Recovery of Low Rank Matrices Under Affine Constraints via a Smoothed Rank Function

Mohammadreza Malek-Mohammadi*, *Student, IEEE*, Massoud Babaie-Zadeh, *Senior Member, IEEE*, Arash Amini, and Christian Jutten, *Fellow, IEEE*

***Abstract*—In this paper, the problem of rank minimization of a matrix under affine constraints is addressed. The state of the art algorithms can recover matrices with a rank much less than what is sufficient for the uniqueness of the solution of this optimization problem. We propose an algorithm based on a smooth approximation of the rank function, which practically improves recoverable limits on the rank. In this method, termed SRF standing for Smoothed Rank Function, the rank function is replaced with a continuous and differentiable approximation. It starts with a rough approximation, and the quality of the approximation improves as the algorithm proceeds.**

On the theoretical side, benefiting from the spherical section property, we will show that the sequence of minimizers of approximating functions converges to the minimum rank solution. On the experimental side, it will be shown that SRF can recover matrices which are unique solutions of the rank minimization problem and yet not recoverable by nuclear norm minimization. Furthermore, it will be demonstrated that, in completing partially observed matrices, the accuracy of SRF is considerably and consistently better than some famous algorithms. In addition, SRF achieves a reduction in the computational cost by one order of magnitude when the number of revealed entries is close to the minimum number of parameters that uniquely represent a low rank matrix.

***Index Terms*—Affine Rank Minimization, Compressive Sensing, Matrix Completion (MC), Nuclear Norm Minimization (NNM), Rank Approximation, Spherical Section Property (SSP).**

I. Introduction

THERE are many applications in signal processing and control theory which involve finding a matrix with minimum rank subject to linear constraints [1]. This task is usually referred to as the affine rank minimization (RM) and includes *Matrix Completion* (MC) as a special case. In the latter, we are interested in reconstructing a low rank matrix from a subset of its entries. If the location of known entries follow certain random laws and the rank of the matrix is sufficiently small, one can uniquely recover the matrix with overwhelming probability [1]–[3].

This work has been supported in part by the Iran Telecommunication Research Center (ITRC) under contract No.: 500/11307.

M. Malek-Mohammadi and M. Babaie-Zadeh are with the Electrical Engineering Department, Sharif University of Technology, Tehran 1458889694, Iran (e-mail: m.rezamm@ieee.org; mbzadeh@yahoo.com).

A. Amini is with Biomedical Imaging Group (BIG), École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland (e-mail: arash.amini@epfl.ch).

C. Jutten is with the GIPSA-Lab, Department of Images and Signals, University of Grenoble and Institut Universitaire de France, France (e-mail: Christian.Jutten@inpg.fr).

One of the celebrated applications of affine rank minimization (or matrix completion) is *Collaborative Filtering* [2]. This technique is applied when a system tries to recommend goods to customers/users based on the available feedbacks of all the customers. In fact, the system learns the user preferences through the feedbacks and identifies similarities between them. As the number of factors affecting the user interests is much less than the total number of customers and products, the matrix whose (i, j) entry represents the rating of the i-th user for the j-th product is expected to be low-rank. This could be efficiently used by the matrix completion techniques to predict the users' ratings for unrated items.

Applications of rank minimization in control theory include *System Identification* [4] and low-order realization of linear systems [5]. In the former, the goal is to find an LTI system with minimum order that fits the available joint input-output observations of a multiple-input multiple-output system [6].

In Wireless Sensor Networks (WSN), due to limited energy resources and transmitting power, the sensors are able to communicate only with their neighboring sensors. These communications (*e.g.,* received powers) determine the pairwise distances between sensors, which partially reveals the matrix of all pairwise distances. To localize the sensors in the network, one needs to estimate their distances from predefined anchor points which in turn requires completion of the distance matrix through the Multi-Dimensional Scaling technique [7]. Interestingly, the rank of the pairwise distance matrix is small compared to its dimension [4].

Other areas to which affine rank minimization is applied include *Machine Learning* [8], *Quantum State Tomography* [9], *Spectrum Sensing* [10], and *Spatial Covariance* matrix completion [4], [11]. The spatial covariance matrix is essential in estimating the directions of arrival of sources impinging on an array of antenna using for example MUSIC [12] or ESPRIT algorithms [13].

The main difficulty of the rank minimization problem is due to the fact that the rank function is discontinuous and non-differentiable. Indeed, the optimization problem is NP-hard, and all available optimizers have doubly exponential complexity [14]. In [15], Fazel proposed to replace the rank of the matrix with its nuclear norm, which is defined as the sum of all singular values (SV). This modification is known to be the tightest convex relaxation of the rank minimization problem [4] and can be implemented using a Semi Definite Program (SDP) [1]. Using similar techniques as in compressed sensing, it is recently shown that under mild conditions and with overwhelming probability, the nuclear norm minimization

(NNM) technique achieves the same solution as the original rank minimization approach [16]–[18].

Other approaches towards rank minimization consist of either alternative solvers instead of SDP in NNM or approximating the rank function using other forms rather than the nuclear norm. The FPCA method belongs to the first category and uses fixed point and Bergman iterative algorithm to solve NNM [19]. Among the examples of the second category, one can name Accelerated Proximal Gradient (APG) [20] and singular value thresholding (SVT) [21]. Further, OptSpace, only applicable to the MC problem, is based on the trimming rows and columns of the incomplete matrix followed by truncation of some singular values of the trimmed matrix [22]. It is also possible to generalize the greedy methods of compressive sensing to the rank minimization problem; for instance, ADMiRA [23] generalizes the CoSaMP [24].

In this work, we introduce an iterative method that is based on approximating the rank function. However, in contrast to previous methods, the approximation is continuous and differentiable, and is made finer in each iteration. Asymptotically, the approximation will coincide with the rank function. It will be shown that this approach finds solutions which are not obtainable by NNM, while they are unique rank minimizers. Our method is inspired by the work of Mohimani et al [25] which uses smoothed ℓ_0-norm[1] to obtain sparse solutions of underdetermined system of linear equations. To generalize the method of [25] to the RM problem, we need to derive the gradient of the approximated rank function in closed form. This is, in fact, one of the contributions of this paper. We further establish analytical guarantees on the performance.

A few preliminary results of this work have been presented in the conference paper [26]. While [26] was only devoted to the matrix completion problem, the current paper focuses on the more general problem of affine rank minimization. Furthermore, here, we present mathematical and experimental convergence analysis, and consider more comprehensive numerical evaluation scenarios.

The reminder of this paper is organized as follows. In Section II, the affine RM problem is formulated, and in Section III, the SRF algorithm is introduced. Section IV is devoted to analyze the convergence properties of the SRF algorithm. Finally, in Section V, some experimental results of our algorithm are provided, and it will be compared empirically against a few well known algorithms, followed by conclusions.

II. Problem Formulation

The affine rank minimization problem generally is formulated as

$$\min_{\mathbf{X}} \operatorname{rank}(\mathbf{X}) \text{ subject to } \mathcal{A}(\mathbf{X}) = \mathbf{b}, \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$ is the decision variable, $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ is a known linear operator, and $\mathbf{b} \in \mathbb{R}^m$ is the observed measurement vector. The special case of matrix completion corresponds to the setting

$$\min_{\mathbf{X}} \operatorname{rank}(\mathbf{X}) \text{ subject to } [\mathbf{X}]_{ij} = [\mathbf{M}]_{ij} \quad \forall (i,j) \in \Omega, \tag{2}$$

where $\mathbf{X}$ is as in (1), $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$ is the matrix whose entries are partially observed, $\Omega \subset \{1, 2, ..., n_1\} \times \{1, 2, ..., n_2\}$ is the set of the indexes of the observed entries of $\mathbf{M}$, and $[\mathbf{X}]_{ij}$ is the (i,j)-th entry of $\mathbf{X}$. Indeed, the constraints $[\mathbf{X}]_{ij} = [\mathbf{M}]_{ij}, \forall (i,j) \in \Omega$ is an affine mapping which keeps some of the entries and discards others.

In the nuclear norm minimization, the rank function is replaced with the nuclear norm of the decision variable, leading to

$$\min_{\mathbf{X}} \|\mathbf{X}\|_* \text{ subject to } \mathcal{A}(\mathbf{X}) = \mathbf{b}, \tag{3}$$

where $\|\mathbf{X}\|_* \triangleq \sum_{i=1}^{r} \sigma_i(\mathbf{X})$ is the nuclear norm, in which r is the rank of the matrix $\mathbf{X}$, and $\sigma_i(\mathbf{X})$ is the i-th largest singular value of the matrix $\mathbf{X}$. There is a strong parallelism between this rank minimization and ℓ_0-norm minimization in compressive sensing [1]. In particular, minimizing the rank is equivalent to minimizing the number of non-zero singular values. Hence, (1) can be reformulated as

$$\min_{\mathbf{X}} \|\boldsymbol{\sigma}(\mathbf{X})\|_0 \text{ subject to } \mathcal{A}(\mathbf{X}) = \mathbf{b}, \tag{4}$$

where $\boldsymbol{\sigma}(\mathbf{X}) = (\sigma_1(\mathbf{X}), ..., \sigma_n(\mathbf{X}))^T$ is the vector of all singular values, $\|\cdot\|_0$ denotes the ℓ_0-norm, and $n = \min(n_1, n_2)$.[2] Likewise, the nuclear norm is the ℓ_1-norm of the singular value vector where ℓ_1-norm of a vector, denoted by $\|\cdot\|_1$, is the sum of the absolute values of its elements. This suggests the alternative form of

$$\min_{\mathbf{X}} \|\boldsymbol{\sigma}(\mathbf{X})\|_1 \text{ subject to } \mathcal{A}(\mathbf{X}) = \mathbf{b} \tag{5}$$

for (3). Based on this strong parallels, many results in compressive sensing theory (see for example [27]–[30]) have been adopted in the rank minimization problem [1], [16], [17], [31].

