논문 2009-46SP-5-3

낮은 계수 행렬의 Compressed Sensing 복원 기법

(Compressed Sensing of Low-Rank Matrices: A Brief Survey on Efficient Algorithms)

이 기 륭*, 예 종 철**

(Kiryung Lee and Jong Chul Ye)

요 약

Compressed sensing은 소수의 선형 관측으로부터 sparse 신호를 복원하는 문제를 언급하고 있다. 최근 벡터 경우에서의 성공적인 연구 결과가 행렬의 경우로 확장되었다. Low-rank 행렬의 compressed sensing은 ill-posed inverse problem을 low-rank 정보를 이용하여 해결한다. 본 문제는 rank 최소화 혹은 low-rank 근사의 형태로 나타내질 수 있다. 본 논문에서는 최근 제안된 여러 가지 효율적인 알고리즘에 대한 survey를 제공한다.

Abstract

Compressed sensing addresses the recovery of a sparse vector from its few linear measurements. Recently, the success for the vector case has been extended to the matrix case. Compressed sensing of low-rank matrices solves the ill-posed inverse problem with the low-rank prior. The problem can be formulated as either the rank minimization or the low-rank approximation. In this paper, we survey recently proposed efficient algorithms to solve these two formulations.

Keywords: Compressed sensing, singular value decomposition, restricted isometry property, low-rank approximation.

I. Introduction

Compressed sensing addresses the problem of reconstructing a sparse vector from a few linear measurements. As an inverse problem, it is ill-posed since the number of measurement is assumed to be much smaller than the dimension of the embedding vector space. However, it has been shown that the use of the sparsity prior leads to a successful recovery. The sparsity level, which is the number of nonzero elements of a given vector, denotes the parsimony of the representation of the vector with respect to the standard basis. This argument has been extended to the matrix case, where the parsimony is denoted by the rank of the matrix. Compressed sensing of low-rank matrices can be interpreted as an extension of compressed sensing for the vector case to the matrix case. There are many applications of compressed sensing of low-rank matrices, which include low-order system identification in control theory, low-dimensional Euclidean embedding in geometry. For more applications, see [1], [2], and the references therein.

The inverse problem of compressed sensing for the matrix case addresses the reconstruction of a matrix $X \in K^{m \times n}$ from its linear measurement $b \in K^p$ given by $b = AX$ where K denotes the scalar field, which can be either R or C, and A is a given linear

* 정회원, 일리노이 주립대학교
(University of Illinois at Urbana- Champaign)
** 정회원, 한국과학기술원
(KAIST)
※ Grant sponsor: Korea Science and Engineering Foundation: Grant number: 2009-0081089
접수일자: 2009년6월21일, 수정완료일: 2009년9월3일

operator $A : K^{m \times n} \rightarrow K^p$. In this paper, we consider two special linear operators: As a direct extension of the compressed sensing for the vector case to the matrix case, the reconstruction of a low-rank matrix from its random linear measurements. Here the linear operator A takes the ensembles of a given matrix and can be considered as the collection of p random $m \times n$ matrices. Matrix completion, also known as the Netflix problem or the collaborative filtering, is a special case of compressed sensing of low-rank matrices where A is given explicitly in the following way. Let $\Omega \subset \{1,\cdots,m\} \times \{1,\cdots,n\}$ be a set of indices of the matrix. Given Ω, a projection operator $P_\Omega : K^{m \times n} \rightarrow K^{m \times n}$ is defined by

$$(P_\Omega M)_{i,j} = \begin{cases} M_{i,j} & \text{if } (i,j) \in \Omega \\ 0 & else \end{cases}$$

Let $p = |\Omega|$. Then a linear operator $R_\Omega : K^{m \times n} \rightarrow K^p$ is defined by assigning $R_\Omega M$ a vector obtained by staking nonzero elements of $P_\Omega M$. The matrix completion corresponds to the case when $A = R_\Omega$.

The algorithms reviewed in this paper are categorized into two classes depending on the formulation. First, the inverse problem can be formulated as the rank minimization problem, which is formally written as:

$$P1: \quad \min_X \ rank(X)$$
$$\text{subject to } AX = b.$$

When the measurement b is corrupted by an additive noise, the formulation is given by

$$P1': \quad \min_X \ rank(X)$$
$$\text{subject to } \|AX - b\|_2^2 \leq \epsilon.$$

An alternative way to solve the inverse problem is to formulate the inverse problem as the minimum rank approximation problem.

$$P2: \quad \min_X \ \|AX - b\|_2^2$$
$$\text{subject to } rank(X) \leq r.$$

Two formulations can be compared in the following perspective: When the minimum rank r of the true solution is known, P2 can handle both the noiseless and the noisy cases in a single formulation. However, in practice, r is not available. The minimum of P2 is monotone decreasing in r and the value can be found efficiently at more cost. Indeed, this is not an issue in many applications where the rank is assumed to be a small constant.

