Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Necessary and Sufficient Conditions for Success of the Nuclear Norm Heuristic for Rank Minimization

Benjamin Recht*, Weiyu Xu**, and Babak Hassibi†

Abstract— **Minimizing the rank of a matrix subject to constraints is a challenging is a challenging problem that arises in many control applications including controller design, realization theory and model reduction. This class of optimization problems, known as rank minimization, is NP-HARD, and for most practical problems there are no efficient algorithms that yield exact solutions. A popular heuristic algorithm replaces the rank function with the nuclear norm—equal to the sum of the singular values—of the decision variable. In this paper, we provide a necessary and sufficient condition that quantifies when this heuristic successfully finds the minimum rank solution of a linear constraint set. We further show that most of the problems of interest in control can be formulated as rank minimization subject to such linear constraints. We additionally provide a probability distribution over instances of the affine rank minimization problem such that instances sampled from this distribution satisfy our conditions for success with overwhelming probability provided the number of constraints is appropriately large. Finally, we give empirical evidence that these probabilistic bounds provide accurate predictions of the heuristic's performance in non-asymptotic scenarios.**

I. INTRODUCTION

Optimization problems involving constraints on the rank of matrices are pervasive in control applications, arising in the context of low-order controller design [7], [12], minimal realization theory [9], and model reduction [2]. Rank minimization is also of interest to a broader optimization community in a variety of applications including inference with partial information [16] and embedding in Euclidean spaces [11]. In certain instances with special structure, the rank minimization problem can be solved via the singular value decomposition or can be reduced to the solution of a linear system [12], [13]. In general, however, minimizing the rank of a matrix subject to convex constraints is NP-HARD. The best exact algorithms for this problem involve quantifier elimination and such solution methods require at least exponential time in the dimensions of the matrix variables.

A popular heuristic for solving rank minimization problems in the controls community is the "trace heuristic" where one minimizes the trace of a positive semidefinite decision variable instead of the rank (see, e.g., [2], [12]). A generalization of this heuristic to non-symmetric matrices introduced by Fazel in [8] minimizes the *nuclear norm*, or the sum of the singular values of the matrix, over the constraint set. When the matrix variable is symmetric and positive semidefinite, this heuristic is equivalent to the trace heuristic, as the trace of a positive semidefinite matrix is equal to the sum of its singular values. The nuclear norm is a convex function and can be optimized efficiently via semidefinite programming. Both the trace heuristic and the nuclear norm generalization have been observed to produce very low-rank solutions in practice, but, until very recently, conditions where the heuristic succeeded were only available in cases that could also be solved by elementary linear algebra [13].

The first non-trivial sufficient conditions that guaranteed the success of the nuclear norm heuristic were provided in [14]. Focusing on the special case where one seeks the lowest rank matrix in an affine subspace, the authors provide a "restricted isometry" condition on the linear map defining the affine subspace which guarantees the minimum nuclear norm solution is the minimum rank solution. Moreover, they provide several ensembles of affine constraints where this sufficient condition holds with overwhelming probability. Their work builds on seminal developments in "compressed sensing" that determined conditions for when minimizing the ℓ_1 norm of a vector over an affine space returns the sparsest vector in that space (see, e.g., [4], [3], [1]). There is a strong parallelism between the sparse approximation and rank minimization settings. The rank of a diagonal matrix is equal to the number of non-zeros on the diagonal. Similarly, the sum of the singular values of a diagonal matrix is equal to the ℓ_1 norm of the diagonal. Exploiting the parallels, the authors in [14] were able to extend much of the analysis developed for the ℓ_1 heuristic to provide guarantees for the nuclear norm heuristic.

Building on a different collection of developments in compressed sensing [5], [6], [17], we present a *necessary* and sufficient condition for the solution of the nuclear norm heuristic to coincide with the minimum rank solution in an affine space. The condition characterizes a particular property of the null-space of the linear map which defines the affine space. To demonstrate why this result is of practical use to the controls community, we also present a reduction of the standard Linear Matrix Inequality (LMI) constrained rank minimization problem to a rank minimization problem with only equality constraints. Moreover, we show that when the linear map defining the constraint set is generated by sampling its entries independently from a Gaussian distribution, the null-space characterization holds with overwhelming probability provided the dimensions of

*Center for the Mathematics of Information, California Institute of Technology, 1200 E California Blvd, Pasadena, CA brecht@ist.caltech.edu

** Electrical Engineering, California Institute of Technology, 1200 E California Blvd, Pasadena, CA weiyu@systems.caltech.edu

†Electrical Engineering, California Institute of Technology, 1200 E California Blvd, Pasadena, CA bhassibi@systems.caltech.edu

978-1-4244-3124-3/08/$25.00 ©2008 IEEE

the equality constraints are of appropriate size. We provide numerical experiments demonstrating that even when matrix dimensions are small, the nuclear norm heuristic does indeed always recover the minimum rank solution when the number of constraints is sufficiently large.