III. The proposed algorithm

A. The main idea

Our approach to solve the rank minimization problem is to approximate the rank with a continuous and differentiable function, and then to use a gradient descent algorithm to minimize it. The approximation is such that the error can be made arbitrarily small. Note that nuclear norm is not differentiable [32] and its approximation error depends on the singular values of the matrix and cannot be controlled.

Instead of using a fixed approximation, we use a family $G_\delta : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^+$ of approximations, where the index δ is a measure of approximation error and reflects the accuracy. The smaller δ, the closer behavior of G_δ to the rank. For instance, G_0 stands for the errorless approximation; *i.e.*, G_0 coincides with the rank function. We constrain the family to be continuous with respect to δ. This helps in achieving the rank minimizer (G_0) by gradually decreasing δ. Besides, to facilitate finding the minimizers of the relaxed problem, we

[1] ℓ_0-norm, not mathematically a vector norm, denotes the number of non-zero elements of a vector.

[2] Note that just r entries of $\boldsymbol{\sigma}(\mathbf{X})$ are non-zero where r is the rank of the matrix $\mathbf{X}$.

require the G_δ's for $\delta > 0$ to be differentiable with respect to the input matrix.

In order to introduce suitable G_δ families, we specify certain families of one-dimensional functions that approximate Kronecker delta function.

Assumption 1: Let $f : \mathbb{R} \to [0,1]$ and define $f_\delta(x) = f(x/\delta)$ for all $\delta > 0$. The class $\{f_\delta\}$ is said to satisfy the Assumption 1, if

(a) f is real, symmetric unimodal, and analytic,
(b) $f(x) = 1 \Leftrightarrow x = 0$,
(c) $f''(0) < 0$, and
(d) $\lim_{|x|\to\infty} f(x) = 0$.

It follows from Assumption 1 that $x = 0$ is the unique mode of all f_δ's. This implies that $f'_\delta(0) = 0$ for $\delta \neq 0$. In addition, $\{f_\delta\}$ converge pointwise to Kronecker delta function as $\delta \to 0$, *i.e.*,

$$\lim_{\delta\to 0} f_\delta(x) = \begin{cases} 0 & \text{if } x \neq 0, \\ 1 & \text{if } x = 0. \end{cases} \tag{6}$$

The class of Gaussian functions, which is of special interest in this paper, is defined as

$$f_\delta(x) = \exp(-\frac{x^2}{2\delta^2}). \tag{7}$$

It is not difficult to verify the constraints of Assumption 1 for this class. Other examples include $f_\delta(x) = 1 - \tanh(\frac{x^2}{2\delta^2})$ and $f_\delta(x) = \frac{\delta^2}{x^2+\delta^2}$.

To extend the domain of $\{f_\delta\}$ to matrices, let define

$$F_\delta(\mathbf{X}) = h_\delta\big(\boldsymbol{\sigma}(\mathbf{X})\big) = \sum_{i=1}^{n} f_\delta\big(\sigma_i(\mathbf{X})\big), \tag{8}$$

where $n = \min(n_1, n_2)$, and $h_\delta : \mathbb{R}^n \to \mathbb{R}$ is defined as $h_\delta(\mathbf{x}) = \sum_{i=1}^n f_\delta(x_i)$. Since f_δ is an approximate Kronecker delta function, $F_\delta(\mathbf{X})$ yields an estimate of the number of zero singular values of $\mathbf{X}$. Consequently, it can be concluded that $\operatorname{rank}(\mathbf{X}) \approx n - F_\delta(\mathbf{X})$, and the RM problem can be relaxed to

$$\min_{\mathbf{X}} \big(G_\delta(\mathbf{X}) = n - F_\delta(\mathbf{X})\big) \text{ subject to } \mathcal{A}(\mathbf{X}) = \mathbf{b}, \tag{9}$$

or equivalently

$$\max_{\mathbf{X}} F_\delta(\mathbf{X}) \text{ subject to } \mathcal{A}(\mathbf{X}) = \mathbf{b}. \tag{10}$$

The advantage of maximizing F_δ compared to minimizing the rank is that F_δ is smooth and we can apply gradient methods. However, for small values of δ where G_δ is a relatively good approximate of the rank function, F_δ has many local maxima, which are likely to trap gradient methods.

To avoid local maxima, we initially apply a large δ. Indeed, we will show in Theorem 2 that under Assumption 1, F_δ becomes convex as $\delta \to \infty$, and (10) will have a unique solution. Then, we gradually decrease δ to improve the accuracy of approximation. For each new value of δ, we initialize the maximization of F_δ with the result of (10) for the previous value of δ. From the continuity of $\{f_\delta\}$ with respect to δ, it is expected that the solutions of (10) for δ_i and δ_{i+1} are close, when δ_i and δ_{i+1} are close. In this fashion, the chance of finding a local maximum instead of a global one is decreased. This approach for optimizing non-convex functions is known as Graduated Non-Convexity (GNC) [33], and was used in [25] to minimize functions approximating the ℓ_0-norm.

B. Gradient Projection

For each δ in the decreasing sequence, to maximize F_δ with equality constraints, we use the 'Gradient Projection' (GP) technique [34]. In GP, the search path in each iteration is obtained by projecting back the ascent (or descent) direction onto the feasible set [34]. In other words, at each iteration, one has $\mathbf{X} \leftarrow \mathcal{P}\big(\mathbf{X} + \mu_j \nabla F_\delta(\mathbf{X})\big)$, where $\mathcal{P}$ denotes the orthogonal projection onto the affine set defined by linear constraints $\mathcal{A}(\mathbf{X}) = \mathbf{b}$, and μ_j is the step size of the j-th iteration. As the feasible set is affine, several methods can be exploited to implement the projection $\mathcal{P}$. For example, one can store the QR factorization of the matrix implementation of $\mathcal{A}$ for fast implementation of the back projection, or alternatively, a least squares problem can be solved at each step [1]. The closed form solution of the least squares problem can be found in Appendix A.

To complete the GP step, we should derive the gradient of the approximating functions with respect to the matrix $\mathbf{X}$. Surprisingly, *although $\sigma_i(\mathbf{X}), i = 1, ..., n$ and $\|\mathbf{X}\|_*$ are not differentiable functions of $\mathbf{X}$ [32], the following theorem shows that one can find functions $F_\delta = h_\delta \circ \boldsymbol{\sigma}(\mathbf{X})$ which are differentiable under the absolutely symmetricity of the h_δ.* Before stating the theorem, recall that a function $f : \mathbb{R}^q \to [-\infty, +\infty]$ is called *absolutely symmetric* [35] if $f(\mathbf{x})$ is invariant under arbitrary permutations and sign changes of the components of $\mathbf{x}$.

Theorem 1: Suppose that $F_\delta : \mathbb{R}^{n_1\times n_2} \to \mathbb{R}$ is represented as $F_\delta(\mathbf{X}) = h_\delta\big(\boldsymbol{\sigma}(\mathbf{X})\big) = h_\delta \circ \boldsymbol{\sigma}(\mathbf{X})$, where $\mathbf{X} \in \mathbb{R}^{n_1\times n_2}$ with the Singular Value Decomposition (SVD) $\mathbf{X} = \mathbf{U}\operatorname{diag}(\sigma_1, ..., \sigma_n)\mathbf{V}^T$, $\boldsymbol{\sigma}(\mathbf{X}) : \mathbb{R}^{n_1\times n_2} \to \mathbb{R}^n$ has the SVs of the matrix $\mathbf{X}$, $n = \min(n_1, n_2)$, and $h_\delta : \mathbb{R}^n \to \mathbb{R}$ is absolutely symmetric. Then the gradient of $F_\delta(\mathbf{X})$ at $\mathbf{X}$ is

$$\frac{\partial F_\delta(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{U}\operatorname{diag}(\boldsymbol{\theta})\mathbf{V}^T, \tag{11}$$

where $\boldsymbol{\theta} = \frac{\partial h_\delta(\mathbf{y})}{\partial \mathbf{y}}|_{\mathbf{y}=\boldsymbol{\sigma}(\mathbf{X})}$ denotes the gradient of h_δ at $\boldsymbol{\sigma}(\mathbf{X})$.

Informal Proof: In [35, Cor. 2.5], it is shown that if a function h_δ is absolutely symmetric and the matrix $\mathbf{X}$ has $\boldsymbol{\sigma}(\mathbf{X})$ in the domain of h_δ, then the subdifferential[3] of F_δ is given by

$$\partial\big(h_\delta \circ \boldsymbol{\sigma}(\mathbf{X})\big) = \{\mathbf{U}\operatorname{diag}(\boldsymbol{\theta})\mathbf{V}^T | \boldsymbol{\theta} \in \partial h_\delta\big(\boldsymbol{\sigma}(\mathbf{X})\big)\}. \tag{12}$$

Since $h_\delta\big(\boldsymbol{\sigma}(\mathbf{X})\big) = \sum_{i=1}^n f_\delta\big(\sigma_i(\mathbf{X})\big)$ is differentiable at $\boldsymbol{\sigma}(\mathbf{X})$, $\partial h_\delta\big(\boldsymbol{\sigma}(\mathbf{X})\big)$ is a singleton and consequently $\partial\big(h_\delta \circ \boldsymbol{\sigma}(\mathbf{X})\big)$ becomes a singleton. When the subdifferential of a non-convex function becomes singleton, the function is intuitively expected to be differentiable with the subgradient as its gradient.[4] Nevertheless, to the best of our knowledge,

[3]To see the definition of subdifferential and subgradient of non-convex functions, refer to [36, Sec. 3].

