Claremont Colleges
Scholarship @ Claremont

CMC Faculty Publications and Research

CMC Faculty Scholarship

3-28-2011

Uniqueness Conditions For Low-Rank Matrix Recovery

Yonina C. Eldar
Israel Institute of Technology

Deanna Needell
Claremont McKenna College

Yaniv Plan
Israel Institute of Technology

Recommended Citation

UNICITY CONDITIONS FOR LOW-RANK MATRIX RECOVERY

Y. C. ELDAR, D. NEEDELL, AND Y. PLAN

ABSTRACT. Low-rank matrix recovery addresses the problem of recovering an unknown low-rank matrix from few linear measurements. Nuclear-norm minimization is a tractible approach with a recent surge of strong theoretical backing. Analagous to the theory of compressed sensing, these results have required random measurements. For example, $m \geq Cnr$ Gaussian measurements are sufficient to recover any rank-r $n \times n$ matrix with high probability. In this paper we address the theoretical question of how many measurements are needed via any method whatsoever — tractible or not. We show that for a family of random measurement ensembles, $m \geq 4nr - 4r^2$ measurements are sufficient to guarantee that no rank-$2r$ matrix lies in the null space of the measurement operator with probability one. This is a necessary and sufficient condition to ensure uniform recovery of all rank-r matrices by rank minimization. Furthermore, this value of m precisely matches the dimension of the manifold of all rank-$2r$ matrices. We also prove that for a fixed rank-r matrix, $m \geq 2nr - r^2 + 1$ random measurements are enough to guarantee recovery using rank minimization. These results give a benchmark to which we may compare the efficacy of nuclear-norm minimization.

1. INTRODUCTION

In the compressed sensing problem, one wishes to recover an unknown vector $x \in \mathbb{R}^n$ from few linear measurements of the form $y = Ax \in \mathbb{R}^m$ where A is an $m \times n$ measurement matrix and $m \ll n$ (see e.g. [8, 1, 9] for tutorials on compressed sensing). This problem is clearly ill-posed until additional assumptions are enforced. A common assumption is that x is *s-sparse*: the support of x is small, $\|x\|_0 = |\operatorname{supp}(x)| \leq s \ll n$. If A is injective on all s-sparse vectors then when x is s-sparse, x will be the solution to

$$(L_0) \qquad \hat{x} = \operatorname*{argmin}_w \|w\|_0 \quad \text{such that} \quad Aw = y.$$

Moreover, for a matrix A to be injective on all s-sparse vectors, we precisely require that its null space be disjoint from the set of all $2s$-sparse vectors. Since there are many classes of matrices satisfying this property with $m = 2s$ rows (see e.g. [6, Theorem 1.1]), this shows that only $2s$ measurements are required to recover all s-sparse vectors $x \in \mathbb{R}^d$! If we consider the problem of weak recovery, where we only wish to recover one *fixed* vector x, $s + 1$ measurements suffice. These are of

course *theoretical* requirements, as the problem (L_0) is a combinatorial optimization problem and is NP-Hard in general (see Sec. 9.2.2 of [19]).

Work in the field of compressed sensing has however provided us with numerically feasible methods for sparse signal recovery. One such method is ℓ_1-minimization which is a relaxation of (L_0):

$$(L_1) \qquad \hat{x} = \operatorname*{argmin}_{z} \|z\|_1 \quad \text{such that} \quad Az = y.$$

It has been shown that for certain measurement matrices A, (L_0) and (L_1) are equivalent [11, 7]. These measurement ensembles can be taken randomly (for example, A can be chosen to have Gaussian entries), and require $m \geq \mathrm{O}(s\log(n/s))$ measurements to guarantee reconstruction of all s-sparse vectors. Thus we require slightly more measurements (from $2s$ to $Cs\log(n/s)$) but can recover via the problem (L_1) which is numerically feasible by linear programming methods. For weak recovery we need only slightly fewer measurements, see [12] for precise thresholds.

