Forty-Fifth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
September 26-28, 2007

Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization

Benjamin Recht Maryam Fazel Pablo A. Parrilo

I. INTRODUCTION

Notions such as order, complexity, or dimensionality can often be expressed by the rank of an appropriate matrix. For example, a low-rank matrix could correspond to a low-order realization of a linear system (see, e.g., [1]), to a low-order controller for a plant (see, e.g., [2]), to a low-degree statistical model for a random process (see, e.g., [3]), or to a low-dimensional embedding of data in Euclidean space (see, e.g., [4]). If the set of feasible models or designs is affine in the matrix variable, choosing the simplest model can be cast as an *affine rank minimization problem*,

$$\begin{array}{ll} \text{minimize} & \operatorname{rank}(X) \\ \text{subject to} & \mathcal{A}(X) = b, \end{array} \tag{1}$$

where $X \in \mathbb{R}^{m\times n}$ is the decision variable, and the linear map $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ and the vector $b \in \mathbb{R}^p$ are given. In certain instances with very special structure, the rank minimization problem can be solved by using the singular value decomposition, or can be exactly reduced to the solution of linear systems [5], [6]. For the general rank minimization problem, a variety of heuristic algorithms based on local optimization, including alternating projections [7] and alternating LMIs [8], have been proposed. However, in general, problem (1) is a challenging nonconvex optimization problem for which all known finite time algorithms have at least doubly exponential running times in both theory and practice.

A recent heuristic introduced in [9] minimizes the *nuclear norm* over the affine subset. For an $m \times n$ matrix X, the nuclear norm is equal to the sum of the singular values of X

$$\|X\|_* := \sum_{i=1}^{r} \sigma_i(X)\,, \tag{2}$$

where $\sigma_i(X)$ denotes the ith largest singular value of X (i.e., square root of the ith largest eigenvalue of $X'X$). The nuclear norm (also known as the Schatten 1-norm, the Ky Fan r-norm, and the trace class norm) is a convex function, can be optimized efficiently, and is the best convex approximation of the rank function over the unit ball of matrices with norm less than one. When the matrix variable is symmetric and positive semidefinite, this heuristic is equivalent to the trace heuristic often used by the control community (see, e.g., [5], [10]). The nuclear norm heuristic has been observed to produce very low-rank solutions in practice, but a theoretical characterization of when it produces the minimum rank solution has not been previously available. This paper provides the first such mathematical characterization.

We delineate a large set of instances where the nuclear norm heuristic solves (1) by building upon the extensive literature on a closely related optimization problem—finding the *sparsest vector* in an affine subspace. This problem is commonly referred to as *cardinality minimization*, since we seek the vector whose support has the smallest cardinality, and is known to be NP-hard [11]. When the matrix variable is constrained to be diagonal, affine rank minimization directly reduces to affine cardinality minimization problem. Moreover, for diagonal matrices, the sum of the singular values is equal to the sum of the absolute values (i.e., the ℓ_1 norm) of the diagonal elements. And since singular values are all positive, the nuclear norm is equal to the ℓ_1 norm of the vector of singular values. Minimization of the ℓ_1 norm is a well-known heuristic for the cardinality minimization problem, and stunning results pioneered by Candès and Tao [12] and Donoho [13] have characterized a vast set of instances for which the ℓ_1 heuristic can be *a priori* guaranteed to yield the optimal solution. These techniques provide the foundations of the recently developed *compressed sensing* or *compressive sampling* frameworks for measurement, coding, and signal estimation. As has been shown by a number of research groups (e.g., [14], [15], [16], [17] to name a few), the ℓ_1 heuristic for cardinality minimization provably recovers the sparsest solution whenever the sensing matrix has certain "basis incoherence" properties, and in particular, when it is randomly chosen according to certain specific ensembles.

The fact that the ℓ_1 heuristic is a special case of the nuclear norm heuristic suggests that these results from the compressed sensing literature might be extended to provide guarantees about the nuclear norm heuristic for the more general rank minimization problem. In this work, we show that this is indeed the case, and the parallels are surprisingly strong. Following the program laid out in the work of Candès and Tao, the main contribution of this paper is the development of a restricted isometry property (RIP), under which the nuclear norm heuristic can be *guaranteed* to produce the minimum-rank solution. Furthermore, as in the case for the ℓ_1 heuristic, we discuss several specific examples

Center for the Mathematics of Information, California Institute of Technology
Control and Dynamical Systems, California Institute of Technology
Laboratory for Information and Decision Systems, Massachusetts Institute of Technology

of matrix ensembles for which RIP holds with overwhelming probability, provided the codimension of the subspace is $\Omega(r(m+n)\log mn)$, where m,n are the dimensions of the matrix, and r is its rank. Finally, we demonstrate that in practice nuclear-norm minimization recovers the lowest rank solutions of affine sets with even fewer constraints than those guaranteed by our mathematical analysis. Our results considerably extend the compressed sensing machinery in a so far undeveloped direction by allowing a much more general notion of parsimonious models that rely on low-rank assumptions instead of cardinality restrictions.