[4]For a convex function, the subdifferential is singleton iff the function is differentiable [37].

there is no formal proof. Provided that this intuition is true, then $\partial\big(h_\delta \circ \boldsymbol{\sigma}(\mathbf{X})\big)$ will be converted to $\nabla\big(h_\delta \circ \boldsymbol{\sigma}(\mathbf{X})\big)$ and equation (11) is obtained. That is why we called this proof informal.

Formal proof: Equation (11) can be obtained directly from the "if part" of [35, Thm. 3.1], which does not require convexity of h_δ as stated in its proof. ■

Corollary 1: For the Gaussian function family given in (7), the gradient of $F_\delta(\mathbf{X})$ at $\mathbf{X}$ is

$$\frac{\partial F_\delta(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{U}\text{diag}(-\frac{\sigma_1}{\delta^2}\mathrm{e}^{-\sigma_1^2/2\delta^2}, ..., -\frac{\sigma_n}{\delta^2}\mathrm{e}^{-\sigma_n^2/2\delta^2})\mathbf{V}^T. \tag{13}$$

Proof: f_δ is an even function for the Gaussian family; therefore, h_δ becomes an absolutely symmetric function. As a result, Theorem 1 proves (13). ■

C. Initialization

Naturally, we initialize the GNC procedure by the solution of (10) corresponding to $\delta \to \infty$. This solution can be found from the following theorem.

Theorem 2: Consider a class of one variable functions $\{f_\delta\}$ satisfying the Assumption 1. For the rank approximation problem (10), let $\widehat{\mathbf{X}} = \arg\min_{\mathcal{A}(\mathbf{X})=\mathbf{b}} \|\mathbf{X}\|_F$, then

$$\lim_{\delta\to\infty} \arg\max_{\mathcal{A}(\mathbf{X})=\mathbf{b}} F_\delta(\mathbf{X}) = \widehat{\mathbf{X}}, \tag{14}$$

where $\|\cdot\|_F$ denotes the matrix Frobenius norm.

There is a simple interpretation of the solution of (10) for the Gaussian family when δ approaches ∞. As $e^{-x} \approx 1 - x$ for small values of x,

$$F_\delta(\mathbf{X}) = \sum_{i=1}^{n} e^{-\sigma_i^2(\mathbf{X})/2\delta^2} \approx n - \sum_{i=1}^{n} \sigma_i^2(\mathbf{X})/\delta^2$$

for $\delta \gg \sigma_i(\mathbf{X})$. Consequently,

$$\arg\max_{\mathcal{A}(\mathbf{X})=\mathbf{b}} F_\delta(\mathbf{X}) \approx \arg\min_{\mathcal{A}(\mathbf{X})=\mathbf{b}} \sum_{i=1}^{n} \sigma_i^2(\mathbf{X}) = \arg\min_{\mathcal{A}(\mathbf{X})=\mathbf{b}} \|\mathbf{X}\|_F.$$

The proof is left to Appendix B.

The following Corollary is an immediate result of the above theorem.

Corollary 2: For the matrix completion problem, the initial solution of the SRF algorithm is $\widehat{\mathbf{X}}$ with the following definition:

$$[\widehat{\mathbf{X}}]_{ij} = \begin{cases} [\mathbf{M}]_{ij} & (i,j) \in \Omega, \\ 0 & (i,j) \notin \Omega. \end{cases}, \tag{15}$$

where $\mathbf{M}$ and Ω are as defined in (2).

D. The Final Algorithm

The final algorithm is obtained by applying the main idea, initial solution, and gradient projection to the Gaussian function given in (7). Fig. 1 depicts the algorithm. In the sequel, we briefly review some remarks about the parameters used in the implementation of the algorithm. Most of these remarks correspond to similar remarks for SL0 algorithm [25], and are presented here for the sake of completeness.

- Initialization:
  1) Let $\widehat{\mathbf{X}}_0 = \arg\min \|\mathbf{X}\|_F$ s.t. $\mathcal{A}(\mathbf{X}) = \mathbf{b}$ as the initial solution.
  2) Choose a suitable decreasing sequence of δ, $\{\delta_1, \delta_2, \ldots\}$; e.g., $\delta_j = c\delta_{j-1}, j \geq 2$.
  3) Choose ϵ as the stopping threshold.
  4) Choose suitable L (Number of internal loop iteration) and μ, and initialize j with 1.
- While $d > \epsilon$
  1) Let $\delta = \delta_j$.
  2) Internal maximization Loop:
     - Initialization: $\mathbf{X} = \widehat{\mathbf{X}}_{j-1}$.
     - For $\ell = 1 \ldots L$ (loop L times):
       a) Compute the SVD of
       $$\mathbf{X} = \mathbf{U}\text{diag}(\sigma_1, \ldots, \sigma_n)\mathbf{V}^T.$$
       b) Let
       $$\mathbf{D} = \mathbf{U}\text{diag}(-\sigma_1 e^{-\sigma_1^2/2\delta^2}, \ldots, -\sigma_n e^{-\sigma_n^2/2\delta^2})\mathbf{V}^T.$$
       c) $\mathbf{X} \leftarrow \mathbf{X} + \mu\mathbf{D}$.
       d) Project $\mathbf{X}$ back onto the feasible set:
       $$\mathbf{X} \leftarrow \mathcal{P}(\mathbf{X}).$$
  3) Set $\widehat{\mathbf{X}}_j = \mathbf{X}$.
  4) $d = \|\widehat{\mathbf{X}}_j - \widehat{\mathbf{X}}_{j-1}\|_F / \sqrt{n_1 n_2}$.
  5) $j \leftarrow j + 1$.
- Final answer is $\widehat{\mathbf{X}} = \widehat{\mathbf{X}}_j$.

Fig. 1. The SRF Algorithm.

Remark 1. It is not necessary to wait for the convergence of the internal steepest ascent loop because as explained in Section III-A for each value of δ, it is just needed to get close to the global maximizer of F_δ to avoid local maxima. Therefore, the internal loop is only repeated for a fixed number of times (L).

Remark 2. After initiating the algorithm with the minimum Frobenius norm, the first value of δ may be set to about two to four times of the largest SV of $\widehat{\mathbf{X}}_0$ (the initial guess). If we take $\delta > 4\max\big(\sigma_i(\widehat{\mathbf{X}})\big)$, then $\exp\big(-\sigma_i^2(\widehat{\mathbf{X}})/2\delta^2\big) > 0.96 \approx 1$ for $1 \leq i \leq n$. Thus, this δ value acts virtually like ∞ for all SVs of $\widehat{\mathbf{X}}_0$. Finally, the decreasing sequence can be adjusted to $\delta_j = c\delta_{j-1}, j \geq 2$, where c generally is chosen between 0.5 and 1.

Remark 3. This remark is devoted to the selection of μ_j, step–size parameter. Typically, in a gradient ascent algorithm, μ_j should be chosen small enough to follow the ascent direction. Furthermore, reducing δ results in more fluctuating behaviour of the rank approximating function. Therefore, to avoid large steps which cause jumps over the maximizer, one should choose smaller value of step–size for smaller values of δ. Following the same reasoning as in [25, Remark 2], a good choice is to decrease μ_j proportional to δ^2, that is, $\mu_j = \mu\delta^2$, where μ is a constant. By letting $\mu_j = \mu\delta^2$, the gradient step can be reduced to

$$\mathbf{X}_j \leftarrow \mathbf{X}_j - \mu\mathbf{U}\text{diag}(\sigma_1 e^{-\sigma_1^2/2\delta^2}, \ldots, \sigma_n e^{-\sigma_n^2/2\delta^2})\mathbf{V}^T.$$

Remark 4. The distance between the solutions at the two

consecutive iterations is the criterion to stop the algorithm. That is, if $d \triangleq \|\widehat{\mathbf{X}}_j - \widehat{\mathbf{X}}_{j-1}\|_F/\sqrt{n_1 n_2}$ is smaller than some tolerance (ϵ), the iterations are ended and $\widehat{\mathbf{X}}_j$ becomes the final solution.

IV. Convergence Analysis

The SRF algorithm consists of two loops. In the internal loop, F_δ is maximized for the current value of δ, and in the external loop, δ is decreased to improve the approximation. Noting that the original problem is NP-Hard and we are dealing with non-convex functions, a complete convergence analysis would be probably very tricky, and is not addressed in this paper. Instead, in the sequel, it is assumed that the internal loop has been converged to the global maximum, and we prove that the final solution converges to the minimum rank solution as δ goes to zero. In other words, it will be shown that the sequence of global maximizers converges to the lowest rank solution. The following results and proofs are not direct extension of the convergence results of [25] and are highly more tricky to obtain, though our exposition follows the same line of presentation.

We start the convergence analysis by the definition of Spherical Section Property (SSP), used in the analysis of uniqueness of the rank and nuclear norm minimization [16], and a lemma which makes this abstract definition clearer.

Definition 1: **Spherical Section Property** [16], [38]. The spherical section constant of a linear operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ is defined as

$$\Delta(\mathcal{A}) = \min_{\mathbf{Z} \in \mathrm{null}(\mathcal{A}) \setminus \{\mathbf{0}\}} \frac{\|\mathbf{Z}\|_*^2}{\|\mathbf{Z}\|_F^2}. \tag{16}$$

Further, $\mathcal{A}$ is said to have the Δ-spherical section property if $\Delta(\mathcal{A}) \geq \Delta$.

Lemma 1: Assume $\mathcal{A}$ has the Δ-spherical section property. Then for any $\mathbf{X} \in \mathrm{null}(\mathcal{A}) \setminus \mathbf{0}$, we have $\mathrm{rank}(\mathbf{X}) \geq \Delta$.