II. NOTATION AND PRELIMINARIES

For a rectangular matrix $X \in \mathbb{R}^{n_1 \times n_2}$, X^* denotes the transpose of X. $\text{vec}(X)$ denotes the vector in $\mathbb{R}^{n_1 n_2}$ with the columns of X stacked on top of one and other.

$\sigma_i(X)$ denotes the i-th largest singular value of X and is equal to the square-root of the i-th largest eigenvalue of XX^*. The rank of X will usually be denoted by r, and is equal to the number of nonzero singular values. For matrices X and Y of the same dimensions, we define the inner product in $\mathbb{R}^{n_1 \times n_2}$ as $\langle X, Y \rangle := \text{trace}(X^*Y) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} X_{ij} Y_{ij}$. The norm associated with this inner product is called the Frobenius (or Hilbert-Schmidt) norm $||\cdot||_F$. The Frobenius norm is also equal to the Euclidean, or ℓ_2, norm of the vector of singular values, i.e.,

$$\|X\|_F := \left(\sum_{i=1}^{r} \sigma_i^2 \right)^{\frac{1}{2}} = \sqrt{\langle X, X \rangle} = \left(\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} X_{ij}^2 \right)^{\frac{1}{2}}.$$

The operator norm (or induced 2-norm) of a matrix is equal to its largest singular value (i.e., the ℓ_∞ norm of the singular values):

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

The nuclear norm of a matrix is equal to the sum of its singular values, i.e.,

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

and is alternatively known by several other names including the Schatten 1-norm, the Ky Fan r-norm, and the trace class norm. These three norms are related by the following inequalities which hold for any matrix X of rank at most r:

$$||X|| \leq ||X||_F \leq ||X||_* \leq \sqrt{r}||X||_F \leq r||X||. \tag{1}$$

We also state the following easily verified fact that will be used extensively throughout.

Lemma 2.1: Suppose X and Y are $n_1 \times n_2$ matrices such that $X^*Y = 0$ and $XY^* = 0$. Then $\|X+Y\|_* = \|X\|_* + \|Y\|_*$.

Indeed, if $X^*Y = 0$ and $XY^* = 0$, we can find a coordinate system in which

$$X = \left\| \begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix} \right\|_* \quad \text{and} \quad Y = \left\| \begin{bmatrix} 0 & 0 \\ 0 & B \end{bmatrix} \right\|_*$$

from which the lemma trivially follows.

III. MAIN RESULTS

Let X be an $n_1 \times n_2$ matrix decision variable. Without loss of generality, we will assume throughout that $n_1 \leq n_2$. Let $\mathscr{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ be a linear map, and let $b \in \mathbb{R}^m$. The main optimization problem under study is

$$\begin{array}{ll} \text{minimize} & \text{rank}(X) \\ \text{subject to} & \mathscr{A}(X) = b\,. \end{array} \tag{2}$$

As described in the introduction, our main concern is when the optimal solution of (2) coincides with the optimal solution of

$$\begin{array}{ll} \text{minimize} & \|X\|_* \\ \text{subject to} & \mathscr{A}(X) = b\,. \end{array} \tag{3}$$

Whenever $m < n_1 n_2$, the null space of $\mathscr{A}$, that is the set of Y such that $\mathscr{A}(Y) = 0$, is not empty. Note that X is an optimal solution for (3) if and only if for every Y in the null-space of $\mathscr{A}$

$$\|X+Y\|_* \geq \|X\|_*\,. \tag{4}$$

The following theorem generalizes this null-space criterion to a critical property that guarantees when the nuclear norm heuristic finds the minimum rank solution of $\mathscr{A}(X) = b$ for all values of the vector b. Our main result is the following

Theorem 3.1: Let X_0 be the optimal solution of (2) and assume that X_0 has rank $r < n_1/2$. Then

1) If for every Y in the null space of $\mathscr{A}$ and for every decomposition
$$Y = Y_1 + Y_2,$$
where Y_1 has rank r and Y_2 has rank greater than r, it holds that
$$\|Y_1\|_* < \|Y_2\|_*,$$
then X_0 is the unique minimizer of (3).
2) Conversely, if the condition of part 1 does not hold, then there exists a vector $b \in \mathbb{R}^m$ such that the minimum rank solution of $\mathscr{A}(X) = b$ has rank at most r and is not equal to the minimum nuclear norm solution.