The goal of this paper is to review the recent algorithms to solve the compressed sensing of low-rank matrices. The effort has been made to provide the motivation and the intuition of the algorithms in an unified framework. The preliminaries and the notations will be introduced in Section II. Section III will review the algorithms that solve the rank minimization problems P1 and P2. In Section IV, the algorithms to solve the low rank approximation problem P2 will be discussed. The comparison of the reviewed algorithms will be given in Section V.

II. Preliminaries

Throughout this paper, we use two vector spaces: the space of column vectors K^p and the space of matrices $K^{m \times n}$. The scalar field K will be either R or C depending on the algorithms. For C^p, the inner product is defined by $< x, y \geqq y^H x$ for $x, y \in C^p$, where y^H denotes the Hermitian transpose of y, and the induced Hilbert-Schmidt norm is the Euclidean or l_2-norm given by $\|x\|_2^2 = < x, x >$ for $x \in C^p$. For $C^{m \times n}$, the inner product is defined by $< X, Y \geqq tr(Y^H X)$ for $X, Y \in C^{m \times n}$, and the induced norm is the Frobenius norm given by $\|X\|_F^2 = < X, X >$ for $X \in C^{m \times n}$. For the real case where $K = R$, the Hermitian transpose is replaced by the transpose denoted by T.

One important tool used in many algorithms reviewed in this paper is the singular value decomposition (SVD) of a matrix. Every matrix $X \in K^{m \times n}$ admits a singular value decomposition

$X=\sum_{k=1}^{\min(m,n)}\sigma_k u_k v_k^H$ where $\{\sigma_k\}_{k=1}^{\min(m,n)}$ denote the singular values in decreasing order, and (u_k, v_k) denotes the pair of the left and right singular vectors associated to σ_k. The singular values are always nonnegative and the singular vectors are orthonormal. Furthermore, Eckart-Young-Mirsky theorem [19] states that $\sum_{k=1}^{s}\sigma_k u_k v_k^H$ is a minimizer to

$$\min_Y\{\|X-Y\|: rank(Y)\le s\}$$

where $\|\ \|$ can be any unitarily invariant norm such as the Frobenius norm or the operator norm. The rank of a given matrix X is defined as the number of nonzero singular values.

At this point, it might be useful to remind the analogous notions for the vector case. For a vector $x\in K^p$, we consider a decomposition $x=\sum_{k=1}^{p}x_k e_k$ with respect to the standard basis $\{e_k\}_{k=1}^{p}$. The sparsity level of x is defined as the number of nonzero x_k's and is also known as the l_0-norm. For both the matrix and the vector cases, the rank and the sparsity level are the pseudo norms that quantify the parsimony in the representation of a matrix X and a vector x, respectively.

Besides the similarity in the measure of the parsimony, there is more analogy between the vector case and the matrix case for the compressed sensing. This analogy has been proposed by Recht, Fazel, and Parrilo[2]. In particular, the useful tools to derive the performance guarantees for various compressed sensing algorithms for the vector case have analogues for the matrix case. The first tool is the rank-restricted isometry property defined in the following way: Given a linear operator $A: K^{m\times n}\to K^p$, the rank-restricted isometry constant $\delta_r(A)$ is defined as the minimum constant that satisfies

$$(1-\delta_r(A))\|X\|_F\le\|AX\|_2\le(1+\delta_r(A))\|X\|_F, \qquad (1)$$

for all $X\in K^{m\times n}$ with $rank(X)\le r$. Intuitively, the property implies that the linear operator A behaves well like an orthogonal system if its domain is restricted as the set of low-rank matrices. Recht et. al. [2] showed that there exist nearly isometric families including the random linear operator that takes the ensembles of a given matrix.

Unfortunately, the linear operator R_Ω in the matrix completion does not satisfy the rank-restricted isometry. Candes and Recht [3] have shown that the linear operator R_Ω can also behave well if its domain is further restricted to the low-rank matrices that are incoherent to R_Ω..

