Appl. Comput. Harmon. Anal. 33 (2012) 309–314

Contents lists available at SciVerse ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Letter to the Editor

Uniqueness conditions for low-rank matrix recovery

Y.C. Eldar [a], D. Needell [b,*], Y. Plan [c]

[a] Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa, Israel
[b] Department of Mathematics and Computer Science, Claremont McKenna College, 850 Columbia Ave, Claremont, CA, United States
[c] Department of Mathematics, 530 Church Street, Ann Arbor, MI, United States

ARTICLE INFO

Article history:
Received 23 July 2011
Accepted 12 April 2012
Available online 17 April 2012
Communicated by Charles K. Chui

Keywords:
Rank minimization
Nuclear-norm minimization
Low-rank matrix recovery
Random matrices
Compressed sensing

ABSTRACT

Low-rank matrix recovery addresses the problem of recovering an unknown low-rank matrix from few linear measurements. There has been a large influx of literature deriving conditions under which certain tractable methods will succeed in recovery, demonstrating that $m \geqslant Cnr$ Gaussian measurements are often sufficient to recover any rank-r $n \times n$ matrix. In this paper we address the theoretical question of how many measurements are needed via any method whatsoever — tractable or not. We show that for a family of random measurement ensembles, $m \geqslant 4nr - 4r^2$ and $m \geqslant 2nr - r^2 + 1$ measurements are sufficient to guarantee strong recovery and weak recovery, respectively, by rank minimization. These results give a benchmark to which we may compare the efficacy of tractable methods such as nuclear-norm minimization.

1. Introduction

Low-rank matrix recovery is a burgeoning topic with various applications including collaborative filtering, machine learning, control theory, quantum state tomography, and triangulation from incomplete distance measurements. Many recovery programs have been proposed, notably nuclear-norm minimization, with impressive performance in theory and practice (see e.g. [1,3,4,6,11,8,10,17,7,2]). For sufficiently random measurements (e.g., Gaussian), a generic result states that Cnr linear measurements are sufficient to recover any rank-r matrix; this holds for a few different methods of recovery (see e.g., [8,2]).

However, in order to precisely ascertain the strength of these different methods, it is important to compare them to a fundamental benchmark. In this paper, we provide this benchmark by analyzing the conditions under which the original low-rank matrix has the lowest rank over all matrices that fit the data.

1.1. Problem setup

We would like to recover a low-rank matrix M from few of its linear measurements, $\mathcal{A}(M) \in \mathbb{R}^m$. The measurement operator is of the form $\mathcal{A} : \mathbb{R}^{n\times n} \to \mathbb{R}^m$ and acts on a matrix M by $(\mathcal{A}(M))_i = \langle A_i, M\rangle$ where A_i are $n \times n$ matrices and $\langle\cdot,\cdot\rangle$ denotes the usual matrix inner product:

$$\langle A, B\rangle \stackrel{\text{def}}{=} \operatorname{trace}\left(A^* B\right).$$

* Corresponding author.
E-mail address: dneedell@cmc.edu (D. Needell).

1063-5203/$ – see front matter
http://dx.doi.org/10.1016/j.acha.2012.04.002

We take our estimate of X to be the matrix which has minimum rank and satisfies the data constraints,

$$\hat{X} = \arg\min_X \operatorname{rank}(X) \quad \text{such that} \quad \mathcal{A}(X) = \mathcal{A}(M). \tag{1.1}$$

This is simply a uniqueness problem; when is M the unique low-rank matrix having these measurements? However, the problem (1.1) is intractable in general.

For reference, we also describe the standard tractable relaxation of this approach, called nuclear-norm minimization. We thus consider the minimization problem

$$\hat{X} = \arg\min_X \|X\|_* \quad \text{such that} \quad \mathcal{A}(X) = \mathcal{A}(M), \tag{1.2}$$

where $\|\cdot\|_*$ denotes the *nuclear norm* which is defined by

$$\|X\|_* = \operatorname{trace}\bigl(\sqrt{X^*X}\bigr) = \sum_{i=1}^{n} \sigma_i.$$

The program (1.2) can be cast as a semidefinite program (SDP) and is therefore numerically feasible.