1.1. **Low-Rank Matrix Recovery.** A related problem to compressed sensing is the problem of low-rank matrix recovery, for which many results have been obtained (see e.g. [2, 4, 5, 10, 18, 15, 17, 25, 13, 3]). In this setting, we would like to recover a matrix M from few of its linear measurements. 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:

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

Given the measurements $\mathcal{A}(M) \in \mathbb{R}^m$, we wish to recover the matrix $M \in \mathbb{R}^{n \times n}$. This is of course again ill-posed for small m in general. However, if we operate under the assumption that M has low rank then the problem can be made well-posed. The question then becomes how large does m need to be in order to guarantee recovery of rank-r matrices and how does one recover such a matrix? Analagous to the program (L_0), one can consider solving

$$(1.1) \qquad \hat{X} = \operatorname*{argmin}_{X} \operatorname{rank}(X) \quad \text{such that} \quad \mathcal{A}(X) = \mathcal{A}(M).$$

This is simply a uniqueness problem; when is M the unique low rank matrix having these measurements? However, as in the case of (L_0), the problem (1.1) is intractible in general.

Instead of solving (1.1), we are often interested in a tractible method which provides worst-case guarantees; that is, guarantees which apply to *all* rank-r matrices whether arbitrary or adversarial. A simple observation allows one to select an appropriate

relaxation of (1.1) that will do just this. The rank of a matrix is the number of non-zero singular values. That is, if σ is the vector of singular values of M, then $\operatorname{rank}(M) = \|\sigma\|_0$. Thus a natural relaxation would be to minimize $\|\sigma\|_1$. We thus consider the minimization problem

$$(1.2) \qquad \hat{X} = \operatorname*{argmin}_{X} \|X\|_* \quad \text{such that} \quad \mathcal{A}(X) = \mathcal{A}(M),$$

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

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

The program (1.2) can be cast as a semidefinite program (SDP) and is therefore numerically feasible. Moreover, it has been shown [20, 25, 21, 3] that $m \geq Cnr$ measurements suffice to recover any $n \times n$ rank-r matrix via (1.2).

A question that does not appear to have been previously addressed is, how many measurements suffice to recover rank-r matrices via the more natural (yet intractible) method (1.1)? In the compressed sensing setting, this question 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 has remained unresolved. 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. In the case of compressed sensing for example, it is clear that to use a tractible method we pay in the number of measurements by a factor of $\log(n/s)$. What is this factor in the matrix recovery framework? How good is nuclear-norm minimization? These are the issues we address in this work.

In this paper we prove that $4nr - 4r^2$ measurements are sufficient to recover all rank$-r$ $n \times n$ matrices using rank minimization almost surely. To recover a fixed rank$-r$ $n \times n$ matrix with probability one, we show that only $2nr - r^2 + 1$ measurements are required. We then compare our results to nuclear norm minimization and show that rank minimization requires less measurements, but only by a constant factor.

2. Uniqueness Results

In this section we provide a detailed summary of our main results.

We consider random operators $\mathcal{A}$, and first ask that for any rank-r matrix M, the solution to (1.1) is $\hat{X} = M$ with probability one. If this were not the case, then there would be some matrices M and M' each with rank-r or less such that $\mathcal{A}(M) = \mathcal{A}(M')$. This means that the rank-$2r$ (or less) matrix $M - M'$ is in the

null space of $\mathcal{A}$. Therefore, 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 $\mathbb{R}^{n \times n}$:

$$(2.1) \qquad\qquad \mathcal{R}' = \{X \in \mathbb{R}^{n \times n} : \operatorname{rank}(X) = 2r\}.$$

We first wish to compute how large m must be so that the null space of $\mathcal{A}$ is disjoint from $\mathcal{R}$. We will then repeat this argument for smaller values of the rank.

It is well known that $\mathcal{R}'$ is a manifold with $4nr - 4r^2$ dimensions. Is $m \geq 4nr - 4r^2$ 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 $r \leq n/2$. When $\mathcal{A} : \mathbb{R}^{n \times n} \to \mathbb{R}^m$ is a Gaussian operator with $m \geq 4nr - 4r^2$, problem (1.1) recovers all rank-r matrices with probability 1.*

Remarks.
1. We actually prove a more general result in Theorem 3.1. In this result we consider any random linear operator $\mathcal{A}$ which takes $m \geq d + 1$ measurements $\langle A_i, X \rangle$ where $\langle A_i, X \rangle$ are independent and do not concentrate around zero. Then Theorem 3.1 shows that any d-dimensional continuously differentiable manifold over the set of real matrices is disjoint (except possibly at the origin) from the null space of $\mathcal{A}$. 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.