Proof: Since $\mathbf{X}$ belongs to $\mathrm{null}(\mathcal{A})$, one can write:

$$\frac{\|\mathbf{X}\|_*}{\|\mathbf{X}\|_F} \geq \sqrt{\Delta} \Rightarrow \|\mathbf{X}\|_* \geq \sqrt{\Delta}\|\mathbf{X}\|_F. \tag{17}$$

It is also known that $\sqrt{\mathrm{rank}(\mathbf{X})}\|\mathbf{X}\|_F \geq \|\mathbf{X}\|_*$, see for example [38]. Putting them together, we have $\|\mathbf{X}\|_* \geq \sqrt{\Delta}\frac{\|\mathbf{X}\|_*}{\sqrt{\mathrm{rank}(\mathbf{X})}} \Rightarrow \mathrm{rank}(\mathbf{X}) \geq \Delta$ or $\mathrm{rank}(\mathbf{X}) \geq \lceil \Delta \rceil$, where $\lceil \Delta \rceil$ denotes the smallest integer greater than or equal to Δ. ∎

The above lemma shows that if Δ is large, the null space of $\mathcal{A}$ does not include low rank matrices. Such subspaces are also known as almost Euclidean subspaces [30], in which the ratio of ℓ_1-norm to ℓ_2-norm of elements cannot be small.

Theorem 3 ([38]): Suppose $\mathcal{A}$ has the Δ-spherical property, and $\mathbf{X}_0 \in \mathbb{R}^{n_1 \times n_2}$ satisfies $\mathcal{A}(\mathbf{X}_0) = \mathbf{b}$. If $\mathrm{rank}(\mathbf{X}_0) < \frac{\Delta}{2}$, then $\mathbf{X}_0$ is the unique solution of problem (1).

Lemma 2: Assume $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ has Δ-spherical section property, and set $n = \min(n_1, n_2)$. Let $\mathbf{X}$ be any element in $\mathrm{null}(\mathcal{A})$ and $(\sigma_1, ..., \sigma_n)$ represent its singular values. Then for any subset $\mathcal{I}$ of $\{1, ..., n\}$ such that $|\mathcal{I}| + \Delta > n$,

$$\frac{\sum_{i \in \mathcal{I}} \sigma_i}{(\sum_{i=1}^n \sigma_i^2)^{0.5}} \geq \sqrt{\Delta} - \sqrt{n - |\mathcal{I}|}, \tag{18}$$

where $|\cdot|$ denotes the cardinality of a set.

Proof: If $\mathcal{I} = \{1, ..., n\}$, then it is clear that $\frac{\sum_{i=1}^n \sigma_i}{(\sum_{i=1}^n \sigma_i^2)^{0.5}} \geq \sqrt{\Delta}$, since the Δ-spherical section property holds. Otherwise, if $|\mathcal{I}| < n$, the Δ-spherical section property implies that

$$\sqrt{\Delta} \leq \frac{\|\mathbf{X}\|_*}{\|\mathbf{X}\|_F} = \frac{\sum_{i=1}^n \sigma_i}{(\sum_{i=1}^n \sigma_i^2)^{0.5}}. \tag{19}$$

For the sake of simplicity, let us define

$$\alpha_i = \frac{\sigma_i}{(\sum_{i=1}^n \sigma_i^2)^{0.5}}. \tag{20}$$

This shows that

$$1 = \sum_{i=1}^n \alpha_i^2 \geq \sum_{i \notin \mathcal{I}} \alpha_i^2 \geq \frac{(\sum_{i \notin \mathcal{I}} \alpha_i)^2}{n - |\mathcal{I}|}, \tag{21}$$

where we used the inequality $\forall \mathbf{z} \in \mathbb{R}^p, \|\mathbf{z}\|_1^2 \leq p\|\mathbf{z}\|_2^2$. Hence, it can be concluded that

$$\sum_{i \notin \mathcal{I}} \alpha_i \leq \sqrt{n - |\mathcal{I}|}. \tag{22}$$

On the other hand, it is known that

$$\sqrt{\Delta} \leq \sum_{i \in \mathcal{I}} \alpha_i + \sum_{i \notin \mathcal{I}} \alpha_i \leq \sum_{i \in \mathcal{I}} \alpha_i + \sqrt{n - |\mathcal{I}|}, \tag{23}$$

which confirms that

$$\frac{\sum_{i \in \mathcal{I}} \sigma_i}{(\sum_{i=1}^n \sigma_i^2)^{0.5}} = \sum_{i \in \mathcal{I}} \alpha_i \geq \sqrt{\Delta} - \sqrt{n - |\mathcal{I}|}. \tag{24}$$

∎

Corollary 3: If $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ has Δ-spherical section property, $n = \min(n_1, n_2)$, and $\mathbf{X} \in \mathrm{null}(\mathcal{A})$ has at most $\lceil \Delta - 1 \rceil$ singular values greater than α, then

$$\|\mathbf{X}\|_F \leq \frac{n\alpha}{\sqrt{\Delta} - \sqrt{\lceil \Delta - 1 \rceil}}. \tag{25}$$

Proof: At least $n - \lceil \Delta - 1 \rceil$ singular values of $\mathbf{X}$ are less than or equal to α. If $\mathcal{I}$ denotes the indices of singular values not greater than α, then by using Lemma 2, we will have

$$\frac{\sum_{i \in \mathcal{I}} \sigma_i}{(\sum_{i=1}^n \sigma_i^2)^{0.5}} \geq \sqrt{\Delta} - \sqrt{n - n + \lceil \Delta - 1 \rceil} \Rightarrow \tag{26}$$

$$\|\mathbf{X}\|_F(\sqrt{\Delta} - \sqrt{\lceil \Delta - 1 \rceil}) \leq \sum_{i \in \mathcal{I}} \sigma_i \leq n\alpha, \tag{27}$$

which proves that

$$\|\mathbf{X}\|_F \leq \frac{n\alpha}{\sqrt{\Delta} - \sqrt{\lceil \Delta - 1 \rceil}}. \tag{28}$$

∎

Lemma 3: Assume $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ has Δ-spherical section property, $f_\delta(\cdot)$ is a member of the class that satisfies Assumption 1, and define F_δ as in (8) and $n = \min(n_1, n_2)$. Let $\mathcal{X} = \{\mathbf{X} | \mathcal{A}(\mathbf{X}) = \mathbf{b}\}$ contain a solution $\mathbf{X}_0$ with $\mathrm{rank}(\mathbf{X}_0) = r_0 < \frac{\Delta}{2}$. Then for any $\widehat{\mathbf{X}} \in \mathcal{X}$ that satisfies

$$F_\delta(\widehat{\mathbf{X}}) \geq n - (\lceil \Delta - 1 \rceil - r_0), \tag{29}$$

we have that

$$\|\mathbf{X}_0 - \widehat{\mathbf{X}}\|_F \leq \frac{n\alpha_\delta}{\sqrt{\Delta} - \sqrt{\lceil \Delta - 1 \rceil}}, \tag{30}$$

where $\alpha_\delta = \left|f_\delta^{-1}(\frac{1}{n})\right|$.

Proof: First, note that due to Assumption 1, $f_\delta(x)$ takes all the values in $]0,1[$ exactly twice; once with a positive x and once with a negative one. Because of the symmetry, the two have the same modulus; therefore, α_δ is well-defined.

Let us denote the singular values of $\mathbf{X}_0$ and $\widehat{\mathbf{X}}$ by $\sigma_1 \geq \cdots \geq \sigma_n$ and $\hat{\sigma}_1 \geq \cdots \geq \hat{\sigma}_n$, respectively. Define $\mathcal{I}_\alpha$ as the set of indices i for which $\hat{\sigma}_i > \alpha$. Now, we have that

$$\begin{aligned} F_\delta(\widehat{\mathbf{X}}) &= \sum_{i=1}^{n} f_\delta(\hat{\sigma}_i) && (31)\\ &= \underbrace{\sum_{i\in\mathcal{I}_\alpha} \underbrace{f_\delta(\hat{\sigma}_i)}_{<\frac{1}{n}}}_{<n\frac{1}{n}=1} + \underbrace{\sum_{i\notin\mathcal{I}_\alpha} \underbrace{f_\delta(\hat{\sigma}_i)}_{\leq 1}}_{\leq n-|\mathcal{I}_\alpha|} && (32)\\ &< n - |\mathcal{I}_\alpha| + 1. \end{aligned}$$

On the other hand, $F_\delta(\widehat{\mathbf{X}}) \geq n - (\lceil\Delta - 1\rceil - r_0)$; therefore,

$$\begin{aligned} n - (\lceil\Delta - 1\rceil - r_0) < n - |\mathcal{I}_\alpha| + 1 \\ \Rightarrow |\mathcal{I}_\alpha| < (\lceil\Delta - 1\rceil - r_0) + 1 \\ \Rightarrow |\mathcal{I}_\alpha| \leq \lceil\Delta - 1\rceil - r_0. \end{aligned}$$

This means that at most $\lceil\Delta - 1\rceil - r_0$ singular values of $\widehat{\mathbf{X}}$ are greater than α_δ. Define

$$\mathbf{H}_0 = \begin{bmatrix} \mathbf{0} & \mathbf{X}_0 \\ \mathbf{X}_0^H & \mathbf{0} \end{bmatrix} \quad , \quad \widehat{\mathbf{H}} = \begin{bmatrix} \mathbf{0} & \widehat{\mathbf{X}} \\ \widehat{\mathbf{X}}^H & \mathbf{0} \end{bmatrix}. \qquad (33)$$