II. Restricted Isometry and Recovery of Low-Rank Matrices

The three vector norms that play significant roles in the compressed sensing framework are the ℓ_1, ℓ_2, and ℓ_∞ norms. When these norms are applied to the singular values of a matrix, they result in unitarily invariant matrix norms. The ℓ_1 norm of the singular values is the nuclear norm, the ℓ_2 norm of the singular values is the Frobenius norm

$$\|X\|_F := \sqrt{\mathrm{Tr}(X'X)} = \left(\sum_{i=1}^r \sigma_i^2\right)^{\frac{1}{2}},$$

and the ℓ_∞ norm of the singular values is the operator norm

$$\|X\| := \sigma_1(X).$$

Most of our results can be derived by the following program. Beginning with a result from compressed sensing, we replace the vector norms with their associated matrix norms, and then extend the proofs in the vector case to the more general matrix setting. This section illustrates just how fruitful this program can be.

Let $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ be a linear map and let X_0 be a matrix of rank r. Set $b := \mathcal{A}(X_0)$, and define the convex optimization problem

$$X^* := \arg\min_X \|X\|_* \qquad \text{s.t.} \quad \mathcal{A}(X) = b. \tag{3}$$

In this section, we will characterize specific cases when we can *a priori* guarantee that $X^* = X_0$. The key conditions will be determined by the values of a sequence of parameters δ_r that quantify the behavior of the linear map $\mathcal{A}$ when restricted to the subvariety of matrices of rank r. The following definition is the natural generalization of the Restricted Isometry Property from vectors to matrices.

Definition 2.1: Let $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ be a linear map. Without loss of generality, assume $m \leq n$. For every integer r with $1 \leq r \leq m$, define the r-restricted isometry constant to be the smallest number $\delta_r(\mathcal{A})$ such that

$$(1-\delta_r(\mathcal{A}))\|X\|_F \leq \|\mathcal{A}(X)\| \leq (1+\delta_r(\mathcal{A}))\|X\|_F \tag{4}$$

holds for all matrices X of rank at most r.

Note that by definition, $\delta_r(\mathcal{A}) \leq \delta_{r'}(\mathcal{A})$ for $r \leq r'$.

The Restricted Isometry Property for sparse vectors was developed by Candès and Tao in [14], and requires (4) to hold with the Euclidean norm replacing the Frobenius norm and rank being replaced by cardinality. Since for diagonal matrices, the Frobenius norm is equal to the Euclidean norm of the diagonal, this definition reduces to the original Restricted Isometry Property of [14] in the diagonal case.[1]

Unlike the case of "standard" compressed sensing, our RIP condition for low-rank matrices cannot be interpreted as guaranteeing all sub-matrices of the linear transform $\mathcal{A}$ of a certain size are well conditioned. Indeed, the set of matrices X for which (4) must hold is *not* a finite union of subspaces, but rather a certain "generalized Stiefel manifold," which is also an algebraic variety (in fact, it is the rth-secant variety of the variety of rank-one matrices). Surprisingly, we are still able to derive analogous recovery results for low-rank solutions of equations when $\mathcal{A}$ obeys this RIP condition. Furthermore, we will see in Section III that many ensembles of random matrices obey the Restricted Isometry Property with small δ_r with high probability for reasonable values of m,n, and p.

The following two recovery theorems will characterize the power of the restricted isometry constants. Both theorems are more or less immediate generalizations from the sparse case to the low-rank case and use only minimal properties of the rank of matrices and the nuclear norm. The first theorem generalizes Lemma 1.3 in [14] to low-rank recovery.

Theorem 2.2: Suppose that $\delta_{2r} < 1$ for some integer $r \geq 1$. Then X_0 is the only matrix of rank at most r satisfying $\mathcal{A}(X) = b$.

The proof of this theorem is an immediate consequence of our definition of the constant δ_r. Assume, on the contrary, that there exists a rank r matrix X satisfying $\mathcal{A}(X) = b$ and $X \neq X_0$. Then $Z := X_0 - X$ is a nonzero matrix of rank at most $2r$, and $\mathcal{A}(Z) = 0$. But then we would have $0 = \|\mathcal{A}(Z)\| \geq (1-\delta_{2r})\|Z\|_F > 0$ which is a contradiction. This is an identical argument to that given by Candès and Tao and no adjustment is necessary in the transition from sparse vectors to low-rank matrices. The key property used is the sub-additivity of the rank: $\mathrm{rank}(A+B) \leq \mathrm{rank}(A) + \mathrm{rank}(B)$. Using a somewhat more complicated argument, we can formulate a weak ℓ_1-type recovery theorem whose proof mimics the approach in [15], but for which a few details need to be adjusted when switching from vectors to matrices.