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 $2nr - r^2$. Recall that in compressed sensing for weak recovery we a require number of measurements equal to at least one more than the sparsity level. 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 $m \geq 2nr - r^2 + 1$, 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

On our way to proving our main results, Theorems 2.1 and 2.2, we will prove a more general result about arbitrary manifolds of real matrices. This result can be extended even further by considering manifolds over more arbitrary Banach spaces and following our proof. 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 $m \geq d + 1$ 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,*

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

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 $m \geq d+1$ is tight in the sense that the result does not generally hold for $m \leq d$. 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 $\text{Null}(A) \cap \mathcal{R} \backslash \{0\} \neq \emptyset$ for any linear operator $\mathcal{A} : \mathbb{R}^{n \times n} \to \mathbb{R}^m$ as long as $m \leq d$.

To prove this result we will utilize 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 now well known (see e.g. Ch. 13 of [16]):

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

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

We are now prepared to prove Theorem 3.1.

Proof of Theorem 3.1. For simplicity we take $m = d + 1$. Since $\mathcal{R}$ is a continuously differentiable manifold, so is $\mathcal{R}\backslash\{0\}$ and this implies that 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 $\phi_i : \mathcal{V}_i \to B_2^d$. In words, there is a homeomorphism ϕ_i from $\mathcal{V}_i$ to the unit Euclidean ball in $\mathbb{R}^d$ (denoted B_2^d) such that ϕ_i and ϕ_i^{-1} are continuously differentiable.

Our strategy is to show that for fixed i, $0 \notin \mathcal{A}(\mathcal{V}_i)$ with probability 1. We will then apply a union bound using the fact that there are only countably many $\mathcal{V}_i$.

Fix an 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

$$(3.1) \qquad \|\phi^{-1}(x) - \phi^{-1}(y)\|_F \le L\|x - y\|_2.$$

Next, let $\overline{B_2^d}$ be an (ε/L)-net for B_2^d with cardinality at most $\left(\frac{3L}{\varepsilon}\right)^d$. This is of course possible by Lemma 3.2. 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 $\overline{b} \in \overline{B_2^d}$ such that

$$\|\overline{b} - \phi(X)\|_2 \le \frac{\varepsilon}{L}.$$

By (3.1) we then have

$$\|\phi^{-1}(\overline{b}) - X\|_F \le L \cdot \|\overline{b} - \phi(X)\|_2 \le L \cdot \frac{\varepsilon}{L} = \varepsilon.$$

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

Using the fact that $\overline{\mathcal{V}}$ is an ε-net for $\mathcal{V}$, we have that for any $X \in \mathcal{V}$, there is an $\overline{X} \in \overline{\mathcal{V}}$ such that $\|X - \overline{X}\|_F \le \varepsilon$. This then implies

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

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

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

---

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

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

$$\mathbb{P}\left(\inf_{X\in\mathcal{V}}\|\mathcal{A}(X)\|_\infty = 0\right) \leq \mathbb{P}\left(\inf_{X\in\mathcal{V}}\|\mathcal{A}(X)\|_\infty \leq \varepsilon\log(1/\varepsilon)\right)$$

$$\leq \mathbb{P}\left(\min_{\overline{X}\in\overline{\mathcal{V}}}\|\mathcal{A}(\overline{X})\|_\infty - \varepsilon\cdot\|\mathcal{A}\|_{F\to\infty} \leq \varepsilon\log(1/\varepsilon)\right).$$

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

$$\mathbb{P}\left(\min_{\overline{X}\in\overline{\mathcal{V}}}\|\mathcal{A}(\overline{X})\|_\infty - \varepsilon\cdot\|\mathcal{A}\|_{F\to\infty} \leq \varepsilon\log(1/\varepsilon)\right)$$

$$(3.2) \qquad \leq \mathbb{P}\left(\min_{\overline{X}\in\overline{\mathcal{V}}}\|\mathcal{A}(\overline{X})\|_\infty \leq 2\varepsilon\log(1/\varepsilon)\right) + \mathbb{P}\Big(\|\mathcal{A}\|_{F\to\infty} > \log(1/\varepsilon)\Big).$$