In fact, $\mathbf{H}_0$ and $\widehat{\mathbf{H}}$ are hermitian matrices that contain the singular values of $\mathbf{X}_0$ and $\widehat{\mathbf{X}}$, respectively, as their n largest eigenvalues and their negatives as the n smallest eigenvalues. Next, we apply Weyl's eigenvalue inequality [39] as

$$\begin{aligned} \lambda_{\lceil\Delta-1\rceil+1}(\mathbf{H}_0 - \widehat{\mathbf{H}}) &\leq \lambda_{r_0+1}(\mathbf{H}_0) + \lambda_{\lceil\Delta-1\rceil-r_0+1}(-\widehat{\mathbf{H}}) \\ &= \sigma_{r_0+1} + \hat{\sigma}_{\lceil\Delta-1\rceil-r_0+1} \\ &= \hat{\sigma}_{\lceil\Delta-1\rceil-r_0+1} \leq \alpha_\delta, \qquad (34) \end{aligned}$$

where $\lambda_i(\cdot)$ stands for the i-th largest eigenvalue. This reveals the fact that, $(\mathbf{X}_0 - \widehat{\mathbf{X}})$ has at most $\lceil\Delta - 1\rceil$ singular values greater than α_δ. Since $(\mathbf{X}_0 - \widehat{\mathbf{X}})$ is in the null space of $\mathcal{A}$, Corollary 3 implies that

$$\|\mathbf{X}_0 - \widehat{\mathbf{X}}\|_F \leq \frac{n\alpha_\delta}{\sqrt{\Delta} - \sqrt{\lceil\Delta - 1\rceil}}. \qquad (35)$$

■

Corollary 4: For the Gaussian function family given in (7), if (29) holds for a solution $\widehat{\mathbf{X}} \in \mathcal{X}$, then

$$\|\widehat{\mathbf{X}} - \mathbf{X}_0\|_F \leq \frac{n\delta\sqrt{2\ln n}}{\sqrt{\Delta} - \sqrt{\lceil\Delta - 1\rceil}}. \qquad (36)$$

Lemma 4: Let $f_\delta, F_\delta, \mathcal{X}$, and, $\mathbf{X}_0$ be as defined in Lemma 3 and assume $\mathbf{X}_\delta$ be the maximizer of $F_\delta(\mathbf{X})$ on $\mathcal{X}$. Then $\mathbf{X}_\delta$ satisfies (29).

Proof: One can write that

$$\begin{aligned} F_\delta(\mathbf{X}_\delta) &\geq F_\delta(\mathbf{X}_0) && (37)\\ &\geq n - r_0 && (38)\\ &\geq n - (\lceil\Delta - 1\rceil - r_0)\,. && (39) \end{aligned}$$

The first inequality comes from the fact that $\mathbf{X}_\delta$ is the maximizer of the $F_\delta(\mathbf{X})$, and the second one is true because $\mathbf{X}_0$ has $(n - r_0)$ singular values equal to zero; thus, in the summation $F_\delta(\mathbf{X}) = \sum_{i=1}^{n} f_\delta(\sigma_i)$, there are $(n - r_0)$ ones. Hence, $F_\delta(\mathbf{X}_0) \geq n - r_0$. To see the last inequality, note that $2r_0 < \Delta$ and

$$\Delta - 1 \leq \lceil\Delta - 1\rceil < \Delta \Rightarrow \Delta \leq \lceil\Delta - 1\rceil + 1 < \Delta + 1. \quad (40)$$

Thus, it can be concluded that $2r_0 < \lceil\Delta - 1\rceil + 1$ which results in $2r_0 \leq \lceil\Delta - 1\rceil$ because $r_0 \in \mathbb{N}$. Finally, $r_0 \leq \lceil\Delta - 1\rceil - r_0$ which implies that $n - (\lceil\Delta - 1\rceil - r_0) \leq n - r_0$. ■

Lemma 4 and Corollary 4 together prove that for the Gaussian family,

$$\lim_{\delta\to 0} \arg\max_{\mathcal{A}(\mathbf{X})=\mathbf{b}} F_\delta(\mathbf{X}) = \mathbf{X}_0.$$

In Theorem 4, we extend this result to all function classes that satisfy Assumption 1.

Theorem 4: Suppose $\mathcal{A} : \mathbb{R}^{n_1\times n_2} \to \mathbb{R}^m$ has Δ-spherical property and $\{f_\delta\}$ satisfies Assumption 1, and define $\mathcal{X}, F_\delta,$ and $\mathbf{X}_0$ as in Lemma 3. If $\mathbf{X}_\delta$ represents the maximizer of $F_\delta(\mathbf{X})$ over $\mathcal{X}$, then

$$\lim_{\delta\to 0} \mathbf{X}_\delta = \mathbf{X}_0.$$

Proof: By combining Lemma 3 and Lemma 4, we obtain that

$$\|\mathbf{X}_0 - \mathbf{X}_\delta\|_F \leq \frac{n\alpha_\delta}{\sqrt{\Delta} - \sqrt{\lceil\Delta - 1\rceil}}, \qquad (41)$$

where $\alpha_\delta = \left|f_\delta^{-1}(\frac{1}{n})\right|$. The consequence of Assumption 1 in (6) shows that for any $\epsilon > 0$ and $0 < x < 1$, one can set δ sufficiently small such that $\left|f_\delta^{-1}(x)\right| < \epsilon$. Therefore,

$$\lim_{\delta\to 0} \alpha_\delta = \lim_{\delta\to 0} \left|f_\delta^{-1}\left(\frac{1}{n}\right)\right| = 0.$$

This yields

$$\lim_{\delta\to 0} \|\mathbf{X}_0 - \mathbf{X}_\delta\|_F = 0.$$

■

V. Numerical Simulations

In this section, the performance of the SRF algorithm is evaluated empirically thorough simulations, and is compared to a few other algorithms. In the first part of numerical experiments, effects of the algorithm parameters ($L, c,$ and ϵ) in reconstruction accuracy are studied. Next, in the second part, the so called *phase transition* curve [1] between perfect recovery and failure is experimentally obtained for the SRF algorithm and is compared to that of the nuclear norm minimization. In the third part of simulations, accuracy and computational load of the SRF algorithm in solving the matrix completion problem are compared to two well known matrix completion algorithms. Finally, in the fourth part, robustness of the SRF against the measurement noise is experimentally verified.

To generate a testing random matrix $\mathbf{X} \in \mathbb{R}^{n_1\times n_2}$ of rank r, the following procedure is used. We generate two random matrices $\mathbf{X}_L \in \mathbb{R}^{n_1\times r}$ and $\mathbf{X}_R \in \mathbb{R}^{r\times n_2}$ whose

entries are independent and identically drawn from a Gaussian distribution with zero mean and unit variance. Then $\mathbf{X}$ is constructed as the product of $\mathbf{X}_L$ and $\mathbf{X}_R$, i.e., $\mathbf{X} = \mathbf{X}_L\mathbf{X}_R$. In the affine rank minimization problems, the affine constraint $\mathcal{A}(\mathbf{X}) = \mathbf{b}$ is converted to $\mathbf{A}\,\mathrm{vec}(\mathbf{X}) = \mathbf{b}$, where $\mathbf{A} \in \mathbb{R}^{m\times n_1 n_2}$ denotes the matrix representation of the linear operator $\mathcal{A}$ and $\mathrm{vec}(\mathbf{X})$ denotes the vector in $\mathbb{R}^{n_1 n_2}$ with the columns of $\mathbf{X}$ stacked on top of one another. All entries of $\mathbf{A}$ are drawn independently and identically from a zero mean, unit variance Gaussian distribution. Moreover, in the matrix completion simulations, the index set Ω of revealed entries is selected uniformly at random. We denote the result of the SRF algorithm by $\widehat{\mathbf{X}}$ and measure its accuracy by $\mathrm{SNR}_{rec} = 20\log_{10}(\|\mathbf{X}\|_F/\|\mathbf{X}-\widehat{\mathbf{X}}\|_F)$ in dB, which is referred to as the *reconstruction SNR*. In addition, by term *easy problems*, we mean problems in which the ratio m/d_r is greater than 3, where $d_r = r(n_1+n_2-r)$ denotes the number of degrees of freedom in a real-valued rank–r matrix [2]. When this ratio is lower than or equal to 3, it is called a *hard* problem.

In all experiments the parameter μ is fixed at 1, and we use a decreasing sequence of δ's according to $\delta_j = c\delta_{j-1}, j > 1$, where $0 < c < 1$ denotes the rate of decay. The value of δ_1 is set twice as large as the largest singular value of the initial estimate. For the sake of simplicity square matrices are tested, so $n_1 = n_2 = n$.

Our simulations are performed in MATLAB 8 environment using an Intel Core i7, 2.6 GHz processor with 8 GB of RAM, under Microsoft Windows 7 operating system.

A. Parameters Effects

Experiment 1. As already discussed in Section III-A, it is not necessary to wait for complete convergence of the internal optimization loop. Instead, a few iterations suffice to only move toward the global maximizer for the current value of δ. Thus, we suggested to do the internal loop for fixed L times. However, the optimal choice of L depends on the aspects of the problem at hand. As a rule of thumb, when the problem becomes harder, i.e., the number of measurements decreases toward the degrees of freedom, larger values of L should be used. Likewise, for easier problems, smaller values of L decrease the computational load of the algorithm, while the accuracy will not degrade very much.