Theorem 2.3: Suppose that $r \geq 1$ is such that $\delta_{5r} < 1/10$. Then $X^* = X_0$.

We will need two technical lemmas, both of which are proven using elementary linear algebra in the appendix. The first shows for any two matrices A and B, we can decompose B as the sum of two matrices B_1 and B_2 such that $\mathrm{rank}(B_1)$ is not too large and such that B_2 and A have orthogonal row and column spaces. In this case, the second lemma shows that $\|A + B_2\|_* = \|A\|_* + \|B_2\|_*$. This will be the key decomposition for proving Theorem 2.3.

[1] In [14], the authors define the restricted isometry properties with squared norms. We note here that the analysis is identical modulo some algebraic rescaling of constants. We choose to drop the squares as it greatly simplifies the analysis.

Lemma 2.4: Let A and B be matrices of the same dimensions. Then there exist matrices B_1 and B_2 such that

1) $B = B_1 + B_2$
2) $\operatorname{rank}(B_1) \leq 2\operatorname{rank}(A)$
3) $AB_2' = 0$ and $A'B_2 = 0$
4) $\langle B_1, B_2 \rangle = 0$

Lemma 2.5: Let A and B be matrices of the same dimensions. If $AB' = 0$ and $A'B = 0$ then $\|A + B\|_* = \|A\|_* + \|B\|_*$.

We now proceed to a proof of Theorem 2.3.

Proof: **[of Theorem 2.3]** By optimality of X^*, we have $\|X_0\|_* \geq \|X^*\|_*$. Let $R := X^* - X_0$. Applying Lemma 2.4 to the matrices X_0 and R, there exist matrices R_0 and R_c such that $R = R_0 + R_c$, $\operatorname{rank}(R_0) \leq 2\operatorname{rank}(X_0)$, and $X_0 R_c' = 0$ and $X_0' R_c = 0$. Then,

$$\begin{aligned} \|X_0\|_* &\geq \|X_0 + R\|_* \\ &\geq \|X_0 + R_c\|_* - \|R_0\|_* \\ &= \|X_0\|_* + \|R_c\|_* - \|R_0\|_*, \end{aligned} \tag{5}$$

where the middle assertion follows from the triangle inequality and the last one from Lemma 2.5. Rearranging terms, we can conclude that

$$\|R_0\|_* \geq \|R_c\|_*. \tag{6}$$

Next we partition R_c into a sum of matrices $R_1, R_2, \ldots$, each of rank at most $3r$. Let $R_c = U \operatorname{diag}(\sigma) V'$ be the singular value decomposition of R_c. For each $i \geq 1$ define the index set $I_i = \{3r(i-1)+1, \ldots, 3ri\}$, and let $R_i := U_{I_i} \operatorname{diag}(\sigma_{I_i}) V_{I_i}'$ (notice that $\langle R_i, R_j \rangle = 0$ if $i \neq j$). By construction, we have

$$\sigma_k \leq \frac{1}{3r} \sum_{j \in I_i} \sigma_j \qquad \forall\, k \in I_{i+1}, \tag{7}$$

which implies $\|R_{i+1}\|_F^2 \leq \frac{1}{3r}\|R_i\|_*^2$. We can then compute the following bound

$$\begin{aligned} \sum_{j \geq 2} \|R_j\|_F &\leq \frac{1}{\sqrt{3r}} \sum_{j \geq 1} \|R_j\|_* \\ &= \frac{1}{\sqrt{3r}} \|R_c\|_* \leq \frac{1}{\sqrt{3r}} \|R_0\|_* \\ &\leq \frac{\sqrt{2r}}{\sqrt{3r}} \|R_0\|_F\,, \end{aligned} \tag{8}$$

where the last inequality follows because $\operatorname{rank}(R_0) \leq 2r$ and $\|X\|_* \leq \sqrt{2r}\|X\|_F$ for all matrices of rank at most $2r$. Finally, note that the rank of $R_0 + R_1$ is at most $5r$, so we may put this all together as

$$\begin{aligned} \|\mathcal{A}(R)\| &\geq \|\mathcal{A}(R_0 + R_1)\| - \sum_{j \geq 2} \|\mathcal{A}(R_j)\| \\ &\geq (1 - \delta_{5r})\,\|R_0 + R_1\|_F - (1 + \delta_{3r}) \sum_{j \geq 2} \|R_j\|_F \\ &\geq \left((1 - \delta_{5r}) - \sqrt{\tfrac{2}{3}}(1 + \delta_{3r}) \right) \|R_0\|_F \\ &\geq \left((1 - \delta_{5r}) - \tfrac{9}{11}(1 + \delta_{3r}) \right) \|R_0\|_F. \end{aligned} \tag{9}$$