III. Rank Minimization

A. Convex Relaxation

The rank of a matrix is an analogue of the sparsity level of a vector. Both are pseudo norms and non-convex. Fazel et. al. [4] proposed a heuristic called the nuclear norm minimization where the nuclear norm is defined by

$$\|X\|_*=\sum_{k=1}^{\min(m,n)}\sigma_k$$

for a matrix X with the singular value decomposition $X=\sum_{k=1}^{m}\sigma_k u_k v_k^H$. The dual norm of the nuclear norm is the operator norm defined by

$$\|X\|_2=\max_k\sigma_k.$$

Indeed, the nuclear norm is a convex surrogate of $rank(X)$ for the set $\{X: \|X\|_2\le 1\}$.

Nuclear norm minimization is a convex relaxation of P1 and is formulated as

$$P3: \quad \min_X \|X\|_*$$
$$\text{subject to } AX=b.$$

Similarly, for the noisy case, a convex relaxation of P1′ is formulated as

$P3'$: $\min_X \|X\|_*$

subject to $\|AX-b\|_2^2 \le \epsilon$.

They are both the analogue of the l_1-norm minimization for the vector case. There is a performance guarantee for P3 (or P3') which will be discussed in Section V with a comparison to other guaranteed algorithms.

B. Semidefinite Programming

For the vector case, the l_1-norm minimization is cast as a linear programming. For the matrix case, the nuclear norm minimization can be formulated as a semidefinite programming (SDP) in the following way [4], [2]. First, the variational characterization of the nuclear norm with respect to its dual norm is given by

$$\|X\|_* = \max_Y \{< X, Y>: \|Y\|_2 \le 1\},$$

By the Schur decomposition, it can be rewritten as

$$\max_Y \quad tr(Y^T X)$$

$$\text{subject to } \begin{bmatrix} I_m & Y \\ Y^T & I_n \end{bmatrix} \ge 0.$$

Here they considered the real case where $K=R$. Otherwise, the objective should be replaced by $|tr(Y^T X)|$ to get an appropriate ordering. Given SDP is a convex optimization with no duality gap and its dual is given by

$$\min_{W_1, W_2} \quad \frac{1}{2}(tr(W_1)+tr(W_2))$$

$$\text{subject to } \begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \ge 0.$$

From this expression of the nuclear norm $\|X\|_*$, nuclear norm minimization problem P3 is a SDP given by

$$\min_{X, W_1, W_2} \quad \frac{1}{2}(tr(W_1)+tr(W_2))$$

$$\text{subject to } \begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \ge 0$$

$$AX = b.$$

The affine constraint in the noiseless case will be replaced by a quadratic constraint for the noisy case. SDP admits an interior point method which solves the problem in a polynomial time. However, the computation of the Hessian matrix prevents the interior point method to scale up to large problems. Alternatively, the first order methods can be used to improve the scalability at the cost of a slower convergence.

C. Projected Subgradient Method

Recht, Fazel, and Parrilo [2] have also discussed the use of the subgradient method to solve P3. The nuclear norm minimization P3 (and P3') can be considered as a constrained optimization problem where the objective function is nonsmooth. For the noiseless case, the feasible set of P3 is a linear variety determined by the affine constraint $AX=b$. For the noisy case, the feasible set of P3' is an ellipsoid determined by the quadratic constraint $\|AX-b\|_2^2 \le \epsilon$. In both cases, the feasible sets are convex and the projection onto these sets can be easily computed. Furthermore, the subdifferential of the nuclear norm at X is explicitly given by

$$\partial\|X\|_* = \{UV^H + W: U^T W = 0, WV = 0, \|W\|_2 \le 1\},$$

where $X = U\Sigma V^H$ denotes the singular value decomposition of X. For the derivation, see [5] and the reference therein. Then the subgradient method generates a sequence of the feasible points with the following update

$$X_{k+1} \leftarrow \Pi(X_k - s_k Y_k), \quad Y_k \in \partial\|X_k\|_*$$

where Π denotes the projection onto the feasible set and s_k is the step size for each iteration. However, the convergence result is not available for this approach.