To see the above rule, the affine rank minimization problem defined in (1) is solved using the SRF algorithm, while changing the parameter L. We put $n = 30, r = 3, \epsilon = 10^{-5}$, and $c = 0.9$. The number of measurements change from 250 to 500 to cover both easy and hard problems. To obtain accurate $\mathrm{SNR}_{\mathrm{rec}}$ estimates, the trials are repeated 100 times. Fig. 2 shows the effects of changing L from 1 to 10. It can be concluded from Fig. 2 that for easy and hard problems, there is a threshold value for L, which choosing L beyond it can only slightly improves reconstruction SNR. However, thorough simulations, we found that increasing the L boosts the computation time almost linearly. For instance, when $m = 500$ and $L = 1$, the average computation time is about 0.5 sec, while this time increases to about 1.2 sec for $L = 5$ and to about 2.2 sec for $L = 10$.

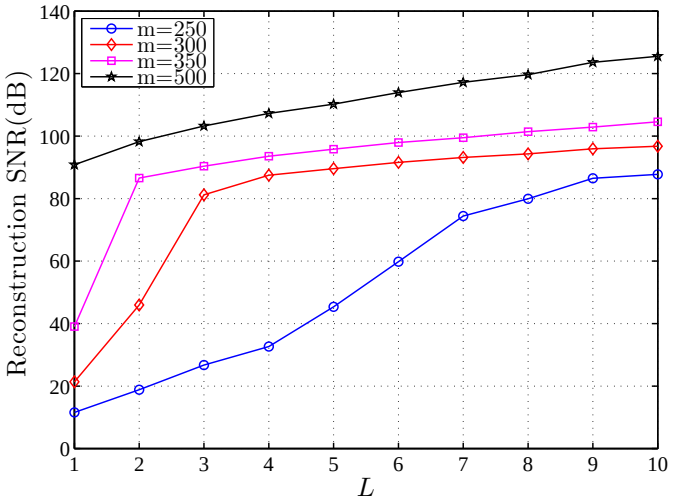

Fig. 2. Averaged $\mathrm{SNR}_{\mathrm{rec}}$ of the SRF algorithm in solving the affine RM problem versus L. Matrix dimensions are fixed to 30×30, and r is set to 3. The parameter c and ϵ are set to 0.9 and 10^{-5}, respectively to have small effect on this analysis. SNR's are averaged over 100 runs.

Experiment 2. The next experiment is devoted to the dependence of the accuracy of the SRF algorithm on the parameter c. In this experiment, the dimensions of the matrix are the same as the previous experiment, and L and ϵ are fixed to 8 and 10^{-5}, respectively. Affine rank minimization and matrix completion problems are solved with two different number of measurements to show the effect on different conditions. c is changed from 0.15 to 0.95, $\mathrm{SNR}_{\mathrm{rec}}$'s are averaged on 100 runs. Fig. 3 depicts the reconstruction SNR versus the scale parameter c for different problems. It is obvious that SNR increases as c approaches 1. However, when c exceeds a critical value, SNR remains almost constant.

Generally, the optimal choice of c depends on the criterion which aimed to be optimized. When accuracy is the key criterion, c should be chosen close to 1, which results in slow decay in the sequence of δ and a higher computational time.

Experiment 3. In this experiment, the effect of ϵ on the accuracy of the algorithm is analyzed. All dimensions and parameters are the same as the experiment 2 except c and ϵ. c is fixed to 0.9, and ϵ is changed from 10^{-1} to 10^{-6}. The result of this experiment is shown in Fig. 4. It is seen that after passing a critical value, logarithmic reconstruction SNR increases almost linearly as ϵ decreases linearly in logarithmic scale. Hence, it can be concluded that ϵ controls the closeness of the final solution to the minimum rank solution.

B. Phase Transition Curve

Experiment 4. To the best of our knowledge, the tightest available bound on the number of required samples for the NNM to find the minimum rank solution is two times greater than that of the rank minimization problem [16]. More precisely, for the given linear operator which has a null space with Δ-spherical section property, (1) has a unique solution if $\mathrm{rank}(\mathbf{X}_0) < \Delta/2$, while (3) and (1) share a common solution if $\mathrm{rank}(\mathbf{X}_0) < \Delta/4$. Our main goal in this experiment is to show that the SRF algorithm can recover the solution

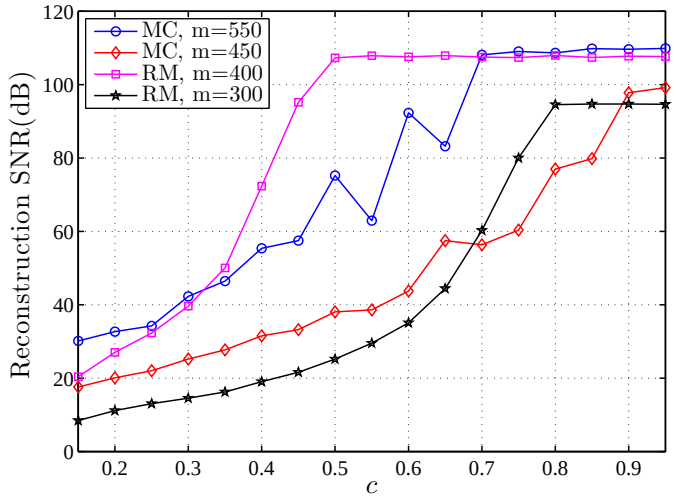

Fig. 3. Averaged $\mathrm{SNR}_{\mathrm{rec}}$ of the SRF algorithm as a function of c. Matrix dimensions are fixed to 30×30 and r is set to 3. The parameter L and ϵ are set to 8 and 10^{-5}, respectively to have small effect on this analysis. SNR's are averaged over 100 runs. 'MC' and 'RM' denote the matrix completion and affine rank minimization problems, respectively. For two MC problems, m is set to 450 and 550, and for two RM problems, is set to 300 and 400.

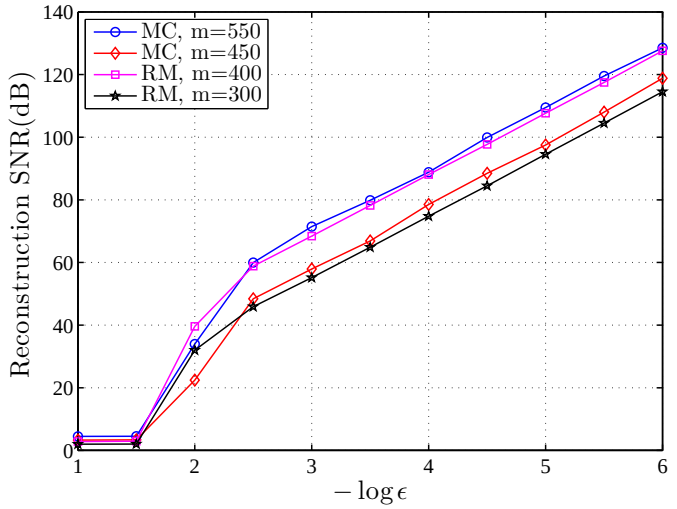

Fig. 4. Averaged $\mathrm{SNR}_{\mathrm{rec}}$ of the SRF algorithm as a function of ϵ. Matrix dimensions are fixed to 30×30 and r is set to 3. The parameter L and c are set to 8 and 0.9, respectively to have small effect on this analysis. SNR's are averaged over 100 runs. ϵ is changed from 10^{-1} to 10^{-6}. 'MC' and 'RM' denote the matrix completion and affine rank minimization problems, respectively. For two MC problems, m is set to 450 and 550, and for two RM problems, is set to 300 and 400.

in situations where the nuclear norm minimization fails. In other words, this algorithm can get closer to the intrinsic bound in recovering of low rank matrices. The computational cost of the SRF algorithm will be compared to an efficient implementation of the nuclear norm minimization in the next experiment.

Like compressive sensing literature, the phase transition can be used to indicate the region of perfect recovery and failure [1]. Fig. 5 shows the results of applying the proposed algorithm on the affine rank minimization. A solution is declared to be recovered if reconstruction SNR is greater than 60 dB. The matrix dimension is $40 \times 40, \epsilon = 10^{-5}, L = 6,$ and $c = 0.9$.

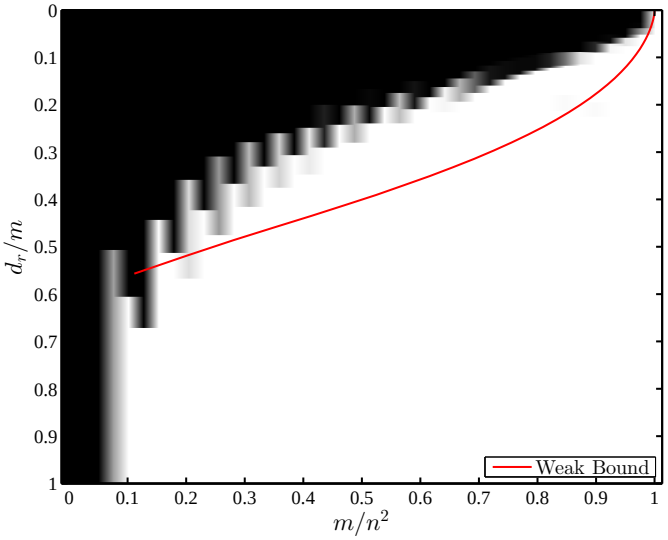

Fig. 5. Phase transition of the SRF algorithm in solving the affine RM problem. $n = 40, \epsilon = 10^{-5}, L = 6, c = 0.9$, and simulations are performed 50 times. Gray-scale color of each cell indicates the rate of perfect recovery. White denotes 100% recovery rate, and black denotes 0% recovery rate. A recovery is perfect if the $\mathrm{SNR}_{\mathrm{rec}}$ is greater than 60 dB. The red trace shows the so called weak bound derived in [17] for the number of required measurements for perfect recovery of low rank matrix using the nuclear norm heuristics.

