SCISPACE
formerly Typeset

Open access • Proceedings Article • DOI:10.1109/ACSSC.2008.5074571

Compressed sensing and robust recovery of low rank matrices — Source link

Maryam Fazel, Emmanuel J. Candès, Benjamin Recht, Pablo A. Parrilo

Institutions: University of Washington, California Institute of Technology, Massachusetts Institute of Technology

Published on: 01 Oct 2008 - Asilomar Conference on Signals, Systems and Computers

Topics: Restricted isometry property, Matrix (mathematics), Singular value decomposition, Compressed sensing and Robustness (computer science)

Related papers:

- Exact Matrix Completion via Convex Optimization
- Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization
- Compressed sensing
- Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information
- A Singular Value Thresholding Algorithm for Matrix Completion

Share this paper:

View more about this paper here: https://typeset.io/papers/compressed-sensing-and-robust-recovery-of-low-rank-matrices-2kc7h81cvp

Compressed Sensing and Robust Recovery of Low Rank Matrices

M. Fazel*, E. Candès†, B. Recht§, P. Parrilo‡
*Electrical Engineering, University of Washington
†Applied and Computational Mathematics Dept., Caltech
§Center for the Mathematics of Information, Caltech
‡Laboratory for Information and Decision Systems, MIT

***Abstract*— In this paper, we focus on compressed sensing and recovery schemes for low-rank matrices, asking under what conditions a low-rank matrix can be sensed and recovered from incomplete, inaccurate, and noisy observations. We consider three schemes, one based on a certain Restricted Isometry Property and two based on directly sensing the row and column space of the matrix. We study their properties in terms of exact recovery in the ideal case, and robustness issues for approximately low-rank matrices and for noisy measurements.**

Keywords: Matrix rank minimization; compressed sensing; singular value decomposition

I. Introduction

In many data acquisition and reconstruction applications, e.g., signal processing, communications, and coding, the data or signal being acquired is sparse. Recent years have seen a renewed interest and significant research progress in the area of Compressed Sensing of sparse signals or vectors (see, e.g., [3], [5], [4]). A central result in this area shows that a sparse signal $x \in \mathbb{R}^n$ with cardinality k may be recovered from a small number (on the order of $k \log n$) of random linear measurements, if the measurement matrix satisfies a certain restricted isometry property [5].

In this paper, we consider a more general notion of *parsimony* in models and data: the notion of matrix rank, which covers vector cardinality as a special case, and arises in a variety of applications. We focus on compressed sensing and recovery schemes for low-rank matrices, asking under what conditions a low-rank matrix can be sensed and recovered from incomplete, inaccurate, and noisy observations. We consider three such schemes, one based on a condition called the Restricted Isometry Property (RIP) for maps acting on matrices, which is reviewed in section II, and two based on directly sensing the row and column space of the matrix. For each scheme, we study its properties in terms of exact recovery in the ideal case, and robustness issues for general matrices (not necessarily low-rank) and for noisy measurements.

Recent results in [12] on finding the minimum rank matrix satisfying linear equality constraints could be considered as a way to sense and recover a low-rank matrix. We take measurements of the unknown but low-rank matrix $X_0 \in \mathbb{R}^{m\times n}$ using a linear map $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ that satisfies the RIP. Then we recover X_0 by minimizing the nuclear norm $\|X\|_*$ over $y = \mathcal{A}(X)$ which can be done by semidefinite programming. It is shown in [12] that for a perfectly low-rank matrix with no measurement noise, the recovery is exact with very high probability, with the number of measurements on the order of $r(m+n)\log(mn)$. This guarantee is universal and works for any matrix.

In the present paper, we extend this result to the more realistic case where the measurements have additive noise and where X_0 is not required to be low-rank, as shown in section II. We prove that the recovery has excellent robustness properties in both cases. Then, section III focuses on another group of sensing and recovery approaches that are based on sensing the row and column spaces of the matrix by multiplying it by random vectors. We present two algorithms, which are computationally cheaper than the RIP-based approach. We discuss whether these algorithms are robust to the ubiquitous additive noise (for example arising from the sensors or quantization), as well as whether they work for general matrices that are not necessarily low-rank and thus whether they can be used as viable sensing methods. We then discuss the properties and relative advantages and disadvantages of the three approaches, and close the paper by describing some future directions.

