vocabularies and fast spatial matching. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*.

Benjamin Recht. 2011. A Simpler Approach to Matrix Completion. *Journal of Machine Learning Research* 12 (2011), 3413–3430.

Shai Shalev-Shwartz, Alon Gonen, and Ohad Shamir. 2011. Large-Scale Convex Minimization with a Low-Rank Constraint. In *Proceedings of International Conference on Machine Learning*.

Shai Shalev-Shwartz and Ambuj Tewari. 2011. Stochastic methods for l 1-regularized loss minimization. *Journal of Machine Learning Research* 12 (2011), 1865–1892.

Nathan Srebro and Adi Shraibman. 2005. Rank, trace-norm and max-norm. In *Learning Theory*. Springer, 545–560.

Jos F. Sturm. 1999. Using SeDuMi 1.02, A Matlab toolbox for optimization over symmetric cones. *Optimization Methods and Software* 11, 1-4 (1999), 625–653.

Kim-Chuan Toh and Sangwoon Yun. 2010. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pacific Journal of Optimization* 6, 615-640 (2010), 15.

Joel A. Tropp and Anna C. Gilbert. 2007. Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit. *IEEE Transactions on Information Theory* 53, 12 (2007), 4655–4666.

RH Tütüncü, KC Toh, and MJ Todd. 2001. SDPT3—a Matlab software package for semidefinite-quadratic-linear programming, version 3.0. *Web page http://www. math. nus. edu. sg/˜ mattohkc/sdpt3.html* (2001).

Shusen Wang, Dehua Liu, and Zhihua Zhang. 2013. Nonconvex relaxation approaches to robust matrix recovery. In *Proceedings of International Joint Conference on Artificial Intelligence*.

Zheng Wang, Ming-Jun Lai, Zhaosong Lu, Wei Fan, Hasan Davulcu, and Jieping Ye. 2014. Rank-One Matrix Pursuit for Matrix Completion. In *Proceedings of International Conference on Machine Learning*.

Stephen J. Wright, Robert D. Nowak, and Mário A.T. Figueiredo. 2009. Sparse Reconstruction by Separable Approximation. *IEEE Transactions on Signal Processing* 57, 7 (2009), 2479–2493.

Harchaoui Zaid, Matthijs Douze, Mattis Paulin, Miroslav Dudik, and Jérôme Malick. 2012. Large-scale image classification with trace-norm regularization. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*.

Debing Zhang, Yao Hu, Jieping Ye, Xuelong Li, and Xiaofei He. 2012a. Matrix completion by truncated nuclear norm regularization. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*.

Tong Zhang. 2009. Adaptive forward-backward greedy algorithm for sparse learning with linear models. In *Proceedings of Advances in Neural Information Processing Systems*.

Xinhua Zhang, Dale Schuurmans, and Yao-liang Yu. 2012b. Accelerated training for matrix-norm regularization: A boosting approach. In *Proceedings of Advances in Neural Information Processing Systems*.

Yi Zhang and Jonathan Koren. 2007. Efficient bayesian hierarchical user modeling for recommendation system. In *Proceedings of International ACM SIGIR Conference on Research and development in information retrieval*.

Ciyou Zhu, Richard H. Byrd, Peihuang Lu, and Jorge Nocedal. 1997. Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-scale Bound-constrained Optimization. *ACM Transactions on Mathmatical Software* 23, 4 (1997), 550–560.

Jinfeng Zhuang, Tao Mei, Steven C. H. Hoi, Xian-Sheng Hua, and Yongdong Zhang. 2015. Community Discovery from Social Media by Low-Rank Matrix Recovery. *ACM Transactions on Intelligent Systems and Technology* 5, 4 (2015), 1–19.