Simulations are repeated 50 times. The gray color of cells indicates the empirical recovery rate. White denotes perfect recovery in all trials, and black shows unsuccessful recovery for all experiments. Furthermore, the thin trace on the figure shows a theoretical bound in recovery of low rank solutions via the nuclear norm minimization found in [17]. In [17], it is shown that this bound is very consistent to the numerical simulations; thus, we use it for the sake of comparison. One can see in Fig. 5 that there is a very clear gap between this bound and phase transition of the SRF algorithm.

C. Matrix Completion

Experiment 5. The accuracy and computational costs of the proposed algorithm in solving the matrix completion problem are analyzed and compared to two other methods. Among many available approaches, FPCA [19] and OptSpace [22] are selected as competitors. FPCA is an efficient implementation of the NNM and can obtain very accurate results with low complexity [19], while OptSpace is based on trimming rows and columns of the incomplete matrix followed by truncation of some singular values of the trimmed matrix [22].

Although CPU time is not an accurate measure of the computational cost, we use it as a rough estimate to compare algorithm complexities. Every simulation is run 100 times, and the results are averaged. We set $\epsilon = 10^{-5}, L = 8,$ and $c = 0.9$. FPCA and OptSpace are run by default parameters.

Table I shows some results of this comparison for easy problems. In all cases, the accuracy of SRF in recovering the solution is considerably higher than other algorithms. Also, we observe that at large matrix dimensions, the run time of the SRF algorithm is comparable or even better than OptSpace. Note that c is set to 0.9 to accommodate the worst case

TABLE I
RESULTS OF APPLYING THREE ALGORITHMS ON EASY MC PROBLEMS.

Algorithm	SRF		OptSpace		FPCA	
$(n, r, m/dr)$	Time	Error	Time	Error	Time	Error
(50,5,4)	0.19	3.17e-8	0.09	6.85e-2	0.28	1.62e-2
(50,5,5)	0.19	8.45e-9	0.08	4.53e-7	0.29	6.81e-1
(100,2,5)	1.48	3.23e-6	0.15	1.93e-2	0.07	3.68e-3
(100,2,10)	0.57	1.96e-7	0.14	8.30e-7	0.07	4.55e-5
(100,5,5)	0.68	1.33e-7	0.30	2.42e-2	0.09	3.69e-5
(100,5,10)	0.71	8.23e-9	65.7	1.37	0.82	6.98e-1
(200,5,5)	4.69	7.10e-7	0.88	2.56e-2	0.19	2.35e-4
(200,5,10)	2.92	2.50e-8	0.64	4.90e-7	0.30	6.42e-6
(200,10,5)	3.25	6.36e-8	2.76	5.78e-7	0.33	1.79e-5
(200,10,10)	3.72	6.44e-9	1.27	2.28e-7	2.82	8.09e-1
(200,20,5)	4.00	4.37e-9	9.64	1.98e-7	2.99	8.40e-1
(300,30,5)	12.6	3.17e-9	75.6	1.56e-7	8.07	8.80e-1

TABLE II
RESULTS OF APPLYING THREE ALGORITHMS ON HARD MC PROBLEMS.

Algorithm	SRF		OptSpace		FPCA	
$(n, r, m/dr)$	Time	Error	Time	Error	Time	Error
(50,5,2,5)	0.30	8.99e-7	0.09	3.85e-1	1.93	2.96e-6
(50,5,3)	0.23	3.21e-7	0.11	1.71e-1	0.06	4.85e-5
(50,10,1.5)	0.62	8.47e-5	1.97	7.40e-1	2.44	3.64e-6
(50,10,2.5)	0.21	1.88e-8	1.10	3.46e-1	2.27	9.97e-8
(50,20,1.25)	0.60	8.88e-6	0.63	7.72e-1	4.45	3.42e-1
(50,20,1.5)	0.27	8.67e-8	12.4	7.24e-1	4.57	2.48e-1
(50,25,1.25)	0.43	4.16e-7	2.55	8.21e-1	4.56	3.82e-1
(100,15,2.5)	0.84	8.10e-8	1.41	4.21e-1	6.02	4.36e-7
(100,40,1.25)	2.07	6.83e-7	19.5	8.74e-1	13.3	3.20e-1
(100,50,1.25)	1.57	2.27e-7	19.5	8.83e-1	14.3	3.62e-1
(200,20,2,5)	4.75	2.05e-7	19.5	3.98e-1	21.2	1.58e-6
(200,100,1.25)	7.24	1.24e-7	19.5	9.35e-1	57.5	3.51e-1
(200,110,1.25)	6.04	2.86e-8	19.5	9.22e-1	57.5	3.63e-1

scenario of hard problems. However, it can be tuned to speed up the SRF method, if the working regime is *a priori* known.

The significant advantage of SRF is in solving hard problems. As Table II shows, in all cases except one, SRF has the best accuracy. Particularly, when the matrix rank increases (for examples in the experiments corresponding to the two bottom rows of Table II), SRF exhibits an accuracy which is 10^6 times better than the closest competitor, without compromising the run time.

D. Noisy Measurements

Experiment 6. Although the SRF algorithm is designed for noiseless measurements, we show experimentally that it is robust against noise. The measurement model can be updated as

$$\mathcal{A}(\mathbf{X}) = \mathbf{b} + \boldsymbol{\nu}$$

in the existence of an additive noise vector $\boldsymbol{\nu}$. To check the robustness, we set up an experiment with $n = 30, r = 3, c = 0.9, \epsilon = 10^{-5}$. We generate entries of $\boldsymbol{\nu}$ as *i.i.d.* realizations of zero-mean Gaussian random variables with unit variance. We also vary m from d_r to $3d_r$, and average SNR_{rec} over 100 trials. Fig. 6 shows the results of SNR_{rec} *vs.* the number of measurements for various measurement SNR values ($\text{SNR}_{meas} = 20\log_{10}(\|\mathbf{b}\|_2/\|\boldsymbol{\nu}\|_2)$). Except for the noiseless case where $\text{SNR}_{meas} = \infty$, we observe that the curves of SNR_{rec} stagnate almost at the same level as SNR_{meas}.

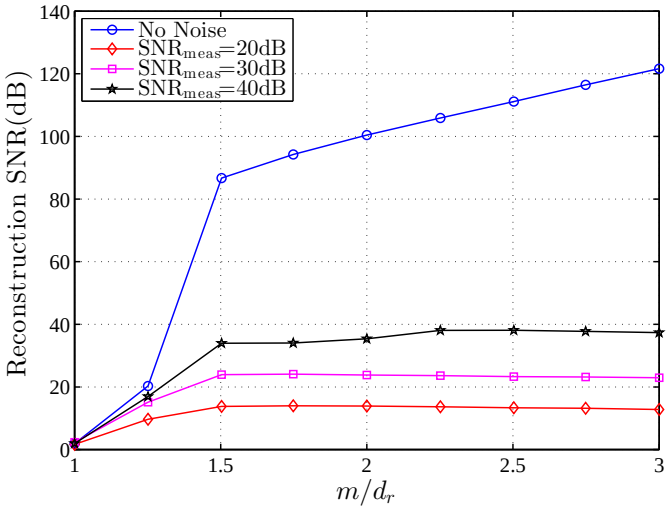

Fig. 6. SNR_{rec} versus number of measurements when there is measurement noise. $n = 30, r = 3, c = 0.9, \epsilon = 10^{-5}$, m is changed from d_r to $3d_r$, and results are averaged on 100 runs.

VI. CONCLUSION

In this work, a rank minimization technique based on approximating the rank function and successively improving the quality of the approximation was proposed. We theoretically showed that the proposed iterative method asymptotically achieves the solution to the rank minimization problem, provided that the middle-stage minimizations are exact. We further examined the performance of this method using numerical simulations. The comparisons against two common methods reveal superiority of the proposed technique in terms of both quality and computational time, especially when the number of affine measurements decreases towards the unique representation lower-bound. By providing examples, we even demonstrate the existence of scenarios in which the conventional nuclear norm minimization fails to recover the unique low rank matrix associated with the linear constraints, while the proposed method succeeds.

APPENDIX A

In this appendix, the closed form least squares solution of the orthogonal back projection onto the feasible set is derived. Let us cast the affine constraints $\mathcal{A}(\mathbf{X}) = \mathbf{b}$ as $\mathbf{A}\,\text{vec}(\mathbf{X}) = \mathbf{b}$. The goal is to find the nearest point in the affine set to the result of the j-th iteration, $\mathbf{X}_j$. Mathematically,

$$\min_{\mathbf{X}} \|\mathbf{X} - \mathbf{X}_j\|_F^2 \text{ subject to } \mathcal{A}(\mathbf{X}) = \mathbf{b}, \tag{42}$$

or equivalently,

$$\min_{\mathbf{X}} \|\,\text{vec}(\mathbf{X}) - \text{vec}(\mathbf{X}_j)\|^2 \text{ subject to } \mathbf{A}\,\text{vec}(\mathbf{X}) = \mathbf{b}, \tag{43}$$

where $\|\cdot\|$ denotes vector ℓ_2-norm. By putting $\mathbf{y} = \text{vec}(\mathbf{X}) - \text{vec}(\mathbf{X}_j)$, the problem (43) can be easily cast as the following least squares problem

$$\min_{\mathbf{y}} \|\mathbf{y}\|_2^2 \text{ subject to } \mathbf{A}\mathbf{y} = \mathbf{b} - \mathbf{A}\,\text{vec}(\mathbf{X}_j). \tag{44}$$

Let $\mathbf{A}^\dagger = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}$ be the Moore-Penrose pseudoinverse of $\mathbf{A}$. Then the least squares solution of (42) will be

$\mathbf{X} = \mathrm{mat}_{n_1,n_2}\left(\mathbf{A}^\dagger\mathbf{b} + [\mathbf{I} - \mathbf{A}^\dagger\mathbf{A}]\,\mathrm{vec}(\mathbf{X}_j)\right)$, where $\mathbf{I}$ denotes the identity matrix, and $\mathrm{mat}_{n_1,n_2}(\cdot)$ reverses the operation of vectorization, i.e., $\mathrm{mat}_{n_1,n_2}\big(\mathrm{vec}(\mathbf{X})\big) = \mathbf{X}$.