This result is of interest for multiple reasons. First, as shown in Section V, many of the rank minimization problems of interest to the controls community can be written in the form of (2). To be precise, we have the following

Theorem 3.2: Let $\mathscr{A}$ be a linear map of $a \times b$ matrices into $\mathbb{R}^c$ and $\mathscr{C}$ maps $a \times b$ matrices into symmetric $d \times d$ matrices. Then the LMI constrained rank minimization problem

$$\begin{array}{ll} \text{minimize} & \text{rank}(X) \\ \text{subject to} & \mathscr{A}(X) = b \\ & \mathscr{C}(X) \succeq 0 \end{array}$$

can be equivalently formulated as

$$\begin{array}{ll} \text{minimize} & \text{rank}(X) + \lambda\, \text{rank}(Z) \\ \text{subject to} & \mathscr{A}(X) = b \\ & Z = \begin{bmatrix} I_d & B \\ B^* & \mathscr{C}(X) \end{bmatrix} \end{array}$$

for any $\lambda > a$. Note that in this is a formulation with a $(a+2d) \times (b+2d)$ dimensional decision variable and a linear map into $c + 2d(a+b) + d^2 - \frac{d}{2}$ dimensions.

Secondly, in Section VI we present a distribution over instances of (2) where the conditions of Theorem 3.1 hold with overwhelming probability. Note that for a linear map $\mathscr{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$, we can always find an $m \times n_1 n_2$ matrix $\mathbf{A}$ such that

$$\mathscr{A}(X) = \mathbf{A}\, \text{vec} X\,. \tag{5}$$

In the case where $\mathbf{A}$ has entries sampled independently from a zero-mean, unit variance Gaussian distribution, then the null space characterization of theorem 3.1 holds with overwhelming probability provided m is large enough. The particular details describing the relationship between the dimensions of the decision variable, the rank of the optimal solution, and the number of equations are described in detail in Section VI.

IV. NECESSARY AND SUFFICIENT CONDITIONS

We first prove our necessary and sufficient condition for success of the nuclear norm heuristic. We will need the following technical lemma which allows us to exploit Lemma 2.1 in our proof.

Lemma 4.1: Let X be an $n_1 \times n_2$ matrix with rank $r < \frac{n_1}{2}$ and Y be an arbitrary $n_1 \times n_2$ matrix. Let P_X^c and P_X^r be the matrices that project onto the column and row spaces of X respectively. Then if $P_X^c Y P_X^r$ has full rank, Y can be decomposed as

$$Y = Y_1 + Y_2,$$

where Y_1 has rank r, and

$$\|X + Y_2\|_* = \|X\|_* + \|Y_2\|_*.$$

Proof: Without loss of generality, we can write X as

$$X = \begin{bmatrix} X_{11} & 0 \\ 0 & 0 \end{bmatrix},$$

where X_{11} is $r \times r$ and full rank. Accordingly, Y becomes

$$Y = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix},$$

where Y_{11} is full rank since $P_X^r Y P_X^c$ is. The decomposition is now clearly

$$Y = \underbrace{\begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{21}Y_{11}^{-1}Y_{12} \end{bmatrix}}_{Y_1} + \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & Y_{22} - Y_{21}Y_{11}^{-1}Y_{12} \end{bmatrix}}_{Y_2}.$$

That Y_1 has rank r follows from the fact that the rank of a block matrix is equal to the rank of a diagonal block plus the rank of its Schur complement (see, e.g., [10, §2.2]). That $\|X_1 + Y_2\|_* = \|X_1\|_* + \|Y_2\|*$ follows from Lemma 2.1. ■

We can now provide a proof of Theorem 3.1.