D. Singular Value Thresholding

To overcome the SDP formulation of the nuclear norm minimization, Cai, Candes, and Shen [6] proposed an efficient algorithm to solve P3, which is based on the proximal method in the optimization. They considered the regularization of the problem by $\|X\|_F^2$.

$$P_\tau: \quad \min_{X\in K^{m\times n}} \quad \tau\|X\|_* + \frac{1}{2}\|X\|_F^2$$

subject to $AX = b$,

First they showed the following relation between the original problem and its regularization. Let X^τ denote the solution to P_τ then X^τ converges to the solution to P1 in Frobenius norm as $\tau\to\infty$. The benefit of the regularized formulation is that the dual problem admits an efficient algorithm. Since P_τ is a convex optimization, there is no duality gap. The Lagrangian is given by

$$L(X,y) = \tau\|X\|_* + \frac{1}{2}\|X\|_F^2 + < y, b - AX >.$$

The dual problem is an unconstrained maximization of the objective function $g(y) = \min_X L(X,y)$, which is concave and non-differentiable. They used a subgradient method to solve the dual problem. A subgradient of min-type function $g(y)$ is given by

$$\frac{\partial}{\partial y} L(\widetilde{X}_y, y) = b - A\widetilde{X}_y$$

where $\widetilde{X}_y = \arg\min_X L(X,y)$.

$$\widetilde{X}_y = \arg\min_X \tau\|X\|_* + \frac{1}{2}\|X - A^* y\|_2^2 \qquad (2)$$

Indeed, the unique solution to (2) is given by $\widetilde{X}_y = D_\tau(A^* y)$. Here $D_\tau(M)$ denotes the singular-value-thresholding operator and is defined by

$$D_\tau(M) = Udiag(\{\max(0, \sigma_k - \tau)\}_{k=1}^{\tau}) V^H$$

where $M = Udiag(\{\sigma_k\}_{k=1}^{\tau}) V^H$ is the singular value decomposition of M.

Algorithm 1 SVT

1: $y \leftarrow 0$
2: **while** stop criterion is false **do**
3: $\quad \hat{X} \leftarrow \mathcal{D}_\tau(\mathcal{A}^* y)$
4: $\quad y \leftarrow y + \delta(b - \mathcal{A}X)$
5: **end while**
6: **return** $\hat{X}$

Ⅳ. Low Rank Approximation

A. Greedy Algorithms

Greedy algorithm such as Matching Pursuit (MP) [8] and Orthogonal Matching Pursuit (OMP) [9] have been proposed as efficient algorithms for the compressed sensing for the vector case. Lee and Bresler [7] introduced the notion of the atomic decomposition to extend the greedy algorithms for the vector case to the matrix case.

A matrix $X\in K^{m\times n}$ satisfies $rank(X) \le r$ if and only if X is spanned by r rank-one matrices. Atomic decomposition denotes the representation of a given matrix as a linear combination of rank-one matrices. The set of atoms O of $K^{m\times n}$ is defined as a refinement of all rank-one matrices such that any two distinct matrices in O are not collinear.

The correlation maximization in greedy algorithms for the vector case is extended to the matrix case by using the atomic decomposition. Matching Pursuit (MP) [8] and Orthogonal Matching Pursuit (OMP) [9] choose the index $k\in\{1,\cdots,n\}$ that maximizes the correlation $|a_k^H(b - A\hat{x})|$ between the k-th column a_k of A and the residual in each iteration, where $\hat{x}$ is the solution of the previous iteration. Given a set Ψ, let P_Ψ denote the projection operator onto the subspace spanned by Ψ in the corresponding embedding space. When $\Psi = \{\psi\}$ is a singleton set, P_ψ will denote P_Ψ. From

$$\begin{aligned} |a_k^H(b - A\hat{x})| &= |< A^H(b - A\hat{x}), e_k >_{K^n}| \\ &= \|P_{e_k} A^H(b - A\hat{x})\|_2 \end{aligned}$$

it follows that maximizing the correlation implies

maximizing the norm of the projection of the image under A^H of the residual $b - A\hat{x}$ onto the selected one dimensional subspace.

The following selection rule generalizes the correlation maximization to the matrix case.

$$\max_{\psi \in O} |< b - A\hat{X}, A\psi >_{K^{m\times n}}| = \max_{\psi \in O} \|P_\psi A^*(b - A\hat{X})\|_F \quad (3)$$

where $A^*: K^p \to K^{m\times n}$ denotes the adjoint operator of A By the Eckart-Young Theorem, the basis of the best subspace is obtained from the singular value decomposition of $M = A^*(b - A\hat{X})$, as $\psi = u_1 v_1^H$, where u_1 and v_1 are the principal left and right singular vectors. Applying the selection rule (3) to update $\hat{X}$ recursively leads to greedy algorithms generalizing MP and OMP to the matrix case.