By assumption $\mathcal{A}(R) = \mathcal{A}(X^* - X_0) = 0$, so if the factor on the right-hand side is strictly positive, $R_0 = 0$, which further implies $R_c = 0$ by (6), and thus $X^* = X_0$. Simple algebra reveals that the right-hand side is positive when $9\delta_{3r} + 11\delta_{5r} < 2$. Since $\delta_{3r} \leq \delta_{5r}$, we immediately have that $X^* = X_0$ if $\delta_{5r} < 1/10$. ■

The rational number $(9/11)$ in the proof of the theorem is chosen for notational simplicity and is clearly not optimal. A slightly tighter bound can be achieved working directly with the second to last line in (9). The most important point is that our recovery condition on δ_{5r} is an absolute constant, independent of m, n, r, and p. We shall discuss in the next section that linear transformations sampled from several families of random matrices with appropriately chosen dimensions have δ_r small with overwhelming probability. The analysis is again similar to the compressive sampling literature, but several details specific to the rank recovery problem need to be employed.

III. Nearly Isometric Families

The following definition characterizes a class of probability distributions obeying certain tail bounds. Our main result is that when we sample linear maps from such distributions, they will obey the Restricted Isometry Property (4) as p, m, and n tend to infinity at appropriate rates.

Definition 3.1: Let $\mathcal{A}$ be a random variable that takes values in linear maps from $\mathbb{R}^{m \times n}$ to $\mathbb{R}^p$. We say that $\mathcal{A}$ is *nearly isometrically distributed* if for all $X \in \mathbb{R}^{m \times n}$

$$\mathbf{E}[\|\mathcal{A}(X)\|^2] = \|X\|_F^2 \tag{10}$$

and for all $0 < \epsilon < 1$ we have,

$$\begin{aligned} \mathbf{P}(|\|\mathcal{A}(X)\|^2 - \|X\|_F^2| \geq \epsilon \|X\|_F^2) \\ \leq 2\exp\left(-\frac{p}{2}(\epsilon^2/2 - \epsilon^3/3) \right) \end{aligned} \tag{11}$$

and for all $t > 0$, we have

$$\mathbf{P}\left(\|\mathcal{A}\| \geq 1 + \sqrt{\frac{mn}{p}} + t \right) \leq \exp(-\gamma p t^2) \tag{12}$$

for some constant $\gamma > 0$.

There are two ingredients for a random linear map to be nearly isometric. First, it must be isometric in expectation. Second, the probability of large distortions of length must be exponentially small. The exponential bound in (11) guarantees union bounds will be small even for rather large sets. This concentration is the typical ingredient required to prove the Johnson-Lindenstrauss Lemma (cf [18], [19]).

The majority of nearly isometric random maps are described in terms of random matrices. For a linear map $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$, we can always write its matrix representation as

$$\mathcal{A}(X) = \mathbf{A} \operatorname{vec}(X)\,, \tag{13}$$

where $\operatorname{vec}(X)$ denotes the vector of X with its columns stacked in order on top of one another, and $\mathbf{A}$ is a $p \times mn$ matrix. We now give several examples of nearly isometric random variables in this matrix representation. The most well-known is the ensemble with independent, identically distributed (i.i.d.) Gaussian entries [19]

$$A_{ij} \sim \mathcal{N}(0, \frac{1}{p}) \,. \tag{14}$$

We also mention the two following ensembles of matrices, described in [18]. One has entries sampled from an i.i.d. symmetric Bernoulli distribution

$$A_{ij} = \begin{cases} \sqrt{\frac{1}{p}} & \text{with probability } \frac{1}{2} \\ -\sqrt{\frac{1}{p}} & \text{with probability } \frac{1}{2} \end{cases}, \tag{15}$$

and the other has zeros in two-thirds of the entries

$$A_{ij} = \begin{cases} \sqrt{\frac{3}{p}} & \text{with probability } \frac{1}{6} \\ 0 & \text{with probability } \frac{2}{3} \\ -\sqrt{\frac{3}{p}} & \text{with probability } \frac{1}{6} \end{cases}. \tag{16}$$