A question that does not appear to have been previously addressed is, how many measurements suffice to recover rank-r matrices via the method (1.1)? This question has remained unresolved when allowing the matrix entries to have unbounded real or complex entries. (In the setting of finite fields, the question has been well addressed, see [18].) Answering this question would not only fill a gap in the literature but also give theoretical bounds on the number of measurements required for low-rank matrix recovery against which those for problem (1.2) may be compared.

2. Uniqueness results

Next we summarize our main results.

Since $M - M'$ is at most rank-$2r$ when M and M' are each rank-r, to guarantee that (1.1) reconstructs all rank-r matrices, a necessary and sufficient condition is that there are no rank-$2r$ (or less) matrices in the null space of $\mathcal{A}$. Thus we examine the following subset of $\mathbb{R}^{n\times n}$:

$$\mathcal{R}' = \bigl\{X \in \mathbb{R}^{n\times n}\colon \operatorname{rank}(X) = 2r\bigr\}. \tag{2.1}$$

It is well known that $\mathcal{R}'$ is a manifold with $4nr - 4r^2$ dimensions. Is $m \geqslant 4nr - 4r^2$ sufficient to guarantee uniform recovery? We will show that the answer is yes! This is summarized by the following theorem.

Below, we call $\mathcal{A}$ a Gaussian operator if each A_i is independent with i.i.d. Gaussian entries.

Theorem 2.1 *(Strong recovery). Let $r \leqslant n/2$. When $\mathcal{A}: \mathbb{R}^{n\times n} \to \mathbb{R}^m$ is a Gaussian operator with $m \geqslant 4nr - 4r^2$, problem* (1.1) *recovers all rank-r matrices with probability* 1.

Remarks. 1. We prove this theorem via a more general result in Theorem 3.1 which holds for any continuously differentiable manifold over the set of real matrices. Theorem 2.1 will follow as a consequence.

2. We consider real-valued matrices but our method can easily be extended to complex-valued matrices as well.

3. To be clear, strong recovery (or universal recovery) requires that one randomly picked measurement ensemble can be successfully used to recover every matrix with rank less than or equal to r.

Our proof technique also allows us to provide a bound on the number of measurements required for weak recovery. Recall that in this framework we are interested in recovering *one fixed matrix M* with high probability. Since M is fixed, we require only that for all rank-r matrices $X \neq M$ that $X - M$ is not in the null space of $\mathcal{A}$. The set of all rank-r matrices is a manifold of dimension $2nr - r^2$. The following result shows that for weak recovery of low-rank matrices we require a number of measurements at least one more than the dimension of the manifold of all rank-r matrices.

Theorem 2.2 *(Weak recovery). Fix a rank-r $n \times n$ real matrix M. When $\mathcal{A}: \mathbb{R}^{n\times n} \to \mathbb{R}^m$ is a Gaussian operator with $m \geqslant 2nr - r^2 + 1$, problem* (1.1) *recovers the matrix M with probability* 1.

As we will see in Section 4, this theorem allows the comparison of rank minimization to the theoretical and empirical results of nuclear-norm minimization in the Gaussian setting.

We prove these results in the next section. In Section 4 we discuss the tightness of these bounds and compare them with results for nuclear-norm minimization.

3. General results and proofs

To prove our main results, Theorems 2.1 and 2.2, we utilize a more general result about arbitrary manifolds of real matrices. For convenience we will restrict ourselves to the Banach space of real matrices. Below, a continuously differentiable manifold is a manifold that may be equipped with a class of atlases having transition maps which are all C^1-diffeomorphisms.

Theorem 3.1. *Let $\mathcal{R}$ be a d-dimensional continuously differentiable manifold over the set of $n \times n$ real matrices. Suppose we take $m \geqslant d+1$ measurements of the form $\langle A_i, X\rangle$ for $X \in \mathcal{R}$, and define the operator $\mathcal{A} : \mathcal{R} \to \mathbb{R}^m$ which takes these measurements, $\mathcal{A} : X \mapsto y$ with $y_i = \langle A_i, X\rangle$. Assume that there exists a constant $C = C(n)$ such that $\mathbb{P}(|\langle A_i, X\rangle| < \varepsilon) < C\varepsilon$ for every X with $\|X\|_F = 1$. Further assume that for each $X \neq 0$ that the random variables $\{\langle A_i, X\rangle\}$ are independent. Then with probability* 1,

$$\mathrm{Null}(\mathcal{A}) \cap \mathcal{R}\backslash\{0\} = \emptyset.$$

Remarks. 1. The requirement that $\mathbb{P}(|\langle A_i, X\rangle| < \varepsilon) < C\varepsilon$ says that the densities of $\langle A_i, X\rangle$ do not spike at the origin. A sufficient condition for this to hold for every X with $\|X\|_F = 1$ is that each A_i has i.i.d. entries with continuous density.