Atomic decomposition also generalize the selection rule in recent greedy algorithms such as CoSaMP [10] and Subspace Pursuit [11]. The selection rule chooses the subset J of $\{1,\cdots,n\}$ with $|J| = s$ defined by

$$|a_k^H(b - A\hat{x})| \geq |a_j^H(b - A\hat{x})|, \quad \forall k \in J, \forall j \notin J. \quad (4)$$

This is equivalent to maximizing

$$\sum_{k\in J} |a_k^H(b - A\hat{x})|^2 = \|P_{\{e_k\}_{k\in J}} A^H(b - A\hat{x})\|_2^2.$$

In other words, selection rule (4) finds the best subspace spanned by s elements in E that maximizes the norm of the projection of $M = A^H(b - A\hat{x})$ onto that s-dimensional subspace.

The following selection rule generalizes the selection rule (4) to the matrix case.

$$\max_{\Psi \subset O} \left\{ \|P_\Psi A^*(b - A\hat{X})\|_F : \ |\Psi| \leq \tau \right\}$$

A basis Ψ of the best subspace is again obtained from the singular value decomposition of $M = A^*(b - A\hat{X})$, as $\Psi = \{\rho_k u_k v_k^H\}_{k=1}^{\tau}$, where u_k and v_k, $k = 1,\cdots,\tau$ are the τ principal left and right singular vectors, respectively and for each k, $\rho_k \in K$ satisfies $|\rho_k| = 1$. Note that Ψ is an orthonormal set although this is not enforced as an explicit constraint in the maximization.

Algorithm 2 MP for Low Rank Approximation

1: $y \leftarrow b, \hat{X} \leftarrow 0$
2: **for** k = 1,...,r **do**
3: $\psi \leftarrow \arg\max_{\psi \in \mathbf{O}} \|P_\psi \mathcal{A}^* y\|_F$
4: $Z \leftarrow \arg\min_X \{\|b - \mathcal{A}X\|_2 : X \in \mathrm{span}(\psi)\}$
5: $\hat{X} \leftarrow \hat{X} + Z$
6: $y \leftarrow y - \mathcal{A}Z$
7: **end for**
8: **return** $\hat{X}$

Algorithm 3 OMP for Low Rank Approximation

1: $y \leftarrow b, \hat{X} \leftarrow 0, \Psi \leftarrow \emptyset$
2: **for** k = 1,...,r **do**
3: $\psi \leftarrow \arg\max_{\psi \in \mathbf{O}} \|P_\psi \mathcal{A}^* y\|_F$
4: $\Psi \leftarrow \Psi \cup \{\psi\}$
5: $\hat{X} \leftarrow \arg\min_X \{\|b - \mathcal{A}X\|_2 : X \in \mathrm{span}(\Psi)\}$
6: $y \leftarrow b - \mathcal{A}\hat{X}$
7: **end for**
8: **return** $\hat{X}$

With the generalized correlation maximization, Lee and Bresler [7], [12] proposed Atomic Decomposition Minimum Rank Approximation (ADMiRA) which is analogous to CoSAMP [10] for the vector case. In a similar way, the subspace pursuit can be extended to the matrix case. ADMiRA refines the estimation of the atoms that spans the approximate low-rank solution. A performance guarantee of ADMiRA has been derived in terms of the rank-restricted isometry. The performance guarantee will be discussed in Section V.

B. Power Factorization

Haldar and Diego [13] proposed using the power factorization to solve P2, which is an alternating optimization method to minimize a bi-convex objective function. A matrix $X \in K^{m\times n}$ satisfies $rank(X) \leq r$ if and only if $X = LR^H$ where $L \in K^{m\times r}$ and $R \in K^{n\times r}$. With this factorization, the objective function in P2 is rewritten as $\frac{1}{2}\|b - A(LR^H)\|_2^2$, which is a bi-convex function in (L,R), i.e. the function is convex in L for fixed R and is also convex in R for fixed L. This implies that once the other variable is fixed the problem

reduces to a least square problem. The convergence of the algorithm is shown in flavor of the monotone decrease of the objective. The optimality of the obtained solution is not guaranteed.