The fact that the top singular value of the matrix $\mathbf{A}$ is concentrated around $1 + \sqrt{D/p}$ for all of these ensembles follows from the work of Yin, Bai, and Krishnaiah, who showed that whenever the entries A_{ij} are i.i.d. with zero mean and finite fourth moment, then the maximum singular value of $\mathbf{A}$ is almost surely $1 + \sqrt{D/p}$ for D sufficiently large [20]. El Karoui uses this result to prove the concentration inequality (12) for all such distributions [21]. The result for Gaussians is rather tight, with $\gamma = 1/2$ (see, e.g., [22]). Finally, note that a random projection also obeys all of the necessary concentration inequalities. Since the norm of a random projection is exactly $\sqrt{D/p}$, (12) holds trivially. The concentration inequality (11) is proven in [19].

Using the concentration inequalities listed in Definition 3.1, Wakin *et al* showed that all of these matrix ensembles obey the RIP conditions for sparse vectors [17]. By an enhancement of their argument, we are able to show that these ensembles also obey our more general RIP conditions for low-rank matrices.

Theorem 3.2: Fix $0 < \delta < 1$. If $\mathcal{A}$ is a nearly isometric random variable, then for every $1 \leq r \leq m$, there exist constants $c_0, c_1 > 0$ depending only on δ such that, with probability at least $1 - \exp(-c_1 p)$, $\delta_r(\mathcal{A}) \leq \delta$ whenever $p \geq c_0 r(m+n)\log(mn)$.

The proof makes use of standard techniques in concentration of measure and can be found in [23]. We first extend the concentration results of [17] to subspaces of matrices. We show that the distortion of a subspace by a linear map is robust to perturbations of the subspace. Finally, we provide an ϵ-net over the set of all subspaces and, using a union bound, show that nearly isometric random variables obey the Restricted Isometry Property (4) with overwhelming probability as the size of the matrices tend to infinity.

Heuristically, the scaling $p = \Omega\left(r(m+n)\log(mn)\right)$ is very reasonable, since a rank r matrix has $r(m+n-r)$ degrees of freedom. This coarse tail bound only provides asymptotic estimates for recovery, and is quite conservative in practice. As we demonstrate next, minimum rank solutions can be determined from between $2r(m+n-r)$ to $4r(m+n-r)$ observations for many practical problems.

Fig. 1. The MIT logo image. The associated matrix has dimensions 46×81 and has rank 5.

IV. Numerical Experiments

To illustrate the scaling of nuclear norm heuristic for a particular matrix M, consider the MIT logo presented in Figure 1. The image has a total of 46 rows and 81 columns (total 3726 elements), and 3 distinct non-zero numerical values corresponding to the colors white, red, and grey. Since the logo only has 5 distinct rows, it has rank 5. For each of the ensembles discussed in Section III, we sampled measurement matrices with p ranging between 700 and 1500, and solved the nuclear norm minimization problem (3).

There are many different approaches to solving (3). As shown in [23], the optimization problem (3) can be reformulated as the semidefinite program

$$\begin{array}{ll} \min\limits_{X, W_1, W_2} & \frac{1}{2}(\mathrm{Tr}(W_1) + \mathrm{Tr}(W_2)) \\ \text{s.t.} & \begin{bmatrix} W_1 & X \\ X' & W_2 \end{bmatrix} \succeq 0 \\ & \mathcal{A}(X) = b \end{array} \tag{17}$$

This optimization problem can be solved with a variety of algorithms. Using a primal-dual interior point method yields highly accurate solutions in a modest amount of time. For example, each semidefinite program can be solved in less than four minutes with the freely available software SeDuMi [24] on a 2.0 GHz Laptop.

Figure 2 plots the Frobenius norm of the difference between the optimal point of the semidefinite program and the true image in Figure 1. We observe a sharp transition to perfect recovery near 1200 measurements which is approximately equal to $2r(m+n-r)$. In Figure 3, we graphically plot the recovered solutions for various values of p under the Gaussian ensemble.

To demonstrate the average behavior of low-rank recovery, we conducted a series of experiments for a variety of the matrix sizes n, ranks r, and numbers of measurements p. For a fixed n, we constructed random recovery scenarios for low-rank $n \times n$ matrices. For each n, we varied p between 0 and n^2 where the matrix is completely discovered. For a fixed n and p, we generated all possible ranks such that $r(2n-r) \leq p$. This cutoff was chosen because beyond that point there would be an infinite set of matrices of rank r satisfying the p equations.