2. The requirement $m \geqslant d+1$ is tight in the sense that the result does not generally hold for $m \leqslant d$. For example, take $\mathcal{R}$ to be the intersection of any $(d+1)$-dimensional linear subspace of $\mathbb{R}^{n\times n}$ with the unit sphere. Then it is not hard to show that $\mathrm{Null}(A) \cap \mathcal{R}\backslash\{0\} \neq \emptyset$ for any linear operator $\mathcal{A} : \mathbb{R}^{n\times n} \to \mathbb{R}^m$ as long as $m \leqslant d$.

Theorem 3.1 can be proved using tools from differential geometry. However, the proof we provide here uses more elementary tools from probability, and we believe it will be more accessible to a wider audience. The strategy of the proof is to cover $\mathcal{R}\backslash\{0\}$ by a collection of small neighborhoods, each of which is 'approximately linear' since $\mathcal{R}\backslash\{0\}$ is a differentiable manifold. Then we show that $\mathrm{Null}(A)$ does not intersect any of these neighborhoods. The notion of approximate linearity is made concrete by specifying a bound on the number of points sufficient to cover each of these neighborhoods to high precision.

The first tool we will use in the proof is a well-known fact about covering numbers. For a set B, norm $\|\cdot\|$ and value ε, we denote by $N(B, \|\cdot\|, \varepsilon)$ the smallest number of balls (with respect to the norm $\|\cdot\|$) of radius ε whose union contains B. This number is called a *covering number*, and the set of balls covering the space (or more precisely the center of these balls) is called an *ε-net*. A bound on the covering number for the unit ball under the Euclidean norm $\|\cdot\|_2$ is well known (see e.g. Chapter 13 of [9]):

Lemma 3.2. *For any $1 > \varepsilon > 0$, we have*

$$N\big(B_2^d, \|\cdot\|_2, \varepsilon\big) \leqslant \left(\frac{3}{\varepsilon}\right)^d.$$

We are now prepared to prove Theorem 3.1.

Proof of Theorem 3.1. It suffices to prove the claim for $m = d+1$. Since $\mathcal{R}$ (and thus $\mathcal{R}\backslash\{0\}$) is a continuously differentiable manifold, there are a countable number of closed[1] sets $\mathcal{V}_i \subset \mathcal{R}\backslash\{0\}$ such that

- $\bigcup \mathcal{V}_i = \mathcal{R}\backslash\{0\}$.
- For each $\mathcal{V}_i$, there exists a C^1-diffeomorphism $\phi_i : \mathcal{V}_i \to B_2^d$, where B_2^d denotes the unit Euclidean ball in $\mathbb{R}^d$. In words, ϕ_i is a homeomorphism from $\mathcal{V}_i$ to the unit ball such that ϕ_i and ϕ_i^{-1} are continuously differentiable.

Our goal will be to show that $0 \notin \mathcal{A}(\mathcal{V}_i)$ with probability 1 for each i. Since there are only countably many i, this will show that $0 \notin \mathcal{R}\backslash\{0\}$ via a union bound.

Fix i, and for convenience set $\phi = \phi_i$ and $\mathcal{V} = \mathcal{V}_i$. Since ϕ^{-1} is continuously differentiable, it is Lipschitz on the closed set B_2^d. Thus there is an $L > 0$ such that

$$\big\|\phi^{-1}(x) - \phi^{-1}(y)\big\|_F \leqslant L\|x - y\|_2. \tag{3.1}$$

By Lemma 3.2, there is an (ε/L)-net for B_2^d with cardinality at most $(\frac{3L}{\varepsilon})^d$. Denote this net by $\overline{B_2^d}$. Then the net $\overline{\mathcal{V}}$ defined by $\overline{\mathcal{V}} = \phi^{-1}(\overline{B_2^d})$ is an ε-net for $\mathcal{V}$. Indeed, for any $X \in \mathcal{V}$, we have $\phi(X) \in B_2^d$ and so there is a $\bar{b} \in \overline{B_2^d}$ such that