Algorithm 4 ADMiRA

1: $\widehat{X} \leftarrow 0,\ \widehat{\Psi} \leftarrow \emptyset$
2: **while** stop criterion is false **do**
3: $\Psi' \leftarrow \arg\max_{\Psi \subset \Omega} \left\{ \left\| P_\Psi \mathcal{A}^*(b - \mathcal{A}\widehat{X}) \right\|_F : |\Psi| \leqslant 2r \right\}$
4: $\widetilde{\Psi} \leftarrow \Psi' \cup \widehat{\Psi}$
5: $\widetilde{X} \leftarrow \arg\min_X \left\{ \|b - \mathcal{A}X\|_2 : X \in \mathrm{span}(\widetilde{\Psi}) \right\}$
6: $\widehat{\Psi} \leftarrow \arg\max_{\Psi \subset \Omega} \left\{ \left\| P_\Psi \widetilde{X} \right\|_F : |\Psi| \leqslant r \right\}$
7: $\widehat{X} \leftarrow P_{\widehat{\Psi}} \widetilde{X}$
8: **end while**
9: **return** $\widehat{X}$

Algorithm 5 SP for Low Rank Approximation

1: $\widehat{X} \leftarrow 0,\ \widehat{\Psi} \leftarrow \emptyset$
2: **while** stop criterion is false **do**
3: $\Psi' \leftarrow \arg\max_{\Psi \subset \Omega} \left\{ \left\| P_\Psi \mathcal{A}^*(b - \mathcal{A}\widehat{X}) \right\|_F : |\Psi| \leqslant r \right\}$
4: $\widetilde{\Psi} \leftarrow \Psi' \cup \widehat{\Psi}$
5: $\widetilde{X} \leftarrow \arg\min_X \left\{ \|b - \mathcal{A}X\|_2 : X \in \mathrm{span}(\widetilde{\Psi}) \right\}$
6: $\widehat{\Psi} \leftarrow \arg\max_{\Psi \subset \Omega} \left\{ \left\| P_\Psi \widetilde{X} \right\|_F : |\Psi| \leqslant r \right\}$
7: $\widehat{X} \leftarrow \arg\min_X \left\{ \|b - \mathcal{A}X\|_2 : X \in \mathrm{span}(\widehat{\Psi}) \right\}$
8: **end while**
9: **return** $\widehat{X}$

C. Manifold Optimization

Keshavan, Oh, and Montanari [14] considered the matrix completion for the real case where the linear operator $\mathcal{A}$ is R_Ω.

Algorithm 6 Power Factorization

1: $L \leftarrow L_0,\ R \leftarrow R_0$
2: **while** stop criterion is false **do**
3: $L \leftarrow \arg\min_L \frac{1}{2} \|b - \mathcal{A}(LR^H)\|_2^2$ for fixed R
4: $R \leftarrow \arg\min_R \frac{1}{2} \|b - \mathcal{A}(LR^H)\|_2^2$ for fixed L
5: **end while**
6: **return** $\widehat{X}$

A matrix $X \in R^{m \times n}$ satisfies $rank(X) \leq r$ if and only if $X = USV^T$ where $U \in R^{m \times r}$ and $V \in R^{n \times r}$ are orthonormal matrices. Here $S \in R^{r \times r}$ need not be a diagonal matrix. With this factorization, P2 can be written as

$$\min_{U,S,V} \left\{ \frac{1}{2} \|R_\Omega(USV^T) - b\|_2^2 : U^T U = I_r, V^T V = I_r \right\}$$

By introducing

$$F(U,V) \equiv \min_S \frac{1}{2} \|R_\Omega(USV^T) - b\|_2^2,$$

the formulation is equivalently rewritten as

$$\min_{U,V} \{ F(U,V) : U^T U = I_r, V^T V = I_r \}.$$

Since $F(U,V)$ satisfies

$$F(UQ,V) = F(U,V) = F(U,VQ)$$

for any r by r orthogonal matrix Q, $F(U,V)$ can be considered to be a function defined on the product of two Grassmann manifolds denoted by $M(m,n) = G(m,r) \times G(n,r)$ where the Grassmann manifold $G(n,r)$ denotes the set of all r-dimensional subspaces in R^n. The problem is now written as

$$\min_{U,V} \{ F(U,V) : (U,V) \in M(m,n) \}.$$

Evaluation of $F(U,V)$ for given (U,V) requires solving a least square problem with r^2 variables.

Although the geometry and algorithms on optimization over manifolds have been well studied [15], the global convergence is not guaranteed in general. They were able to derive a performance guarantee that guarantees the global convergence in the following two steps. They first find a good initial point by the operation called *trim*. The *trim* operator replaces the columns and rows with sufficiently many known entries by zeros. As a result, the trimmed matrix is located in a small neighborhood of the global optimum where there is no other stationary point. The convergence to the global optimum is then guaranteed by the gradient descent algorithm over the Grassmann manifold with a line search.