A. Notation

Our notation is mostly standard. For $X \in \mathbb{R}^{m\times n}$, $\sigma_i(X)$ denotes the i-th largest singular value of X. The operator norm of a matrix is the largest singular value $\|X\| := \sigma_1(X)$, and the nuclear norm (or Schatten 1-norm) of a matrix is equal to the sum of its singular values $\|X\|_* := \sum_{i=1}^r \sigma_i(X)$, where r is the rank of X. $X^\dagger$ denotes the pseudo-inverse of X given by $V\Sigma^{-1}U^T$, where $X = U\Sigma V^T$ is the singular value decomposition of X in reduced form (i.e., $\Sigma \in \mathbb{R}^{r\times r}$.

II. Sensing and recovery based on the Restricted Isometry Property

Let $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ be a linear map, and let $X_0 \in \mathbb{R}^{m\times n}$ be the matrix to be sensed. Consider a sensing and recovery scheme that consists of two phases:

- **Sensing:** use the sensing map $\mathcal{A}$ to make linear measurements $y = \mathcal{A}(X_0)$
- **Recovery:** Given y (and $\mathcal{A}$), construct X as

$$\hat{X} := \arg\min_X \|X\|_* \qquad \text{subject to} \quad \mathcal{A}(X) = y. \quad (1)$$

In this section, we first review some known results [12] about when X_0 can be recovered under using this scheme, and how the number of measurements p required for exact recovery scales with the dimension and the rank of the matrix. Existing results have considered only the ideal case, where X_0 is perfectly low-rank and the measurements y are free from noise. We then consider the more realistic cases that relax both of these assumptions. We prove that matrix recovery by nuclear norm minimization is remarkably stable and recovers the original matrix with an error that is bounded by a constant factor of the noise level (i.e., the 2-norm of the noise vector). In the case where the matrix is not perfectly low-rank, we prove that we recover the closest rank-r approximation with an error bounded by a constant factor times the error we would have if X was completely known.

The recovery problem can be cast as a semidefinite program [2], and various algorithms can be used for its solution—see [12] and references therein, as well as [11] for a recent fast, large scale algorithm for a certain class.

1) Ideal case: We begin by reviewing key results in [12] indicating when we can a priori guarantee that $X = X_0$. The main conditions are determined by the values of parameters δ_r that quantify the behavior of the linear map $\mathcal{A}$ when restricted to the set of matrices of rank at most r. The following definition is the natural generalization of the Restricted Isometry Property introduced in [5] from vectors to matrices.

Definition 1: Let $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ be a linear map, and $m \le n$. For every r with $1 \le r \le m$, define the r-restricted isometry constant to be the smallest number $\delta_r(\mathcal{A})$ such that

$$(1-\delta_r)\|X\|_F \le \|\mathcal{A}(X)\| \le (1+\delta_r)\|X\|_F \qquad (2)$$

holds for all matrices X of rank at most r.

The following two recovery theorems demonstrate the role of the restricted isometry constants.

Theorem 2: [12] If $\delta_{2r} < 1$ for some integer $r \ge 1$, then X_0 is the only matrix of rank at most r satisfying $\mathcal{A}(X) = b$.

Theorem 3: [12] If $r \ge 1$ is such that $\delta_{5r} < 1/10$, then $X = X_0$.

Thus the RIP provides a sufficient condition for exact recovery. For a linear map $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$, we can write its matrix representation as $\mathcal{A}(X) = \mathbf{A}\,\mathrm{vec}(X)$, where $\mathbf{A}$ is a $p\times mn$ matrix. When such matrices are chosen from a certain class of probability distributions obeying certain tail bounds, called nearly isometric random families, the associated linear maps will obey the Restricted Isometry Property (2) with very high probability, whenever

$$p \ge c_0 r(m+n)\log(mn).$$

See theorem 4.2 in [12] for more details.

2) General X_0 and noisy measurements: In this section, we are concerned with the situation where we have linear measurements of a general matrix, which are contaminated with additive noise. The setup is as follows: we observe

$$y = \mathcal{A}(X_0) + z, \qquad (3)$$

where z is a perturbation obeying $\|z\| \le \varepsilon$. To recover X_0, we propose solving

$$\begin{array}{ll} \text{minimize} & \|X\|_* \\ \text{subject to} & \|y - \mathcal{A}(X)\| \le \varepsilon, \end{array} \qquad (4)$$

which is a convex optimization problem, and can be cast as a semidefinite program [2]. We denote the solution by $\hat{X}$. Here, we do not make any assumption about X_0. In particular, X_0 is not necessarily low-rank, and we will set $X_{0,r}$ to be the matrix of rank r that best approximates X_0 in the nuclear, Frobenius, or operator norm; that is, the truncated SVD of X_0. The following two theorems hold.