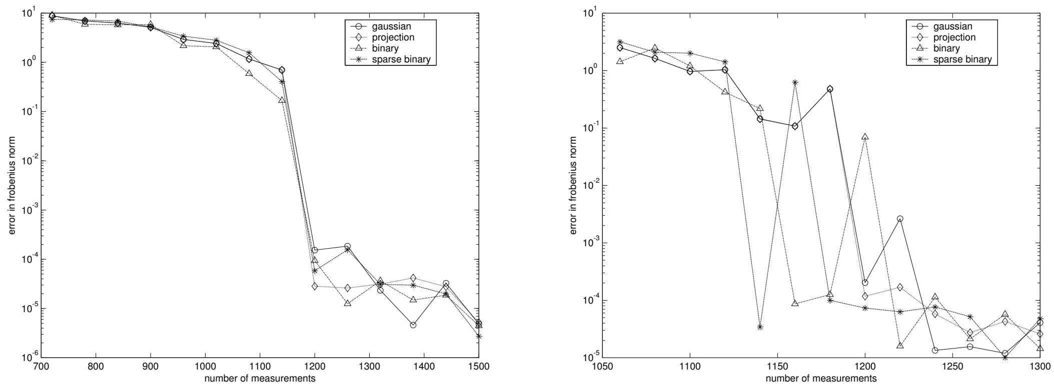

Fig. 2. (left) Error, as measured by the Frobenius norm, between the recovered image and the ground truth. Observe that there is a sharp transition to near zero error at around 1200 measurements. (right) Zooming in on this transition, we see fluctuation between high and low error when between 1125 and 1225 measurements are available.

For each (n, p, r) triple, we repeated the following procedure ten times. A matrix of rank r was generated by choosing two random $n \times r$ factors Y_L and Y_R with i.i.d. random entries and setting $Y_0 = Y_L Y_R'$. A matrix $\mathbf{A}$ was sampled from the Gaussian ensemble with p rows and n^2 columns. Then the nuclear norm minimization

$$\begin{array}{ll} \min\limits_{X} & \|X\|_* \\ \text{s.t.} & \mathbf{A}\,\mathrm{vec}(X) = \mathbf{A}\,\mathrm{vec}(Y_0) \end{array} \qquad (18)$$

was solved using the SDP solver SeDuMi on the formulation (17). We declared Y_0 to be recovered if $\|X - Y_0\|_F / \|Y_0\|_F < 10^{-3}$. Figure 4 shows the results of these experiments for $n = 30$ and 40. The color of the cell in the figures reflects the empirical recovery rate of the ten runs (scaled between 0 and 1). White denotes perfect recovery in all experiments, and black denotes failure for all experiments.

These experiments demonstrate that the logarithmic factors and constants present in our scaling results are somewhat conservative. For example, as one might expect, low-rank matrices are perfectly recovered by nuclear norm minimization when $p = n^2$ as the matrix is uniquely determined. Moreover, as p is reduced slightly away from this value, low-rank matrices are still recovered 100 percent of the time for most values of r. Finally, we note that despite the asymptotic nature of our analysis, our experiments demonstrate excellent performance with low-rank matrices of size 30×30 and 40×40 matrices, showing that the heuristic is practical even in low-dimensional settings.

Intriguingly, Figure 4 also demonstrates a "phase transition" between perfect recovery and failure. As observed in several recent papers by Donoho and his collaborators (See e.g. [25]), the random sparsity recovery problem has two distinct connected regions of parameter space: one where the sparsity pattern is perfectly recovered, and one where no sparse solution is found. Not surprisingly, Figure 4 illustrates an analogous phenomenon in rank recovery. Computing explicit formulas for the transition between perfect recovery and failure is left for future work.

V. Discussion and Future Developments

Having illustrated the natural connections between affine rank minimization and affine cardinality minimization, we were able to draw on these parallels to determine scenarios where the nuclear norm heuristic exactly solves the rank minimization problem. These scenarios directly generalize conditions for which the ℓ_1 heuristic succeeded and ensembles of linear maps for which these conditions hold. Furthermore, our experimental results displayed similar recovery properties to those demonstrated in the empirical studies of ℓ_1 minimization. Inspired by the success of this program, we are excited about pursuing several directions that are natural continuations of this work. These include producing alternative measurement ensembles to those described in Section III, providing analysis of the nuclear norm heuristic when the measurements are noisy, and discussing numerical alternatives to interior point methods.

Finally, we note that sparsity and low-rank are only two particular measures of parsimony. Our work suggests that there may be other kinds of easy-to-describe parametric models where the search for parsimonious models is amenable to exact solutions via convex optimization techniques. Characterizing this broader notion of parsimonious modeling is an exciting line of future inquiry.

References

[1] M. Fazel, H. Hindi, and S. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proceedings of the American Control Conference*, 2001.