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

$$\mathbb{P}\left(\min_{\overline{X}\in\overline{\mathcal{V}}}\|\mathcal{A}(\overline{X})\|_\infty \leq 2\varepsilon\log(1/\varepsilon)\right) \leq |\overline{\mathcal{V}}|\cdot\mathbb{P}\left(\|\mathcal{A}(\overline{X})\|_\infty \leq 2\varepsilon\log(1/\varepsilon)\right)$$

$$= |\overline{\mathcal{V}}|\cdot\mathbb{P}\left(\max\{|z_1|,\ldots,|z_m|\} \leq 2\varepsilon\log(1/\varepsilon)\right)$$

$$\leq \left(\frac{3L}{\varepsilon}\right)^d\cdot\prod_{i=1}^m\Big(\mathbb{P}\left(|z_i| \leq 2\varepsilon\log(1/\varepsilon)\right)\Big),$$

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,

$$\mathbb{P}\left(|z_i| \leq 2\varepsilon\log(1/\varepsilon)\right) = \mathbb{P}\left(|\langle A_i, X\rangle| \leq 2\varepsilon\log(1/\varepsilon)\right)$$

$$= \mathbb{P}\left(\left|\left\langle A_i, \frac{X}{\|X\|_F}\right\rangle\right| \leq \frac{2\varepsilon\log(1/\varepsilon)}{\|X\|_F}\right).$$

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

$$\left(\frac{3L}{\varepsilon}\right)^d\cdot\prod_{i=1}^m\Big(\mathbb{P}\left(|z_i| \leq 2\varepsilon\log(1/\varepsilon)\right)\Big) \leq \left(\frac{3L}{\varepsilon}\right)^d\cdot(4C'\varepsilon\log(1/\varepsilon))^m$$

$$= C''\varepsilon^{m-d}\cdot(\log(1/\varepsilon))^m$$

$$= C''\varepsilon\cdot(\log(1/\varepsilon))^m,$$

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.

$\square$

Now we turn to proving our main result Theorem 2.1. To prove this theorem, it will be useful to view the space of rank-$2r$ unit norm matrices as a smooth manifold. Then Theorem 2.1 will follow as a corollary of Theorem 3.1. We denote by $\|\cdot\|_F$ the usual Frobenius norm for matrices.

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

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

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

Proof. We will first show that the space of rank-$2r$ matrices (with arbitrary Frobenius norm) is a smooth manifold. This is a well-known result but we sketch the proof. Then we will show that the intersection of this space and the sphere of all unit norm matrices is transverse which will yield the desired result. To this end, let $\mathcal{R}'$ be as in (2.1) and let $\mathcal{M} = \{A \in \mathbb{R}^{n \times n}\}$ be the set of all $n \times n$ matrices. Let $\mathcal{G}$ be the lie group consisting of the cross product of the general linear group with itself:

$$\mathcal{G} = GL(n, \mathbb{R}) \times GL(n, \mathbb{R}).$$

For an element $(g, h) \in \mathcal{G}$, let it act on elements $A \in \mathcal{R}'$ by $(g, h)A = gAh^{-1}$. Since this is just matrix multiplication, this action is clearly continuous. Moreover, it is transitive. Indeed, let $A, B \in \mathcal{R}'$ Since A is rank $2r$, there are $g, h \in GL(n, \mathbb{R})$ such that

$$gAh^{-1} = \begin{bmatrix} \mathbb{I}_{2r} & \\ & \mathbb{O}_{n-2r} \end{bmatrix} \overset{\text{def}}{=} D,$$