Proof: We begin by proving the converse. Assume the condition of part 1 is violated, i.e., there exists some Y, such that $\mathscr{A}(Y) = 0$, $Y = Y_1 + Y_2$, $\text{rank}(Y_2) > \text{rank}(Y_1) = r$, yet $\|Y_1\|_* > \|Y_2\|_*$. Now take $X_0 = Y_1$ and $b = \mathscr{A}(X_0)$. Clearly, $\mathscr{A}(-Y_2) = b$ (since Y is in the null space) and so we have found a matrix of higher rank, but lower nuclear norm.

For the other direction, assume the condition of part 1 holds. Now use Lemma 4.1 with $X = X_0$ and $Y = X_* - X_0$. That is, let P_X^c and P_X^r be the matrices that project onto the column and row spaces of X_0 respectively and assume that $P_{X_0}^c(X_* - X_0)P_{X_0}^r$ has full rank. Write $X_* - X_0 = Y_1 + Y_2$ where Y_1 has rank r and $\|X_0 + Y_2\|_* = \|X_0\|_* + \|Y_2\|_*$. Assume further that Y_2 has rank larger than r (recall $r < n/2$). We will consider the case where $P_{X_0}^c(X_* - X_0)P_{X_0}^r$ does not have full rank and/or Y_2 has rank less than or equal to r in the appendix. We now have:

$$\begin{aligned} \|X_*\|_* &= \|X_0 + X_* - X_0\|_* \\ &= \|X_0 + Y_1 + Y_2\|_* \\ &\geq \|X_0 + Y_2\|_* - \|Y_1\|_* \\ &= \|X_0\|_* + \|Y_2\|_* - \|Y_1\|_* \quad \text{by Lemma 4.1.} \end{aligned}$$

But $\mathscr{A}(Y_1 + Y_2) = 0$, so $\|Y_2\|_* - \|Y_1\|_*$ non-negative and therefore $\|X_*\|_* \geq \|X_0\|_*$. Since X_* is the minimum nuclear norm solution, implies that $X_0 = X_*$. ■

For the interested reader, the argument for the case where $P_{X_0}^r(X_* - X_0)P_{X_0}^c$ does not have full rank or Y_2 has rank less than or equal to r can be found in the appendix.

V. REDUCTION TO THE AFFINE CASE

The preceding result only analyzes the affine rank minimization problem and do not extend to the case of arbitrary convex constraints. However, the affine case is far more general than it appears at first glance. For example, we can again use the fact that the rank of a block symmetric matrix is equal to the rank of a diagonal block plus the rank of its Schur complement to cast any LMI in X as a rank constraint. Indeed, given $\mathscr{C}(X) \in \mathscr{S}^{d\times d}$, its positive semidefiniteness can be equivalently expressed through a rank constraint, since $\mathscr{C}(X) \succeq 0$ if and only if

$$\text{rank}\left(\begin{bmatrix} I_d & B \\ B^* & \mathscr{C}(X) \end{bmatrix}\right) = d$$

for some $B \in \mathbb{R}^{d\times d}$. That is, if there exist matrices X and B satisfying the equality above, then $f(X) = B^*B \succeq 0$. We can also impose the rank constraint $\text{rank}(\mathscr{C}(X)) \leq r$ by choosing B to be of size $r \times d$ and having I_r in the $(1,1)$ block. Certainly, this is not an efficient way to solve standard LMIs for which polynomial time algorithms already exist, but this example allows us to reformulate rank constrained LMIs as linearly constrained LMIs and may allow us to characterize for which LMIs the nuclear norm heuristic succeeds.

Consider the LMI constrained rank minimization problem

$$\begin{array}{ll} \text{minimize} & \text{rank}(X) \\ \text{subject to} & \mathscr{A}(X) = b \\ & C(X) \succeq 0 \end{array} \tag{6}$$

where X, the decision variable is an $a \times b$ matrix (without loss of generality, $a \leq b$), $\mathscr{A}$ is some linear map of $a \times b$ matrices into $\mathbb{R}^c$ and $\mathscr{C}$ maps $a \times b$ matrices into symmetric $d \times d$ matrices. We can reformulate this problem into affine form by noting that is equivalent to

$$\begin{array}{ll} \text{minimize} & \text{rank}(X) + \lambda\,\text{rank}(Z) \\ \text{subject to} & \mathscr{A}(X) = b \\ & Z = \begin{bmatrix} I_d & B \\ B^* & \mathscr{C}(X) \end{bmatrix} \end{array} \tag{7}$$

for any $\lambda > a$. Note that by dimension counting, the block diagonal decision variable in (7) is $(a+2d) \times (b+2d)$. Also, by constraint counting, we see that there are a total of $c + 2d(a+b) + d^2 - \frac{d}{2}$ equations needed to both specify $\mathscr{A}(X) = b$ and to define Z.