[1] Note that in general these sets are open, but by writing each (open) $\mathcal{V}_i$ as a countable union of closed sets (for example $\mathcal{V}_i = \bigcup_{j=1\ldots\infty} \phi^{-1}(\{x\colon \|x\|_2 \leqslant 1 - 1/j\})$) we observe that we can choose them to be closed.

$$\left\|\bar{b}-\phi(X)\right\|_2 \leqslant \frac{\varepsilon}{L}.$$

By (3.1) we then have

$$\left\|\phi^{-1}(\bar{b})-X\right\|_F \leqslant L\cdot\left\|\bar{b}-\phi(X)\right\|_2 \leqslant L\cdot\frac{\varepsilon}{L}=\varepsilon.$$

Since $\phi^{-1}(\bar{b})\in\phi^{-1}(\overline{B_2^d})$, this shows that $\overline{\mathcal{V}}$ is an ε-net for $\mathcal{V}$.

Therefore, for any $X\in\mathcal{V}$, there is an $\overline{X}\in\overline{\mathcal{V}}$ such that $\|X-\overline{X}\|_F\leqslant\varepsilon$. This then implies

$$\begin{aligned}\left\|\mathcal{A}(X)\right\|_\infty &\geqslant \left\|\mathcal{A}(\overline{X})\right\|_\infty - \left\|\mathcal{A}(X-\overline{X})\right\|_\infty\\ &\geqslant \left\|\mathcal{A}(\overline{X})\right\|_\infty - \|\mathcal{A}\|_{F\to\infty}\|X-\overline{X}\|_F\\ &\geqslant \left\|\mathcal{A}(\overline{X})\right\|_\infty - \varepsilon\cdot\|\mathcal{A}\|_{F\to\infty},\end{aligned}$$

where $\|\cdot\|_{F\to\infty}$ denotes the operator norm from the Frobenius norm to the supremum norm, $\|\cdot\|_\infty$. Optimizing over all $X\in\mathcal{V}$ and $\overline{X}\in\overline{\mathcal{V}}$ yields

$$\inf_{X\in\mathcal{V}}\left\|\mathcal{A}(X)\right\|_\infty \geqslant \min_{\overline{X}\in\overline{\mathcal{V}}}\left\|\mathcal{A}(\overline{X})\right\|_\infty - \varepsilon\cdot\|\mathcal{A}\|_{F\to\infty}.$$

We can then bound the probability (over the random choice of $\mathcal{A}$) by:

$$\begin{aligned}\mathbb{P}\Big(\inf_{X\in\mathcal{V}}\left\|\mathcal{A}(X)\right\|_\infty=0\Big) &\leqslant \mathbb{P}\Big(\inf_{X\in\mathcal{V}}\left\|\mathcal{A}(X)\right\|_\infty\leqslant\varepsilon\log(1/\varepsilon)\Big)\\ &\leqslant \mathbb{P}\Big(\min_{\overline{X}\in\overline{\mathcal{V}}}\left\|\mathcal{A}(\overline{X})\right\|_\infty - \varepsilon\cdot\|\mathcal{A}\|_{F\to\infty}\leqslant\varepsilon\log(1/\varepsilon)\Big).\end{aligned}$$

Conditioning on whether $\|\mathcal{A}\|_{F\to\infty}>\log(1/\varepsilon)$ and using the law of total probability yields

$$\begin{aligned}&\mathbb{P}\Big(\min_{\overline{X}\in\overline{\mathcal{V}}}\left\|\mathcal{A}(\overline{X})\right\|_\infty - \varepsilon\cdot\|\mathcal{A}\|_{F\to\infty}\leqslant\varepsilon\log(1/\varepsilon)\Big)\\ &\quad\leqslant \mathbb{P}\Big(\min_{\overline{X}\in\overline{\mathcal{V}}}\left\|\mathcal{A}(\overline{X})\right\|_\infty\leqslant 2\varepsilon\log(1/\varepsilon)\Big)+\mathbb{P}\big(\|\mathcal{A}\|_{F\to\infty}>\log(1/\varepsilon)\big).\end{aligned}\tag{3.2}$$