where $\mathbb{I}_{2r}$ denotes the $2r \times 2r$ identity matrix and $\mathbb{O}_{n-2r}$ denotes the $n - 2r \times n - 2r$ matrix of zeros. Similarly, there are $y, z \in GL(n, \mathbb{R})$ such that yBz^{-1} equals this same block matrix. Thus $gAh^{-1} = yBz^{-1}$ and so $A = g^{-1}yBz^{-1}h$ which proves transitivity since $(g^{-1}y, z^{-1}h) \in \mathcal{G}$. Next let $\mathcal{H}$ be the stabilizer of the matrix D under the action of $\mathcal{G}$. Since the action of $\mathcal{G}$ is continuous and transitive, the stabilizer $\mathcal{H}$ is a closed subgroup of the lie group $\mathcal{G}$ and thus $\mathcal{H}$ is a closed lie subgroup. Therefore $\mathcal{G}/\mathcal{H}$ is a smooth manifold. But $\mathcal{H}$ is the stabilizer of D under $\mathcal{G}$ and so since the action is transitive, $\mathcal{G}/\mathcal{H}$ must be isomorphic to the orbit of D under the action of $\mathcal{G}$. But the orbit of D is precisely our set $\mathcal{R}'$ and thus $\mathcal{R}'$ is also a smooth manifold.

We now wish to show that $\mathcal{R}$ is also a smooth manifold by viewing $\mathcal{R}$ as the intersection of $\mathcal{R}'$ and the sphere

$$\mathcal{S} = \{A \in \mathcal{M} : \|A\|_F = 1\}.$$

It is clear that $\mathcal{S}$ is a smooth manifold (it is a sphere of the smooth manifold $\mathcal{M}$) and by above $\mathcal{R}'$ is also a smooth manifold. Since $\mathcal{R} = \mathcal{R}' \cap \mathcal{S}$ and both $\mathcal{R}'$ and $\mathcal{S}$ are smooth manifolds, to show that $\mathcal{R}$ is also a smooth manifold it suffices to show that this intersection is transverse. That is, we need to show that for any $A \in \mathcal{R}' \cap \mathcal{S}$, the direct sum of the tangent space of $\mathcal{S}$ at A and $\mathcal{R}'$ at A is equal to the tangent space of $\mathcal{M}$ of A:

$$T_A(\mathcal{S}) \oplus T_A(\mathcal{R}') = T_A(\mathcal{M}).$$

Since $\mathcal{S}$ has codimension 1 in $\mathcal{M}$, it will suffice to show that $T_A(\mathcal{R}')$ contains a vector in the direction normal to the sphere. Now for any $A \in \mathcal{R}'$ and $k \neq 0$, the matrix kA will clearly also have rank $2r$ and thus be contained in $\mathcal{R}'$. Therefore $\mathcal{R}'$ contains the entire line L through the origin containing A (excluding the actual origin itself). Then since $L \subset \mathcal{R}'$, we have $L = T_A(L) \subset T_A(\mathcal{R}')$. Thus we must indeed have $T_A(\mathcal{S}) \oplus T_A(\mathcal{R}') = T_A(\mathcal{M})$. Therefore the intersection $\mathcal{R} = \mathcal{S} \cap \mathcal{R}'$ is transverse and so it is a smooth manifold.

Finally, it is well known that the dimension of the manifold $\mathcal{R}'$ is $4nr - 4r^2$ (see [14, Chapter 8]) and the codimension of $\mathcal{S}$ in $\mathcal{M}$ is 1. Thus $\mathrm{codim}(\mathcal{R}) = \mathrm{codim}(\mathcal{S}) + \mathrm{codim}(\mathcal{R}') = n^2 - (4nr - 4r^2) + 1$ and so the dimension of $\mathcal{R}$ is $4nr - 4r^2 - 1$.

$\square$

We finally show that Theorem 2.1 and Theorem 2.2 follow as corollaries.

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 $m \geq 4nr - 4r^2$ 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, $\mathrm{Null}(\mathcal{A}) \cap \mathcal{R} = \emptyset$. 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. $\square$

Proof of Theorem 2.2. Let $\mathcal{W} = \{X - M : \mathrm{rank}(X) = r\}$. 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