[2] L. El Ghaoui and P. Gahinet, "Rank minimization under LMI constraints: A framework for output feedback problems," in *Proceedings of the European Control Conference*, 1993.

[3] J. D. M. Rennie and N. Srebro, "Fast maximum margin matrix factorization for collaborative prediction," in *Proceedings of the International Conference of Machine Learning*, 2005.

[4] N. Linial, E. London, and Y. Rabinovich, "The geometry of graphs and some of its algorithmic applications," *Combinatorica*, vol. 15, pp. 215–245, 1995.

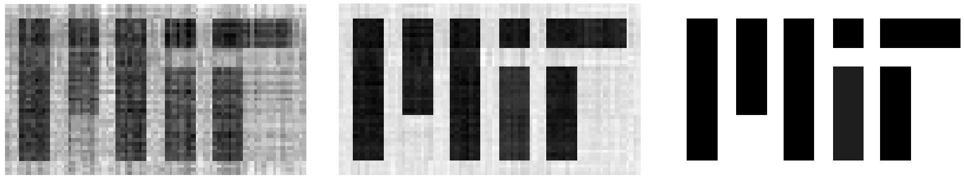

Fig. 3. Example recovered images using the Gaussian ensemble. (left) 700 measurements. (middle) 1100 measurements (right) 1250 measurements. The total number of pixels is $46 \times 81 = 3726$. Note that the error is plotted on a logarithmic scale.

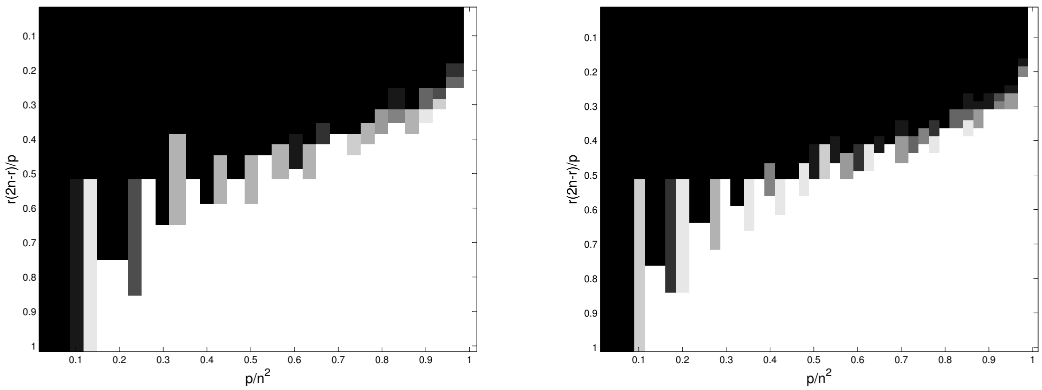

Fig. 4. For each (n, p, r) triple, we repeated the following procedure ten times. A matrix of rank r was generated by choosing two random $n \times r$ factors Y_L and Y_R with i.i.d. random entries and set $Y_0 = Y_L Y_R'$. We select a matrix $\mathbf{A}$ from the Gaussian ensemble with p rows and n^2 columns. Then we solve the nuclear norm minimization subject to $\mathbf{A}\,\mathrm{vec}(X) = \mathbf{A}\,\mathrm{vec}(Y_0)$ We declare Y_0 to be recovered if $\|X - Y_0\|_F / \|Y_0\|_F < 10^{-3}$. The results are shown for (left) $n = 30$ and (right) $n = 40$. The color of each cell reflects the empirical recovery rate (scaled between 0 and 1). White denotes perfect recovery in all experiments, and black denotes failure for all experiments.

[5] M. Mesbahi and G. P. Papavassilopoulos, "On the rank minimization problem over a positive semidefinite linear matrix inequality," *IEEE Transactions on Automatic Control*, vol. 42, no. 2, pp. 239–243, 1997.

[6] P. A. Parrilo and S. Khatri, "On cone-invariant linear matrix inequalities," *IEEE Trans. Automat. Control*, vol. 45, no. 8, pp. 1558–1563, 2000.

[7] K. M. Grigoriadis and E. B. Beran, "Alternating projection algorithms for linear matrix inequalities problems with rank constraints." in *Advances in Linear Matrix Inequality Methods in Control*, L. El Ghaoui and S. Niculescu, Eds. SIAM, 2000, ch. 13, pp. 251–267.

[8] R. E. Skelton, T. Iwasaki, and K. Grigoriadis, *A Unified Algebraic Approach to Linear Control Design*. Taylor and Francis, 1998.

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

[10] C. Beck and R. D'Andrea, "Computational study and comparisons of LFT reducibility methods," in *Proceedings of the American Control Conference*, 1998.