Clearly, for ε small, the second term in the last line of (3.2) is negligible. Thus it remains to bound the first term. Letting $z_1,\dots,z_m$ be the coordinates of $\mathcal{A}(\overline{X})$ for a given $\overline{X}\in\overline{\mathcal{V}}$, we have:

$$\begin{aligned}\mathbb{P}\Big(\min_{\overline{X}\in\overline{\mathcal{V}}}\left\|\mathcal{A}(\overline{X})\right\|_\infty\leqslant 2\varepsilon\log(1/\varepsilon)\Big) &\leqslant |\overline{\mathcal{V}}|\cdot\mathbb{P}\big(\left\|\mathcal{A}(\overline{X})\right\|_\infty\leqslant 2\varepsilon\log(1/\varepsilon)\big)\\ &= |\overline{\mathcal{V}}|\cdot\mathbb{P}\big(\max\{|z_1|,\dots,|z_m|\}\leqslant 2\varepsilon\log(1/\varepsilon)\big)\\ &\leqslant \left(\frac{3L}{\varepsilon}\right)^d\cdot\prod_{i=1}^m\big(\mathbb{P}\big(|z_i|\leqslant 2\varepsilon\log(1/\varepsilon)\big)\big),\end{aligned}$$

where in the last line we have utilized the independence of all $z_i=\langle A_i,\overline{X}\rangle$ and the size of the net $\mathcal{V}$.

Now,

$$\begin{aligned}\mathbb{P}\big(|z_i|\leqslant 2\varepsilon\log(1/\varepsilon)\big) &= \mathbb{P}\big(\big|\langle A_i,X\rangle\big|\leqslant 2\varepsilon\log(1/\varepsilon)\big)\\ &= \mathbb{P}\left(\left|\left\langle A_i,\frac{X}{\|X\|_F}\right\rangle\right|\leqslant\frac{2\varepsilon\log(1/\varepsilon)}{\|X\|_F}\right).\end{aligned}$$

Since $\mathcal{V}$ is closed and does not contain zero, the Frobenius norm of any $X\in\mathcal{V}$ is bounded uniformly away from zero. This combined with the assumption that $\mathbb{P}(|\langle A_i,X\rangle|<\varepsilon)<C\varepsilon$ for every X with $\|X\|_F=1$ yields

$$\begin{aligned}\left(\frac{3L}{\varepsilon}\right)^d\cdot\prod_{i=1}^m\big(\mathbb{P}\big(|z_i|\leqslant 2\varepsilon\log(1/\varepsilon)\big)\big) &\leqslant \left(\frac{3L}{\varepsilon}\right)^d\cdot\big(4C'\varepsilon\log(1/\varepsilon)\big)^m\\ &= C''\varepsilon^{m-d}\cdot\big(\log(1/\varepsilon)\big)^m\\ &= C''\varepsilon\cdot\big(\log(1/\varepsilon)\big)^m,\end{aligned}$$

where C, C' and C'' are constants which do not depend on ε. The last equality follows since $m=d+1$. Taking ε to zero once again makes this last term vanish. Thus the probability that the null space of $\mathcal{A}$ intersects $\mathcal{V}$ is zero. Since there are only countably many $\mathcal{V}_i$, the probability that the null space of $\mathcal{A}$ intersects any of these sets is also zero. □

Next we show that Theorems 2.1 and 2.2 follow as corollaries to Theorem 3.1 via the following lemma.

Lemma 3.3. *The space of rank-r matrices with fixed Frobenius norm,*

$$\mathcal{R} = \left\{X \in \mathbb{R}^{n\times n}\colon \operatorname{rank}(X) = 2r, \|X\|_F = 1\right\},$$

is a smooth manifold with dimension $4nr - 4r^2 - 1$.

The proof of Lemma 3.3 is a standard exercise in differential geometry so we now turn to the proof of Theorem 2.1 and Theorem 2.2 via the general result, Theorem 3.1.