Theorem 4: Assume that $\delta_{5r} < 1/10$, then $\hat{X}$ obtained from (4) obeys

$$\|\hat{X} - X_0\|_F \le C_0 \cdot \frac{\|X_0 - X_{0,r}\|_*}{\sqrt{r}} + C_1 \cdot \varepsilon, \qquad (5)$$

where C_0 and C_1 are two (small) constants depending only on the isometry constants.

The proof is given in appendix I. In a nutshell, the error is the sum of two terms: an approximation term, namely, $\|X_0 - X_{0,r}\|_*/\sqrt{r}$ which one would have if we had perfectly noiseless data, and a measurement-error term proportional to the noise level ε. This extends the result of [12] since in the setup studied there we have $\varepsilon = 0$ and $X_0 = X_{0,r}$ so that the right-hand side in (5) vanishes and the reconstruction is exact.

Whereas the previous result involved mixed norms, the next theorem is based only on the nuclear norm.

Theorem 5: Assume that $\delta_{5r} < 1/10$ and noiseless data ($\varepsilon = 0$). Then

$$\|\hat{X} - X_0\|_* \le C_0' \cdot \|X_0 - X_{0,r}\|_* \qquad (6)$$

where C_0' is a small constant depending only on the isometry constants.

In other words the reconstruction is quantitatively as good as if one knew the r-largest left and right singular vectors together with the r largest singular values of the unknown matrix X_0. (One can get better conditions than $\delta_{5r} < 1/10$, see [], but we will not discuss this here). See Appendix II for the proof.

III. Methods based on sensing the row and column space

In this section we consider a different approach to sensing and recovery of low-rank matrices. The idea is to sense the low-dimensional row-space and/or column-space directly, by multiplying the matrix by random vectors. For example, let $X_0 \in \mathbb{R}^{m\times n}$ be of rank r, and pick a set of r random vectors $s_i \in \mathbb{R}^m$ with i.i.d. Gaussian entries. Then Xs_i gives a nonzero vector (with probability one) in the column space or range of X_0, and r of these vectors span the column space (with probability one). Thus we have "sensed" the column space, and can do the same with the row space. Unlike the RIP-based approach in section II, these schemes are not "universal" in the sense that they do not work for *all*

matrices X_0 (e.g., we cannot recover X_0 if some s_i falls in the nullspace of X_0).

The basic idea of random vector multiplication to find low-dimensional approximations to the row or column space has already been successfully used in designing fast randomized algorithms for matrix operations such as the SVD [14], [15]. In that context, the goal is to approximate the SVD of a given matrix with a much lower computational cost than standard SVD algorithms, while bounding the approximation error compared to the exact SVD. Various approaches have been taken in the literature: some algorithms sample rows, columns, submatrices, or entries of the matrix [7], [1], [6], [13], others form random linear combination of rows or columns, or compute orthogonal projections of row or columns onto a lower dimensional space [14], [15], [10].

Our goal in this paper is indeed quite different from approximating matrix SVD, but we find some ideas from this literature useful for our purposes. We propose two schemes for sensing and recovery of low-rank matrices. The first scheme is based on the approach of Woolfe et al, [15], and the second one on the relative-error SVD algorithm of Sarlos [14].

We seek simple sensing schemes to gather or measure a small amount of information about the matrix $X \in \mathbb{R}^{m\times n}$, then store or transmit this information to be used to recover the original matrix. It is important to note that in the recovery phase we do not have access to the original matrix or other information—the matrix should be retrieved solely from the measurements. In this section, we adapt and extend existing results to this new setting, and begin to address the important issue of robustness to measurement noise in addition to approximation error bounds.