[11] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal of Computing*, vol. 24, no. 2, pp. 227–234, 1995.

[12] E. J. Candès, "Compressive sampling," in *International Congress of Mathematicians. Vol. III*. Eur. Math. Soc., Zürich, 2006, pp. 1433–1452.

[13] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[14] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[15] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications of Pure and Applied Mathematics*, vol. 59, pp. 1207–1223, 2005.

[16] ——, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[17] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Preprint*, 2007, http://dsp.rice.edu/cs/jlcs-v03.pdf.

[18] D. Achlioptas, "Database-friendly random projections: Johnson-Lindenstrauss with binary coins," *Journal of Computer and Systems Science*, vol. 66, no. 4, pp. 671–687, 2003, special issue of invited papers from PODS'01.

[19] S. Dasgupta and A. Gupta, "An elementary proof of a theorem of Johnson and Lindenstrauss," *Random Structures and Algorithms*, vol. 22, no. 1, pp. 60–65, 2003.

[20] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah, "On the limit of the largest eigenvalue of the large dimensional sample covariance matrix," *Probability Theory and Related Fields*, vol. 78, pp. 509–512, 1988.

[21] N. El Karoui, "New results about random covariance matrices and statistical applications," Ph.D. dissertation, Stanford University, 2004.

[22] K. R. Davidson and S. J. Szarek, "Local operator theory, random matrices and Banach spaces," in *Handbook on the Geometry of Banach spaces*, W. B. Johnson and J. Lindenstrauss, Eds. Elsevier Scientific, 2001, pp. 317–366.

[23] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," 2007, preprint available at http://lanl.arxiv.org/abs/0706.4138.

[24] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11-12, pp. 625–653, 1999.

[25] D. L. Donoho and J. Tanner, "Neighborliness of randomly projected simplices in high dimensions," *Proc. Natl. Acad. Sci. USA*, vol. 102, no. 27, pp. 9452–9457, 2005.

APPENDIX

Proof: [of Lemma 2.4] Consider a full singular value decomposition of A

$$A = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V',$$

and let $\hat{B} := U'BV$. Partition $\hat{B}$ as

$$\hat{B} = \begin{bmatrix} \hat{B}_{11} & \hat{B}_{12} \\ \hat{B}_{21} & \hat{B}_{22} \end{bmatrix}.$$

Defining now

$$B_1 := U \begin{bmatrix} \hat{B}_{11} & \hat{B}_{12} \\ \hat{B}_{21} & 0 \end{bmatrix} V',$$
$$B_2 := U \begin{bmatrix} 0 & 0 \\ 0 & \hat{B}_{22} \end{bmatrix} V',$$

it can be easily verified that B_1 and B_2 satisfy the conditions (1)–(4). ■

Proof: [of Lemma 2.5] Partition the singular value decompositions of A and B to reflect the zero and non-zero singular vectors

$$A = \begin{bmatrix} U_{A1} & U_{A2} \end{bmatrix} \begin{bmatrix} \Sigma_A & \\ & 0 \end{bmatrix} \begin{bmatrix} V_{A1} & V_{A2} \end{bmatrix}'$$
$$B = \begin{bmatrix} U_{B1} & U_{B2} \end{bmatrix} \begin{bmatrix} \Sigma_B & \\ & 0 \end{bmatrix} \begin{bmatrix} V_{B1} & V_{B2} \end{bmatrix}'.$$

The condition $AB' = 0$ implies that $V_{A1}'V_{B1} = 0$, and similarly, $A'B = 0$ implies that $U_{A1}'U_{B1} = 0$. Hence, there exist matrices U_C and V_C such that $U_{AB} := [U_{A1}\, U_{B1}\, U_C]$ and $V_{AB} := [V_{A1}\, V_{B1}\, V_C]$ are orthogonal matrices. Thus, the following are valid singular value decompositions for A and B:

$$A = U_{AB} \begin{bmatrix} \Sigma_A & & \\ & 0 & \\ & & 0 \end{bmatrix} V_{AB}'$$
$$B = U_{AB} \begin{bmatrix} 0 & & \\ & \Sigma_B & \\ & & 0 \end{bmatrix} V_{AB}'.$$

In particular, we have that

$$A + B = \begin{bmatrix} U_{A1} & U_{B1} \end{bmatrix} \begin{bmatrix} \Sigma_A & \\ & \Sigma_B \end{bmatrix} \begin{bmatrix} V_{A1} & V_{B1} \end{bmatrix}'.$$

This shows that the singular values of $A + B$ are equal to the union (with repetition) of the singular values of A and B. Hence, $\|A + B\|_* = \|A\|_* + \|B\|_*$ as desired. ■