The equivalence can be seen as follows. Let p_1^* denote the optimum of (6) and p_2^* denote the optimum of (7). Certainly, for any optimal solution X of (6), we can factor $\mathscr{C}(X) = B^*B$ to construct a Z with rank d, implying $p_1^* + \lambda d \geq p_2^*$.

Conversely, let X_0 be an optimal solution for (6). Then there exists a $B_0 \in \mathbb{R}^{d\times d}$ such that

$$Z_0 := \begin{bmatrix} I_d & B_0 \\ B_0^* & \mathscr{C}(X_0) \end{bmatrix} \succeq 0$$

and has rank d. That is, the pair (X_0, Z_0) is feasible for (7). Let X_1 and Z_1 be feasible for (7) and suppose $\operatorname{rank}(Z_1) > d$. Then

$$\begin{aligned} \operatorname{rank}(X_1) + \lambda \operatorname{rank}(Z_1) &\geq (\operatorname{rank}(X_1) + \lambda) + \lambda \operatorname{rank}(Z_0) \\ &\geq \operatorname{rank}(X_0) + \lambda \operatorname{rank}(Z_0) \qquad (8) \\ &= p_1^* + \lambda d\,. \end{aligned}$$

Now, if (X_2, Z_2) was feasible for (7) and $\operatorname{rank}(Z_2) = d$, then X_2 would be feasible for (6) and would have rank at most p_1^*. Therefore (X_0, Z_0) is an optimal solution for (7). Note that if we have an upper bound on the rank of the optimal X for (6), then the same argument reveals that any λ greater than that *a priori* rank bound will also suffice to guarantee that (6) and (7) have the same optimal solutions.

Using this equivalence, we may apply the analysis tools developed here to determine if the minimum rank solution is found when using nuclear norm heuristic.

VI. PROBABILISTIC GENERATION OF CONSTRAINTS SATISFYING NULL-SPACE CHARACTERIZATION

We now present a family of random equality constraints under which the nuclear norm heuristic succeeds with overwhelming probability. For simplicity of notation in the theorem statements, we consider the case of square matrices. These results can be then translated into rectangular matrices by padding with rows/columns of zeros to make the matrix square. We define the random ensemble of $d_1 \times d_2$ matrices $\mathfrak{G}(d_1, d_2)$ to be the Gaussian ensemble, with each entry sampled i.i.d. from a Gaussian distribution with zero-mean and variance one. We also denote $\mathfrak{G}(d,d)$ by $\mathfrak{G}(d)$.

The first result characterizes when a particular low-rank matrix can be recovered from a random linear system via nuclear norm minimization.

Theorem 6.1 (Weak Bound): Let X_0 be an $n \times n$ matrix of rank $r = \beta n$. Let $\mathscr{A} : \mathbb{R}^{n\times n} \to \mathbb{R}^{\mu n^2}$ denote the random linear transformation

$$\mathscr{A}(X) = \mathbf{A}\operatorname{vec}(X)$$

where $\mathbf{A}$ is sampled from $\mathfrak{G}(\mu n^2, n^2)$. Then whenever

$$\mu \geq 1 - \frac{64}{9\pi^2}\left((1-\beta)^{3/2} - \beta^{3/2}\right)^2 \qquad (9)$$

there exists a numerical constant $c_w(\mu,\beta) > 0$ such that with probability exceeding $1 - e^{-c_w(\mu,\beta)n^2}$,

$$X_0 = \arg\min\{\|Z\|_* \ : \ \mathscr{A}(Z) = \mathscr{A}(X_0)\}\,.$$

In particular, if β and μ satisfy (9), then nuclear norm minimization will recover X_0 from a random set of μn^2 constraints drawn from the Gaussian ensemble almost surely as $n \to \infty$.

The second theorem characterizes when the nuclear norm heuristic succeeds at recovering *all* low rank matrices.