Algorithm 7 Spectral Matrix Completion

1: $Z \leftarrow \mathtt{trim}(R_\Omega^* b)$
2: $Z_r = U_0 S_0 V_0^H \leftarrow \arg\min_X \{ \|Z - X\|_F^2 : \mathrm{rank}(X) \leqslant r \}$
3: $(\widehat{U}, \widehat{V}) \leftarrow \arg\min_{U,V} \{ \widetilde{F}(U,V) : (U,V) \in M(m,n) \}$ starting from (U_0, V_0)
4: $\left(\widehat{S} \leftarrow \arg\min_S \frac{1}{2} \left\| b - R_\Omega(\widehat{U} S \widehat{V}^H) \right\|_2^2 \right)$
5: **return** $\widehat{X} = \widehat{U} \widehat{S} \widehat{V}^H$

V. Comparison of algorithms

The algorithms reviewed in this paper can be compared by using two criteria. The first criterion is the performance guarantee and the other criterion is the computational complexity.

A. Performance Guarantee

There two types of the performance guarantees. The first one is given in terms of the rank-restricted isometry. These performance guarantee is restricted to the applications with linear operator satisfying the rank-restricted isometry. The reconstruction of low-rank matrices from its random ensembles corresponds to this case. Let X_0 denote the matrix to be reconstructed from the linear measurement $b = AX_0$. First, Recht et. al. showed the rank minimization P1 can uniquely determine X_0.

Theorem 5.1: [2] If $\delta_{2r}(A) < 1$, then X_0 is the only matrix such that $rank(X) \leq r$ and $AX = b$.

This theorem also applies to P2. However, the rank minimization is NP-hard and it is desired to have the performance guarantee on the efficient algorithms. The algorithms to solve the nuclear norm minimization (SDP or SVT) and ADMiRA have been proved to have the performance guarantee. For the ideal case where X_0 is exactly low-rank and there is no noise in the measurement, X_0 can be reconstructed in polynomial time under certain conditions on the linear operator. In practice, X_0 may be approximately low-rank and the measurement can be corrupted with an additive noise. For this more general case, the distance between the solution obtained by the algorithms and X_0 can be bounded by the best rank-r approximation error and the strength of the noise with scaling by constants.

Theorem 5.2: [16] If $\delta_{5r}(A) \leq 0.1$, then the approximate solution $\hat{X}$ to P3' satisfies

$$\|X - \hat{X}\|_F \leq C_1 \frac{1}{\sqrt{r}} \|X - X_r\|_* + C_2 \|v\|,$$

where the constants C_1 and C_2 only depend on the rank-restricted isometry constant.

This theorem applies to SDP formulation of P3' and SVT.

Theorem 5.3: [7] Assume that A satisfies $\delta_{4r}(A) \leq 0.04$. Then, after at most $6(r+1)$ iterations, ADMiRA provides a rank-r approximation $\hat{X}$ of X which satisfies

$$\|X - \hat{X}\|_F \leq 20\left(\|X - X_r\|_F + \frac{1}{\sqrt{r}}\|X - X_r\|_* + \|v\|_2\right)$$

where ε is the unrecoverable energy.

The required condition for the performance guarantee of ADMiRA is slightly more demanding compared to that of the nuclear norm minimization. However, the verification of the condition is more difficult than the reconstruction problem itself and there is no known polynomial time algorithm to verify the condition. It has been only shown that some random linear operators satisfy the condition asymptotically with high probability. In very high dimension where the asymptotic works, the difference in the order of the restricted-isometry constant does not make significant difference in the performance. In the mean time, ADMiRA is the only known iterative algorithm such that the number of iteration is bounded by a constant only dependent on r.

The other kind of the performance guarantee is derived for the matrix completion. Candes and Recht [3] showed that the linear operator R_Ω for the matrix completion does not satisfy the rank-restricted isometry. Instead of the rank-restricted isometry, Candes and Recht [3] showed that the matrix completion by P3 can be successful for the low-rank and incoherent matrices. They have shown that some generic models have the incoherence to the linear operator R_Ω.. OptSpace by Keshavan et. al. is another guaranteed algorithm for the matrix completion.

B. Computational Complexity

Applications of the compressed sensing for the

matrix case requires the scalability of the algorithms to manage large scale problems. For the nuclear norm minimization, SVT is an efficient algorithm with the scalability whereas SDP cannot be used for large problems. ADMiRA, Power Factorization, and OptSpace are also efficient algorithms with the scalability. Interestingly, these efficient algorithms share common components that dominate the other computation.