A. Approach I: sensing row and column spaces

1) Ideal case: We first consider the case where the matrix $X_0 \in \mathbb{R}^{m\times n}$ to be sensed has rank exactly r, and the product of X_0 and X_0^T with given sensing vectors is measured without noise. We consider the following simple scheme:

- **Sensing:** Make linear measurements of X_0 as
$$Y = SX_0 \quad \text{and} \quad \tilde{Y} = \tilde{S}X_0^T, \tag{7}$$
where $S \in \mathbb{R}^{r\times m}$ and $\tilde{S} \in \mathbb{R}^{r\times n}$ are the sensing matrices.
- **Recovery:** Given Y, $\tilde{Y}$ (and S), construct
$$X = \tilde{Y}^T(S\tilde{Y}^T)^{-1}Y. \tag{8}$$

For the sensing matrices, we pick, for example, matrices with i.i.d. Gaussian entries. Other choices are also possible, such as the so-called SRFT matrices [15] consisting of randomly selected rows of the product of a discrete Fourier transform matrix and a random diagonal matrix. An SRFT construction allows the fast application of S and $\tilde{S}$ to arbitrary vectors. Note that $S\tilde{Y}^T$ is invertible with probability one.

The recovery step is based on the fact that any rank k matrix with row space spanned by the rows of Y and column space spanned by columns of $\tilde{Y}^T$ can be written as $X = \tilde{Y}^T WY$ for some $W \in \mathbb{R}^{r\times r}$. (To implement the recovery efficiently, we could first carry out QR decompositions for $Y^T = Q_1R_1$ and $\tilde{Y}^T = Q_2R_2$, then obtain $W \in \mathbb{R}^{r\times r}$ by solving $RQ_2W = YQ_1$ by some efficient linear equation solver.)

Lemma 6: (Exact recovery) Suppose entries of S are i.i.d. Gaussian. If $\mathrm{rank}(X_0) = r$, the algorithm described by (7) and (8) yields $X = X_0$ with probability one.

Proof: We observe that $SX = SX_0$ or $S(X-X_0) = 0$, which means either $X = X_0$ holds, or all rows of S (which is from the Gaussian or SRFT ensembles, for example) fall in the nullspace of $(X-X_0)^T$ implying that S would belong to a set of measure zero. ■

Note that this scheme requires a total of $r(m+n)$ measurements to recover X_0, which is smaller than the number required by the RIP-based method discussed in the previous section. The recovery is also far simpler compared to solving an SDP for the RIP-based scheme.

2) When X_0 is not low-rank: In this case, we are interested in recovering the best rank-r approximation of the general matrix X_0. To do this, we make a slight (and intuitive) change in the recovery phase of the above scheme. Instead of the whole row or column space of X_0, we need to find only the span of the top k singular vectors, so given measurements Y and $\tilde{Y}$, we form their truncated r-term SVD denoted by $Y_r^T = U_r\Sigma_r V_r^T$ and $\tilde{Y}_r^T = \tilde{U}_r\tilde{\Sigma}_r\tilde{V}_r^T$ and construct $\hat{X}_r$ as
$$\hat{X}_r = \tilde{Y}_r^T(S\tilde{Y}_r^T)^{\dagger}Y_r. \tag{9}$$
The scheme we obtained is in fact equivalent to the SVD algorithm of Woolfe, Liberty, Rokhlin and Tygert [15], where they take the number of sensing vectors to be $l > r$. They show the following approximation error bound that applies to the reconstruction (9).

Theorem 7: [15, sec. 5.2] If $S \in \mathbb{R}^{l\times m}$ and $\tilde{S} \in \mathbb{R}^{l\times n}$ are SRFT matrices, and if there exist $\alpha, \beta > 0$ satisfying
$$\frac{\alpha^2\beta}{(\alpha-1)^2}(2r^2) \le l < m,$$
then the approximation error bound
$$\|\hat{X}_r - X_0\| \le C\sqrt{\max\{m,n\}}\,\sigma_{r+1}(X_0), \tag{10}$$
holds with probability at least $1-3/\beta$. Constant C depends on α (see [15] for precise value).

3) Measurement noise: Consider the case with additive noise in the measurements,
$$Y = SX + \delta Y, \quad \tilde{Y} = \tilde{S}X^T + \delta\tilde{Y}, \tag{11}$$
where the noise terms are assumed to have bounded norm $\|\delta Y\| \le \varepsilon_1$ and $\|\delta\tilde{Y}\| \le \varepsilon_2$. The analysis of the case where both sets of measurements are noisy may be complicated, but we point out that in the special case where one set of measurements is noise-free, we can bound the error $\|\hat{X} - X_0\|$ in terms of noise level and $\sigma_{\min}(S)$. Furthermore, the distribution of $\sigma_{\min}(S)$ is known for various random matrices, for example it is concentrated around $1-\sqrt{l/m}$ for a Gaussian matrix. See [8] for more details and discussion of several special cases and their error bounds.