Theorem 6.2 (Strong Bound): Let $\mathscr{A}$ be defined as in Theorem 6.1. Define the two functions

$$\begin{aligned} f(\beta,\varepsilon) &= \frac{8}{3\pi}\frac{(1-\beta)^{3/2} - \beta^{3/2} - 4\varepsilon}{1+4\varepsilon} \\ g(\beta,\varepsilon) &= \sqrt{2\beta(2-\beta)}\log\left(\frac{3\pi}{2\varepsilon}\right)\,. \end{aligned}$$

Then there exists a numerical constant $c_s(\mu,\beta) > 0$ such that with probability exceeding $1 - e^{-c_s(\mu,\beta)n^2}$, for all $n \times n$ matrices X_0 of rank $r \leq \beta n$,

$$X_0 = \arg\min\{\|Z\|_* \ : \ \mathscr{A}(Z) = \mathscr{A}(X_0)\}$$

whenever

$$\mu \geq 1 - \sup_{\substack{\varepsilon>0 \\ f(\beta,\varepsilon)-g(\beta,\varepsilon)>0}} \left(f(\beta,\varepsilon) - g(\beta,\varepsilon)\right)^2\,. \qquad (10)$$

In particular, if β and μ satisfy (9), then nuclear norm minimization will recover all rank r matrices from a random set of μn^2 constraints drawn from the Gaussian ensemble almost surely as $n \to \infty$.

The strategy of the proofs of these theorems is to show that $\mathscr{A}$ obeys the null-space criteria of Equation (4) and Theorem 3.1 respectively with overwhelming probability. The proofs can be found in the full version of this paper [15]. Noting that the null space of $\mathscr{A}$ is spanned by Gaussian vectors, we use bounds from probability on Banach Spaces to show that the sufficient conditions are met.

Figure 1 plots the bound from Theorems 6.1 and 6.2. We call (9) the *Weak Bound* because it is a condition that depends on the optimal solution of (2). On the other hand, we call (10) the *Strong Bound* as it guarantees the nuclear norm heuristic succeeds no matter what the optimal solution. The Weak Bound is the only bound that can be tested experimentally, and, in the next section, we will show that it corresponds well to experimental data. Moreover, the Weak Bound provides guaranteed recovery over a far larger region of (β,μ) parameter space. Nonetheless, the mere existence of a Strong Bound is surprising in of itself and results in a much better bound than what was available from previous results (c.f., [14]).

VII. NUMERICAL EXPERIMENTS

We now show that these asymptotic estimates hold even for small values of n. We conducted a series of experiments for a variety of the matrix sizes n, ranks r, and numbers of measurements m. As in the previous section, we let $\beta = \frac{r}{n}$ and $\mu = \frac{m}{n^2}$. For a fixed n, we constructed random recovery scenarios for low-rank $n \times n$ matrices. For each n, we varied μ between 0 and 1 where the matrix is completely determined. For a fixed n and μ, we generated all possible ranks such that $\beta(2-\beta) \leq \mu$. This cutoff was chosen because beyond that point there would be an infinite set of matrices of rank r satisfying the m equations.

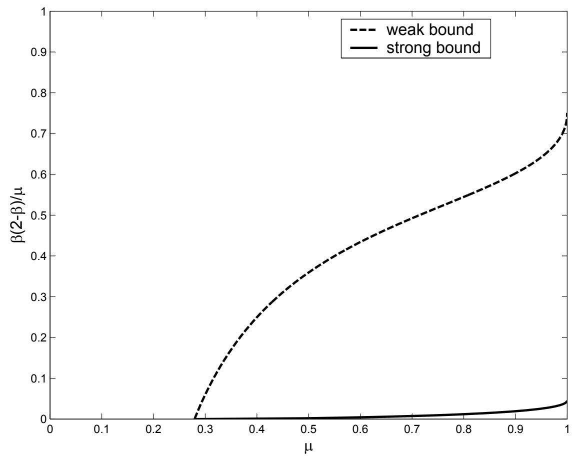


Fig. 1. The Weak Bound (9) versus the Strong Bound (10).

For each (n, μ, β) triple, we repeated the following procedure 10 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 m rows and n^2 columns. Then the nuclear norm minimization

$$\begin{array}{ll} \text{minimize} & \|X\|_* \\ \text{subject to} & \mathbf{A}\operatorname{vec} X = \mathbf{A}\operatorname{vec} Y_0 \end{array}$$

was solved using the freely available software SeDuMi [18] using the semidefinite programming formulation described in [14]. On a 2.0 GHz Laptop, each semidefinite program could be solved in less than two minutes for 40×40 dimensional X. We declared Y_0 to be recovered if $\|X - Y_0\|_F / \|Y_0\|_F < 10^{-3}$.