Proof of Theorem 2.1. By Lemma 3.3, $\mathcal{R}$ is a smooth manifold of dimension $d = 4nr - 4r^2 - 1$ and note that clearly $\mathcal{R} = \mathcal{R}\backslash\{0\}$. Let $\mathcal{A}$ be the operator taking $m \geqslant 4nr - 4r^2$ Gaussian measurements $\langle A_i, X\rangle$ for $X \in \mathcal{R}$ and A_i (for $i = 1, 2, \ldots m$) having i.i.d. Gaussian entries. Then all $\langle A_i, X\rangle$ are independent and have (the same) continuous density. Therefore by Theorem 3.1, $\operatorname{Null}(\mathcal{A}) \cap \mathcal{R} = \emptyset$. Applying Theorem 3.1 for all ranks between 1 and $2r$, we see that there is no matrix of rank $2r$ or less in the null space of $\mathcal{A}$. Thus when M has rank r (or less) there can be no other matrix X with $\mathcal{A}(X) = \mathcal{A}(M)$ having the same or lower rank. This proves that (1.1) must recover the matrix M and completes the proof. □

Proof of Theorem 2.2. Let $\mathcal{W} = \{X - M\colon \operatorname{rank}(X) = r\}$. Note the proof of Lemma 3.3 explicitly shows that the space of all matrices of a fixed rank r is a smooth manifold of dimension $2nr - r^2$. Since $\mathcal{W}$ is a shift of this space, it is also a smooth manifold of the same dimension. Then by Theorem 3.1, we have that with probability one

$$\mathcal{W}\backslash\{0\} \cap \operatorname{Null}(\mathcal{A}) = \emptyset.$$

Repeating this for ranks 1 through r, we get that with probability one

$$\mathcal{W}'\backslash\{0\} \cap \operatorname{Null}(\mathcal{A}) = \emptyset \tag{3.3}$$

where $\mathcal{W}' = \{X - M\colon \operatorname{rank}(X) \leqslant r\}$. Now let X be the solution of the rank minimization problem (1.1). Since M has rank r and is a feasible matrix, $\operatorname{rank}(X) \leqslant r$ as well. Thus $X - M \in \mathcal{W}'$. But since $\mathcal{A}(X) = \mathcal{A}(M)$, $X - M \in \operatorname{Null}(A)$. Thus by (3.3) it must be that $X - M = 0$ which shows $X = M$ is the recovered matrix. □

4. Discussion

In this paper we discuss the number of measurements required to recover rank-r matrices via rank minimization. This is useful both from a purely theoretical point of a view and also provides a benchmark with which to compare tractable methods such as nuclear-norm minimization. Cnr measurements suffice to (provably) recover $n \times n$ rank-r matrices using nuclear-norm minimization [12,17,13,2]. In [13] explicit formulas and graphs are given from which bounds on the constant C can be derived. Even more recent results in [14] prove that $6nr$ measurements suffice for weak recovery and $16nr$ measurements suffice for strong recovery. New work in [15,16] also shows weak recovery when $m \geqslant 6nr - 3r^2$. In addition, numerical results indicate that weak recovery requires about $4nr - 2r^2$ Gaussian measurements [13, Fig. 1]. Thus according to these results, rank minimization does succeed with somewhat fewer measurements. We emphasize that this should not be a surprise — nuclear-norm minimization is a tractable method whereas rank minimization is an intractable method whose guarantees give us theoretical bounds with which to compare. In fact, the price to pay for a tractable method in low-rank matrix recovery seems to be a very reasonable one.

As discussed above, our general manifold result, Theorem 3.1, is tight. However, this does not imply that its consequences, Theorems 2.1 and 2.2, are tight since the set of matrices of fixed rank is not a linear subspace. A simple lower bound on the required number of measurements has been obtained [2] by observing that the set of rank-$2r$ matrices have (many) linear subsets of dimension $2nr$ (e.g., consider the set of matrices whose last $n - 2r$ rows contain zeros); thus $2nr$ measurements are necessary. However, this is a considerable restriction on the underlying manifold, suggesting that it leads to a loose lower bound. We conjecture that the strong recovery requirement, $m \geqslant 4nr - 4r^2$ from Theorem 2.1, is tight because the number of measurements required matches the dimension of the underlying manifold. In the case of the weak recovery requirement $m \geqslant 2nr - r^2 + 1$ given by Theorem 2.2, we require m to be one greater than the dimension of the underlying manifold. However, we once again conjecture this to be tight at least within an additive factor of one for the same reason. Furthermore, this requirement matches that lower bound derived in [18] for matrices whose elements are contained in finite fields.