B. Approach II: random row-space projection

1) Ideal case: In this section we outline a scheme that is based on the approximate SVD algorithm of Sarlos [14]. Suppose $\mathrm{rank}(X_0) = r$, we again perform two sets of measurements of X_0. The difference between this scheme and the previous one is that here the output of the first set is used as the "sensing matrix" for the second set. Thus this method needs to access X_0 and X_0^T to obtain the two sets of measurements sequentially. The second set of measurements are in fact quadratic in X_0. We again have several choices for the sensing matrix $S \in \mathbb{R}^{r\times m}$, for example we can pick S with i.i.d. Gaussian entries. It is also possible to use structured matrices that are faster to apply, for example the SRFT matrix mentioned earlier. We consider the following scheme:

- **Sensing:** Make linear measurement

$$Y = SX_0, \quad \text{followed by} \quad Z = YX_0^T. \tag{12}$$

- **Recovery:** Given measurements Y, Z (and S), construct

$$\hat{X}^T = Y^\dagger Z. \tag{13}$$

The recovery step can be implemented efficiently using a QR decomposition of Y; see [14] for details.

A geometric interpretation is as follows: using $\hat{X}^T = Y^\dagger Y X_0^T = (SX_0)^\dagger (SX_0) X_0^T$ and noting that $Y^\dagger Y$ is the orthogonal projection matrix onto the range of Y, we see that the estimate $\hat{X}$ is given by the projection of each row of X_0 onto the row-space of SX_0, which is spanned by random linear combinations of the rows if X_0. That is, each row of X_0 is approximated by its closest vector in the row-space of SX. Note that the projection can also be expressed as $\hat{X}^T = X_0 U U^T$, where $Y^T = U\Sigma V^T$ is the SVD.

Lemma 8: (Exact recovery) Suppose entries of S are i.i.d. Gaussian. If $\mathrm{rank}(X_0) = r$, the scheme described in (12) and (13) yields $\hat{X} = X_0$ with probability one.

Proof: Let s_i denote the ith row or S (for example, from a Gaussian or SRFT ensemble). If $\mathrm{rank}(X_0) = r$ the set of random vectors $X_0^T s_i$, $i = 1, \ldots, r$ are linearly independent with probability one, which implies that row-space of SX_0 is equal to row-space of X_0 with probability one, and projecting X_0 onto its own row-space gives X_0. ■

2) When X_0 is not low-rank: When $\mathrm{rank}(X_0) > r$, a modification of this scheme gives the algorithm of Sarlos [14], in which the measurements are as above but the estimate is given by $\hat{X}_r = (Y^\dagger Z)_r$, i.e., the truncated r-term SVD of $\hat{X}$. For SRFT matrices, the following bound was shown in [15] as an aside to their proof for the bound in (10).

Lemma 9: [15, sec. 5.2] Suppose S is an SRFT matrix and there are $\alpha, \beta > 1$ such that

$$\frac{\alpha^2\beta}{(\alpha-1)^2}(2r^2) \le l < m.$$

Then,

$$\|\hat{X} - X_0\| = \|X_0 U_r U_r^T - X_0\| \le C\sqrt{m}\sigma_{r+1}(X_0)$$

holds with probability at least $1 - 1/\beta$. Constant C depends on α (see [15] for the precise version of the bound and constant C).

Sarlos in [14, sec. 4] gave the first relative error bound for this estimate (in Frobenius norm). His result implies the bound above.

Note that the probability that the bound holds can be boosted by repeating the measurements with several independent sensing matrices, which is a standard trick for boosting success probability from constant to any desired level in randomized algorithms (see, e.g., [14]).

3) Measurement noise: If $\mathrm{rank}(X_0) = r$, columns of X are the least-norm solutions to the set of under-determined linear equations $YX^T = Z$. Thus, in this case noise analysis boils down to sensitivity analysis of the solution to this set of equations with respect to norm-bounded perturbations in Z and Y. This is a well-studied problem in numerical linear algebra, and we can apply known results from this area (e.g., [9]) to obtain error bounds. We derive and discuss such bounds in [8].