Figure 2 displays 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 10 runs (scaled between 0 and 1). White denotes perfect recovery in all experiments, and black denotes failure for all experiments. It is remarkable to note that not only are the plots very similar for $n = 30$ and $n = 40$, but that the Weak Bound falls completely within the white region and is an excellent approximation of the boundary between success and failure for large β.

VIII. CONCLUSIONS AND FUTURE WORK

We have presented a necessary and sufficient condition for the nuclear norm heuristic (and hence also the trace heuristic) to find the lowest rank solution of an affine set, and also shown how to reformulate LMI constrained rank minimization problems in the affine form. It would be interesting to directly formulate necessary and sufficient conditions for the LMI constrained rank minimization problem (6) that do not require such an affine reformulation. Along the same lines, it would be interesting to provide random instances of LMI constrained rank minimization problems that satisfy such necessary and sufficient conditions with high probability. Future work should also investigate if the probabilistic analysis that provides the bounds in Theorems 6.1 and 6.2 can be further tightened at all.

REFERENCES

[1] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, 2008, to Appear. Preprint available at http://dsp.rice.edu/cs/jlcs-v03.pdf.

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

[3] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

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

[5] 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.

[6] ——, "Sparse nonnegative solution of underdetermined linear equations by linear programming," *Proc. Natl. Acad. Sci. USA*, vol. 102, no. 27, pp. 9446–9451, 2005.

[7] 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.

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

[9] 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.

[10] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. New York: Cambridge University Press, 1991.

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

[12] 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.

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

[14] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum rank solutions of matrix equations via nuclear norm minimization," Submitted. Preprint Available at http://www.ist.caltech.edu/ brecht/publications.html.

[15] B. Recht, W. Xu, and B. Hassibi, "Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization," California Institute of Technology, Tech. Rep., 2008, preprint available at http://arxiv.org/abs/0809.1260.

[16] 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.

[17] M. Stojnic, W. Xu, and B. Hassibi, "Compressed sensing - probabilistic analysis of a null-space characterization," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2008.

[18] 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.

APPENDIX

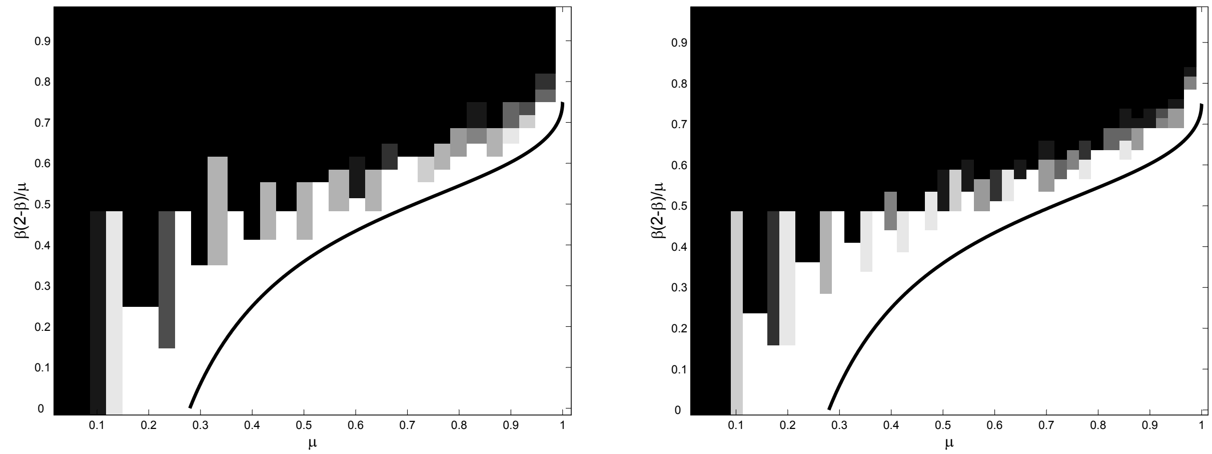
A. Rank-deficient case of Theorem 3.1

In an appropriate basis, we may write

$$X_0 = \begin{bmatrix} X_{11} & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad X_* - X_0 = Y = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}$$