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

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

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

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

4. DISCUSSION

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. As we did in the compressed sensing case, it is of course insightful to compare rank minimization and nuclear-norm minimization. To (provably) recover $n \times n$ rank-r matrices using nuclear-norm minimization, one needs Cnr measurements. As discussed in [3], by observing that the space of rank-r matrices has a subspace which consists of all rank-r matrices whose last $n - r$ rows are zero, one sees that at least $2nr$ measurements are required to recover all rank-r matrices. In [21] explicit formulas and graphs are given from which bounds on the constant C can be derived. Even more recent results in [22] prove that $6nr$ measurements suffice for weak recovery and $16nr$ measurements suffice for strong recovery. New work in [23, 24] also shows weak recovery when $m \geq 6nr - 3r^2$. In addition, numerical results indicate that weak recovery requires about $4nr - 2r^2$ Gaussian measurements [21, Figure 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 tractible method whereas rank minimization is an intractible method whose guarantees give us theoretical bounds with which to compare. In fact, the price to pay for a tractible 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. We conjecture that the strong recovery requirement, $m \geq 4nr - 4r^2$ 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 $m \geq 2nr - 2r^2 + 1$ 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.

Our results in conjunction with work on nuclear-norm minimization show how close nuclear-norm minimization guarantees are to those of the intractible 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] R. Baraniuk. Compressive sensing. *IEEE Sig. Proc. Mag.*, 24(4):118–121, 2007.

[2] J.F. Cai, E. J. Candes, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.*, 20(4):1956–1982, 2008.

[3] E. J. Candes and Y. Plan. Tight oracle bounds for low-rank matrix recovery from a minimal number of random measurements. *IEEE Trans. Inform. Theory*, 2009. to appear.

[4] E. J. Candes and Y. Plan. Matrix completion with noise. *Proc. of the IEEE*, 98(6):925–936, 2010.

[5] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.

[6] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete Fourier information. *IEEE Trans. Info. Theory*, 52(2):489–509, Feb. 2006.

[7] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51:4203–4215, 2005.

[8] E. J. Candès and M. Wakin. An introduction to compressive sampling. *IEEE Signal Proc. Mag.*, 25(2):21–30, 2008.

[9] Compressed sensing webpage. http://www.dsp.ece.rice.edu/cs/.

[10] W. Dai and O. Milenkovic. SET: an algorithm for consistent matrix completion. In *ICASSP 2010 IEEE Int. Conf.*, pages 3646–3649. IEEE, 2010.

[11] D. L. Donoho and P. B. Stark. Uncertainty principles and signal recovery. *SIAM J. Appl. Math.*, 49(3):906–931, June 1989.

[12] D. L. Donoho and J. Tanner. *J. of the AMS*, 22.

[13] D. Gross. Recovering low-rank matrices from few coefficients in any basis. to appear, 2009.

[14] J.M. Lee. *Introduction to smooth manifolds*. Springer Verlag, 2003.

[15] K. Lee and Y. Bresler. Admira: Atomic decomposition for minimum rank approximation. *Information Theory, IEEE Transactions on*, 56(9):4402–4416, 2010.

[16] G. G. Lorentz, M. von Golitschek, and Y. Makovoz. *Constructive Approximation: Advanced Problems*, volume 304. Springer, Berlin, 1996.

[17] S. Ma, D. Goldfarb, and L. Chen. Fixed point and Bregman iterative methods for matrix rank minimization. *Mathematical Programming*, pages 1–33, 2009.

[18] R. Meka, P. Jain, and I.S. Dhillon. Guaranteed rank minimization via singular value projection. Preprint, 2009.

[19] S. Muthukrishnan. *Data Streams: Algorithms and Applications*. Now Publishers, Hanover, MA, 2005.

[20] S. Negahban, P. Ravikumar, M.J. Wainwright, and B. Yu. A unified framework for high-dimensional analysis of *M*-estimators with decomposable regularizers. 2010. Submitted.

[21] S. Oymak and B. Hassibi. New Null Space Results and Recovery Thresholds for Matrix Rank Minimization. Submitted, 2010.

[22] S. Oymak and B. Hassibi. Tight Recovery Thresholds and Robustness Analysis for Matrix Rank Minimization. In preparation, 2010.

[23] P.A. Parrilo, A.S. Willsky, V. Chandrasekaran, and B. Recht. The Convex Geometry of Linear Inverse Problems. Submitted, 2010.

[24] B. Recht, Mar. 2011. Personal communication.

[25] B. Recht, M. Fazel, and P.A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. Preprint, 2007.