Finally, we point to the analogous theory in compressed sensing — the problem of recovering a *sparse* vector from linear measurements. Here, an *s-sparse* vector $x \in \mathbb{R}^d$ is one with at most s non-zero entries. In this setting, one needs $2s$ and $s + 1$ measurements for strong and weak recovery, respectively, and this recovery is via an intractable method (see e.g. [5, Theorem 1.1]). The bounds on the number of measurements given by Theorems 2.1 and 2.2 of $4nr - 4r^2$ and $2nr - r^2 + 1$ are analogous to the bounds of $2s$ and $s + 1$ in compressed sensing. In the compressed sensing setting, the question about theoretical requirements was easy to answer because the set of s-sparse vectors is the union of a finite number of linear subspaces. In the matrix recovery problem, however, this question had remained unresolved.

Our results in conjunction with work on nuclear-norm minimization show how close nuclear-norm minimization guarantees are to those of the intractable problem of rank minimization. While rank minimization requires fewer measurements, it is not at all an unreasonable amount to pay in order to solve the problem via a computationally feasible method.

References

[1] J.F. Cai, E.J. Candès, Z. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optim. 20 (4) (2008) 1956–1982.
[2] E.J. Candès, Y. Plan, Tight oracle bounds for low-rank matrix recovery from a minimal number of random measurements, IEEE Trans. Inform. Theory 57 (4) (2010) 2342–2359.
[3] E.J. Candès, Y. Plan, Matrix completion with noise, Proc. IEEE 98 (6) (2010) 925–936.
[4] E.J. Candès, B. Recht, Exact matrix completion via convex optimization, Found. Comput. Math. 9 (6) (2009) 717–772.
[5] E.J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete Fourier information, IEEE Trans. Inform. Theory 52 (2) (2006) 489–509.
[6] W. Dai, O. Milenkovic, SET: an algorithm for consistent matrix completion, in: ICASSP 2010 IEEE Int. Conf., IEEE, 2010, pp. 3646–3649.
[7] D. Gross, Recovering low-rank matrices from few coefficients in any basis, IEEE Trans. Inform. Theory 57 (3) (2011) 1548–1566.
[8] K. Lee, Y. Bresler, Admira: Atomic decomposition for minimum rank approximation, IEEE Trans. Inform. Theory 56 (9) (2010) 4402–4416.
[9] G.G. Lorentz, M. von Golitschek, Y. Makovoz, Constructive Approximation: Advanced Problems, Springer, Berlin, 1996.
[10] S. Ma, D. Goldfarb, L. Chen, Fixed point and Bregman iterative methods for matrix rank minimization, Math. Program. (2009) 1–33.
[11] R. Meka, P. Jain, I.S. Dhillon, Guaranteed rank minimization via singular value projection, in: Proc. Neural Information Processing Systems (NIPS), December 2010, pp. 937–945.
[12] S. Negahban, P. Ravikumar, M.J. Wainwright, B. Yu, A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers, in: Advances in Neural Information Processing Systems (NIPS), 2009.
[13] S. Oymak, B. Hassibi, New null space results and recovery thresholds for matrix rank minimization, available at http://arxiv.org/abs/1011.6326, 2010, submitted for publication.
[14] S. Oymak, B. Hassibi, Tight recovery thresholds and robustness analysis for matrix rank minimization, submitted for publication.
[15] P.A. Parrilo, A.S. Willsky, V. Chandrasekaran, B. Recht, The convex geometry of linear inverse problems, available at http://arxiv.org/abs/1012.0621, 2010.
[16] B. Recht, personal communication, March 2011.
[17] B. Recht, M. Fazel, P.A. Parrilo, Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization, SIAM Rev. 52 (3) (2010) 471–501.
[18] V.Y.F. Tan, L. Balzano, S.C. Draper, Rank minimization over finite fields: Fundamental limits and coding-theoretic interpretations, IEEE Trans. Inform. Theory 58 (4) (2012) 2018–2039.