If Y_{11} and $Y_{22} - Y_{21}Y_{11}^{-1}Y_{12}$ have full rank, then all our previous arguments apply. Thus, assume that at least one of them is not full rank. Nonetheless, it is always possible to find an *arbitrarily small* $\varepsilon > 0$ such that

$$Y_{11} + \varepsilon I \quad \text{and} \quad \begin{bmatrix} Y_{11} + \varepsilon I & Y_{12} \\ Y_{21} & Y_{22} + \varepsilon I \end{bmatrix}$$



(a) (b)

Fig. 2. Random rank recovery experiments for (a) $n = 30$ and (b) $n = 40$. The color of each cell reflects the empirical recovery rate. White denotes perfect recovery in all experiments, and black denotes failure for all experiments. In both frames, we plot the Weak Bound (9), showing that the predicted recovery regions are contained within the empirical regions, and the boundary between success and failure is well approximated for large values of β.

are full rank. This, of course, is equivalent to having $Y_{22} + \varepsilon I - Y_{21}(Y_{11}+\varepsilon I)^{-1}Y_{12}$ full rank. We can write

$$\begin{aligned}
\|X_*\|_* &= \|X_0 + X_* - X_0\|_* \\
&= \left\| \begin{bmatrix} X_{11} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \right\|_* \\
&\geq \left\| \begin{bmatrix} X_{11} - \varepsilon I & 0 \\ 0 & Y_{22} - Y_{21}(Y_{11}+\varepsilon I)^{-1}Y_{12} \end{bmatrix} \right\|_* \\
&\quad - \left\| \begin{bmatrix} Y_{11} + \varepsilon I & Y_{12} \\ Y_{21} & Y_{21}(Y_{11}+\varepsilon I)^{-1}Y_{12} \end{bmatrix} \right\|_* \\
&= \|X_{11} - \varepsilon I\|_* \\
&\quad + \left\| \begin{bmatrix} 0 & 0 \\ 0 & Y_{22} - Y_{21}(Y_{11}+\varepsilon I)^{-1}Y_{12} \end{bmatrix} \right\|_* \\
&\quad - \left\| \begin{bmatrix} Y_{11} + \varepsilon I & Y_{12} \\ Y_{21} & Y_{21}(Y_{11}+\varepsilon I)^{-1}Y_{12} \end{bmatrix} \right\|_* \\
&\geq \|X_0\|_* - r\varepsilon \\
&\quad + \left\| \begin{bmatrix} \varepsilon I - \varepsilon I & 0 \\ 0 & Y_{22} - Y_{21}(Y_{11}+\varepsilon I)^{-1}Y_{12} \end{bmatrix} \right\|_* \\
&\quad - \left\| \begin{bmatrix} Y_{11} + \varepsilon I & Y_{12} \\ Y_{21} & Y_{21}(Y_{11}+\varepsilon I)^{-1}Y_{12} \end{bmatrix} \right\|_* \\
&\geq \|X_0\|_* - 2r\varepsilon \\
&\quad + \left\| \begin{bmatrix} -\varepsilon I & 0 \\ 0 & Y_{22} - Y_{21}(Y_{11}+\varepsilon I)^{-1}Y_{12} \end{bmatrix} \right\|_* \\
&\quad - \left\| \begin{bmatrix} Y_{11} + \varepsilon I & Y_{12} \\ Y_{21} & Y_{21}(Y_{11}+\varepsilon I)^{-1}Y_{12} \end{bmatrix} \right\|_* \\
&\geq \|X_0\|_* - 2r\varepsilon,
\end{aligned}$$

where the last inequality follows from the condition of part 1 and noting that

$$\begin{aligned}
X_0 - X_* = &\begin{bmatrix} -\varepsilon I & 0 \\ 0 & Y_{22} - Y_{21}(Y_{11}+\varepsilon I)^{-1}Y_{12} \end{bmatrix} \\
&+ \begin{bmatrix} Y_{11} + \varepsilon I & Y_{12} \\ Y_{21} & Y_{21}(Y_{11}+\varepsilon I)^{-1}Y_{12} \end{bmatrix},
\end{aligned}$$

lies in the null space of $\mathscr{A}(\cdot)$ and the first matrix above has rank more than r. But, since ε can be arbitrarily small, this implies that $X_0 = X_*$.