IV. Conclusions and Future Directions

We discussed compressed sensing of matrices and presented three approaches, one based on the Restricted Isometry Property and two based on directly sensing row and column spaces. We showed that sensing a matrix with a linear map satisfying the RIP and recovery via nuclear norm minimization is robust both when the measurements are noisy and when the matrix is not exactly low-rank, with a recovery error that is on the order of the best achievable error. We then presented two approaches based on sensing the low-dimensional row and column spaces, that are easy to carry out and less computationally demanding in sensing and especially recovery compared to the RIP-based approach. We explored their robustness properties. Bounds on approximation error, for the case when the matrix is not perfectly low-rank, already exist for the closely related problem of fast randomized SVD; which we adapted to our problem. We also briefly discussed bounds on the effect of noise.

Current and future directions include detailed noise analysis, bias-variance tradeoff, and characterizing the unrecoverable part of the matrix with respect to the noise level. Numerical study of the schemes and comparisons will also be pursued. In addition, we will explore concrete application areas, for example, wireless channel sensing and feedback, identification of linear operators and dynamical systems via input/output measurements.

Appendix I

Proof of Theorem 4. The proof parallels that of Candès, Romberg and Tao about the recovery of nearly sparse vectors from a limited number of measurements [4]. Set $R = X - X_0$ and observe that by the triangle inequality,

$$\|\mathcal{A}(R)\| \le \|\mathcal{A}(X) - y\| + \|y - \mathcal{A}(X_0)\| \le 2\varepsilon, \tag{14}$$

since X_0 is feasible for the problem (4). Let $X_0 = X_{0,r} + \Delta$. Following lemma 3.4 in [12], we can decompose R as $R = R_0 + R_c$, where

$$\operatorname{rank}(R_0) \le 2\operatorname{rank}(X_{0,r}), \quad X_{0,r}R_c^* = 0, \quad \text{and} \quad X_{0,r}^* R_c = 0.$$

We have

$$\begin{aligned}\|X_0 + R\|_* &\ge \|X_{0,r} + R_c\|_* - \|\Delta\|_* - \|R_0\|_* \\ &= \|X_{0,r}\|_* + \|R_c\|_* - \|\Delta\|_* - \|R_0\|_*,\end{aligned}$$

where the equality $\|X_{0,r}+R_c\|_* = \|X_{0,r}\|_* + \|R_c\|_*$ follows from $X_{0,r}R_c^* = 0$ and $X_{0,r}^* R_c = 0$ (lemma 2.3 in [12]). Since by definition, $\|X_0 + R\|_* \le \|X_0\|_*$, this gives

$$\|R_c\|_* \le \|R_0\|_* + 2\|\Delta\|_*, \tag{15}$$

where we used the fact that $\|X_0\|_* = \|X_{0,r}\|_* + \|\Delta\|_*$.

Next, we use a classical estimate developed in [4] (see also [12]). Decompose R_c into a sum of matrices $R_1, R_2, \ldots$, each of rank at most $3r$. For each i define the index set $I_i = \{3r(i-1)+1, ..., 3ri\}$, and let $R_i := U_{I_i}\operatorname{diag}(\sigma_{I_i})V_{I_i}^*$; that is, R_1 is the part of R_c corresponding to the $3r$ largest singular values, R_2 is the part corresponding to the next $3r$ largest and so on. A now standard computation (see (3.6) in [12]) shows that

$$\sum_{j\ge 2} \|R_j\|_F \le \frac{1}{\sqrt{3r}} \|R_c\|_*, \tag{16}$$

and thus

$$\sum_{j\ge 2} \|R_j\|_F \le \sqrt{\frac{2}{3}} \|R_0\|_F + \frac{2}{\sqrt{3}} \frac{\|\Delta\|_*}{\sqrt{r}},$$

since $\|R_0\|_* \le \sqrt{2r}\,\|R_0\|_F$ by the Cauchy-Schwarz inequality. The triangle inequality and the restricted isometry property now give

$$\begin{aligned}\|\mathcal{A}(R)\| &\ge \|\mathcal{A}(R_0 + R_1)\| - \sum_{j\ge 2} \|\mathcal{A}(R_j)\| \\ &\ge (1-\delta_{5r})\|R_0 + R_1\|_F - (1+\delta_{3r}) \sum_{j\ge 2} \|R_j\|_F.\end{aligned}$$