The first component is the singular value decomposition. Indeed, it is not necessary to compute the whole singular value decomposition. Computing the truncated SVD for a few leading singular values suffices. The input matrix for SVD are often sparse depending on the property of the linear operator A For that purpose, the efficient software such as PROPACK [17] has been used for SVT and ADMiRA. Alternative way to compute SVD of large matrices is to use the randomized algorithms such as one proposed in [18].

The other component is the least square problem. ADMiRA, Power Factorization, and OptSpace have the least square problem in their iterations. In particular, for OptSpace, the least square problem is required to evaluate the objective function. There are many efficient algorithms to solve the least square problem including Richardson iteration and conjugate gradient method.

If we solve the least square problem by using the Penrose-Moore pseudo inverse, which also involves the singular value decomposition, the complexity of these two problems can be regarded comparable. The gain of using improved methods seem to be similar for both cases. If it is in the case, the complexity of one iteration of these algorithms (SVT, ADMiRA, FP, OptSpace) can be said all comparable and the overall complexity will be dominated by the number of iterations.

REFERENCES

[1] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, Stanford University, 2002.

[2] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," Arxiv preprint arXiv:0706.4138, 2007.

[3] E. Candes and B. Recht, "Exact Matrix Completion via Convex Optimization," Arxiv preprint arXiv:0805.4471, 2008.

[4] M. Fazel, H. Hindi, and S. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in Proceedings American Control Conference, vol. 6, 2001, pp. 4734 - 4739.

[5] G. Watson, "On matrix approximation problems with Ky Fank norms," Numerical Algorithms, vol. 5, no. 5, pp. 263 - 272, 1993.

[6] J. Cai, E. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," Arxiv preprint arXiv:0810.3286, 2008.

[7] K. Lee and Y. Bresler, "Efficient and guaranteed rank minimization by atomic decomposition," Arxiv preprint arXiv:0901.1898, 2009.

[8] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," IEEE Transactions on Signal Processing, vol. 41, no. 12, pp. 3397 - 3415, 1993.

[9] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in Proc. Asilomar Conf. on Signals, Systems, and Computers, vol. 1, Pacific Grove, CA, Nov. 1993, pp. 40 - 44.

[10] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," Arxiv preprint arXiv:0803.2392, 2008.

[11] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing: closing the gap between performance and complexity," Arxiv preprint arXiv:0803.0811, 2008.

[12] K. Lee and Y. Bresler, "ADMiRA: Atomic Decomposition for Minimum Rank Approximation ," Arxiv preprint arXiv:0905.0044, 2009.

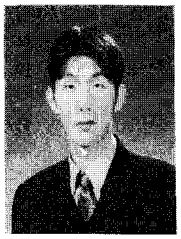
[13] J. Haldar and D. Hernando, "Rank-Constrained Solutions to Linear Matrix Equations using PowerFactorization," will appear in IEEE Signal Processing Letters.

[14] R. Keshavan, S. Oh, and A. Montanari, "Matrix Completion from a Few Entries," Arxiv preprint arXiv:0901.3150, 2009.

[15] A. Edelman, T. Arias, and S. Smith, "The

Geometry of Algorithms with Orthogonality Constraints," SIAM journal on matrix analysis and applications, vol. 20, no. 2, pp. 303 - 353, 1999.
[16] M. Fazel, E. Candes, B. Recht, and P. Parrilo, "Compressed sensing and robust recovery of low rank matrices," in Proc. Asilomar Conf. on Signals, Systems, and Computers, Pacific Grove, CA, Oct. 2008.
[17] R. Larsen, "PROPACK - Software for large and sparse SVD calculations," Available from http://soi.stanford.edu/ rmunk/PROPACK/.
[18] S. Har-Peled, "Low rank matrix approximation in linear time," note, 2006.
[19] L. Mirsky, "Symmetric gauge functions and unitarily invariant norms," The Quarterly Journal of Mathematics, vol. 11, no. 1, pp. 50-59, 1960.

저 자 소 개



이 기 륭(정회원)
2000년 서울대학교 전기공학부 학사 졸업.
2002년 서울대학교 전기컴퓨터 공학부 석사 졸업.
<주관심분야 : 신호처리>

예 종 철(정회원)
1993년 서울대학교 전자공학과 학사 졸업
1995년 서울대학교 전자공학과 석사 졸업
1999년 Purdue Univ. school of electrical and computer engineering 박사 졸업.
2004년 ~ 현재 한국과학기술원 바이오및뇌공학과 부교수
<주관심분야 : Compressed sensing, Bio imaging and signal processing, MRI, CT, Optical imaging, NIRS>