Let $e_0 = \|\Delta\|_*/\sqrt{r}$. Because R_0 is orthogonal to R_1, we have that $\|R_0 + R_1\|_F \ge \|R_0\|_F$, which gives

$$\begin{aligned}&(1-\delta_{5r})\|R_0 + R_1\|_F \le \\ &\|\mathcal{A}(R)\| + (1+\delta_{3r}) \left[\sqrt{\frac{2}{3}} \|R_0 + R_1\|_F + \frac{2}{\sqrt{3}} e_0\right].\end{aligned}$$

Therefore,

$$\|R_0 + R_1\|_F \le C_0' \, e_0 + C_1' \, \varepsilon, \tag{17}$$

where $C_0' = \frac{1}{\alpha}\frac{2}{\sqrt{3}}(1+\delta_{3r})$, $C_1' = \frac{2}{\alpha}$ and

$$\alpha = (1-\delta_{5r}) - \sqrt{\frac{2}{3}}(1+\delta_{3r}) > 0,$$

where the positivity is guaranteed by the condition of the theorem, namely, $\delta_{5r} < 1/10$. We conclude with

$$\begin{aligned}\|R\|_F &\le \|R_0 + R_1\|_F + \sum_{j\ge 2} \|R_j\|_F \\ &\le \left(1 + \sqrt{\frac{2}{3}}\right) \|R_0 + R_1\|_F + \frac{2}{\sqrt{3}} e_0.\end{aligned}$$

Substituting (17) in this last inequality establishes our claim.

Appendix II

Proof of Theorem 5. The proof of (6) is similar to that of (5) and we only indicate where the argument differs. We start as in the previous section and from (14), we conclude that

$$\|R\|_* \le 2\|R_0\|_* + 2\|\Delta\|_*. \tag{18}$$

We continue as before and obtain $\|R_0\|_F \le C_0' \frac{\|\Delta\|_*}{\sqrt{r}}$ thanks to (17) since $\varepsilon = 0$. Applying Cauchy-Schwarz gives

$$\|R_0\|_* \le \sqrt{2r}\|R_0\|_F \le C_0'\sqrt{2}\,\|\Delta\|_*, \tag{19}$$

and, therefore, it follows from (18) and (19) that

$$\|R\|_* \le 2(\sqrt{2}C_0' + 1)\,\|\Delta\|_*,$$

which is what we needed to establish.

References

[1] D. Achlioptas and F. McSherry. Fast computation of low-rank approximations. *Journal of the ACM*.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2003.

[3] E. J. Candès. Compressive sampling. In *International Congress of Mathematicians. Vol. III*, pages 1433–1452. Eur. Math. Soc., Zürich, 2006.

[4] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, Aug. 2006.

[5] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.

[6] A. Deshpande and S. Vempala. Adaptive sampling and fast low-rank matrix approximation. In *Proc. Workshop on Randomization and Computation (RANDOM)*, 2006.

[7] P. Drineas, M. W. Mahoney, and R. Kannan. Fast Monte Carlo algorithms for matrices II: Computing a low rank approximation to a matrix. *SIAM Journal on computing*, 36:158–183, 2006.

[8] M. Fazel, E. Candès, B. Recht, and P. A. Parrilo. Compressed sensing of low-rank matrices. In preparation.

[9] G. H. Golub and C. F. V. Loan. *Matrix Computations*. Johns Hopkins University Press, 1996.

[10] S. Har-Peled. Low rank matrix approximation in linear time. Manuscript, 2006.

[11] E. J. C. J-F Cai and Z. Shen. A singular value thresholding algorithm for matrix completion. Submitted for publication, 2008.

[12] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization. 2007. Submitted to *SIAM Review*.

[13] M. Rudelson and R. Vershynin. Sampling from large matrices: an approach through geometric functional analysis. *Journal of the ACM*.

[14] T. Sarlos. Improved approximation algorithms for large matrices via random projections. In *Proc. FOCS 2006, Symposium on Foundations of Computer Science*, pages 143–152, Oct. 2006.

[15] F. Woolfe, E. Liberty, V. Rokhlin, and M. Tygert. A fast randomized algorithm for the approximation of matrices. *Appl. Comput. Harmon. Anal.*, 25:335–366, 2008.