APPENDIX B
PROOF OF THEOREM 2

Proof: Let $\mathbf{X}_\delta = \arg\max_{\mathcal{A}(\mathbf{X})=\mathbf{b}} F_\delta(\mathbf{X})$. To prove $\lim_{\delta\to\infty}\mathbf{X}_\delta = \widehat{\mathbf{X}}$, we first focus on singular values $\sigma_i(\mathbf{X}_\delta)$. Due to Assumption 1, it is known that $\lim_{\delta\to\infty} F_\delta(\widehat{\mathbf{X}}) = n$. Thus, for any $\epsilon \geq 0$, one can set δ large enough such that $F_\delta(\widehat{\mathbf{X}}) \geq n - \epsilon$. Note that for any $1 \leq i \leq n$, we have that

$$n - 1 + f_\delta\big(\sigma_i(\mathbf{X}_\delta)\big) \geq F_\delta(\mathbf{X}_\delta) \geq F_\delta(\widehat{\mathbf{X}}) \geq n - \epsilon, \quad (45)$$

or

$$f_\delta\big(\sigma_i(\mathbf{X}_\delta)\big) \geq 1 - \epsilon. \quad (46)$$

This implies that $\sigma_i(\mathbf{X}_\delta) \leq |f_\delta^{-1}(1-\epsilon)| = \delta|f^{-1}(1-\epsilon)|$. Hence,

$$0 \leq \lim_{\delta\to\infty}\frac{\sigma_i(\mathbf{X}_\delta)}{\delta} \leq \big|f^{-1}(1-\epsilon)\big|, \quad \forall\ 0 < \epsilon < 1. \quad (47)$$

By considering the above inequality for $\epsilon \to 0$, we conclude that

$$\lim_{\delta\to\infty}\frac{\sigma_i(\mathbf{X}_\delta)}{\delta} = 0, \quad 1 \leq i \leq n. \quad (48)$$

Using the Taylor expansion, we can rewrite $f(\cdot)$ as

$$f(s) = 1 - \gamma s^2 + g(s), \quad (49)$$

where $\gamma = -\frac{1}{2}f''(0)$ and

$$\lim_{s\to 0}\frac{g(s)}{s^2} = 0. \quad (50)$$

In turn, $F_\delta(\cdot)$ can be rewritten as

$$\begin{aligned} F_\delta(\mathbf{X}) &= \sum_{i=1}^{n} f_\delta\big(\sigma_i(\mathbf{X})\big) \\ &= n - \frac{\gamma}{\delta^2}\sum_{i=1}^{n}\sigma_i^2(\mathbf{X}) + \sum_{i=1}^{n} g(\sigma_i(\mathbf{X})/\delta). \end{aligned} \quad (51)$$

This helps us rewrite $F_\delta(\mathbf{X}_\delta) \geq F_\delta(\widehat{\mathbf{X}})$ in the form

$$\begin{aligned} \frac{\gamma}{\delta^2}\sum_{i=1}^{n}\sigma_i^2(\mathbf{X}_\delta) - \sum_{i=1}^{n} g(\sigma_i(\mathbf{X}_\delta)/\delta) \leq \\ \frac{\gamma}{\delta^2}\sum_{i=1}^{n}\sigma_i^2(\widehat{\mathbf{X}}) - \sum_{i=1}^{n} g(\sigma_i(\widehat{\mathbf{X}})/\delta), \end{aligned}$$

or similarly,

$$\begin{aligned} \|\boldsymbol{\sigma}(\mathbf{X}_\delta)\|^2 - \|\boldsymbol{\sigma}(\widehat{\mathbf{X}})\|^2 &\leq \frac{\sum_{i=1}^{n} g\big(\sigma_i(\mathbf{X}_\delta)/\delta\big) - g\big(\sigma_i(\widehat{\mathbf{X}})/\delta\big)}{\gamma\,\delta^{-2}} \\ &\leq \frac{\|\boldsymbol{\sigma}(\mathbf{X}_\delta)\|^2}{\gamma}\sum_{i=1}^{n}\frac{\big|g\big(\sigma_i(\mathbf{X}_\delta)/\delta\big)\big|}{\big(\sigma_i(\mathbf{X}_\delta)/\delta\big)^2} \\ &\quad + \frac{\|\boldsymbol{\sigma}(\widehat{\mathbf{X}})\|^2}{\gamma}\sum_{i=1}^{n}\frac{\big|g\big(\sigma_i(\widehat{\mathbf{X}})/\delta\big)\big|}{\big(\sigma_i(\widehat{\mathbf{X}})/\delta\big)^2}. \end{aligned}$$

Recalling $\|\boldsymbol{\sigma}(\mathbf{X})\|^2 = \|\mathbf{X}\|_F^2$, we can write that

$$\|\mathbf{X}_\delta\|_F^2 \leq \|\widehat{\mathbf{X}}\|_F^2\frac{1 + \frac{1}{\gamma}\left(\sum_{i=1}^{n}\left|\frac{g\big(\sigma_i(\widehat{\mathbf{X}})/\delta\big)}{\big(\sigma_i(\widehat{\mathbf{X}})/\delta\big)^2}\right|\right)}{\left|1 - \frac{1}{\gamma}\left(\sum_{i=1}^{n}\left|\frac{g\big(\sigma_i(\mathbf{X}_\delta)/\delta\big)}{\big(\sigma_i(\mathbf{X}_\delta)/\delta\big)^2}\right|\right)\right|}. \quad (52)$$

We also have

$$\lim_{\delta\to\infty}\sigma_i(\widehat{\mathbf{X}})/\delta = 0 \quad \overset{(50)}{\Longrightarrow} \lim_{\delta\to\infty}\frac{g\big(\sigma_i(\widehat{\mathbf{X}})/\delta\big)}{\big(\sigma_i(\widehat{\mathbf{X}})/\delta\big)^2} = 0, \quad (53)$$

$$\lim_{\delta\to\infty}\sigma_i(\mathbf{X}_\delta)/\delta = 0 \quad \overset{(50)}{\Longrightarrow} \lim_{\delta\to\infty}\frac{g\big(\sigma_i(\mathbf{X}_\delta)/\delta\big)}{\big(\sigma_i(\mathbf{X}_\delta)/\delta\big)^2} = 0. \quad (54)$$

Application of (53) and (54) in (52) results in

$$\lim_{\delta\to\infty}\|\mathbf{X}_\delta\|_F^2 \leq \|\widehat{\mathbf{X}}\|_F^2. \quad (55)$$

According to the definition of $\widehat{\mathbf{X}}$, we have $\|\mathbf{X}_\delta\|_F^2 \geq \|\widehat{\mathbf{X}}\|_F^2$ and $\lim_{\delta\to\infty}\|\mathbf{X}_\delta\|_F^2 \geq \|\widehat{\mathbf{X}}\|_F^2$. Combining this result with (55), we obtain

$$\lim_{\delta\to\infty}\|\mathbf{X}_\delta\|_F^2 = \|\widehat{\mathbf{X}}\|_F^2. \quad (56)$$

Also, any matrix in $\mathrm{null}(\mathcal{A})$ is perpendicular to $\widehat{\mathbf{X}}$ since it is the minimum Frobenius-norm solution of the $\mathcal{A}(\mathbf{X}) = \mathbf{b}$. Thus,

$$\|\mathbf{X}_\delta\|_F^2 = \|\widehat{\mathbf{X}}\|_F^2 + \|\mathbf{X}_\delta - \widehat{\mathbf{X}}\|_F^2. \quad (57)$$

In summary, we conclude that $\lim_{\delta\to\infty}\|\mathbf{X}_\delta - \widehat{\mathbf{X}}\|_F^2 = 0$ which establishes $\lim_{\delta\to\infty}\mathbf{X}_\delta = \widehat{\mathbf{X}}$. ■

ACKNOWLEDGMENT

The authors would like to thank Mr. Ghasemi for his help in obtaining preliminary results.

REFERENCES

[1] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 55, pp. 471–501, 2010.

[2] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[3] B. Recht, W. Xu, and B. Hassibi, "Null space conditions and thresholds for rank minimization," *Mathematical Programming*, vol. 127, no. 1, pp. 175–202, 2011.

[4] E. J. Candés and Y. Plan, "Matrix completion with noise," *Proceedings of IEEE*, vol. 98, no. 6, pp. 925–936, 2010.

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

[6] Zh. Liu and L. Vandenberghe, "Interior-point method for nuclear norm approximation with application to system identification," *SIAM. J. Matrix Anal. & Appl.*, vol. 31, no. 3, pp. 1235–1256, 2009.

[7] T. Cox and M. A. A. Cox, *Multidimensional Scaling*, Chapman and Halle, 1994.

[8] Y. Amit, M. Fink, N. Srebro, and S. Ullman, "Uncovering shared structures in multiclass classification," in *Proceedings of the 24 International Conference on Machine Learning*, vol. 24.

[9] D. Gross, Y. K. Liu, S. T. Flammia, S. Becker, and J. Eisert, "Quantum state tomography via compressed sensing," *Physical review letters*, vol. 105, no. 15, pp. 150401, 2010.

[10] J. Meng, W. Yin, H. Li, E. Hossain, and Z. Han, "Collaborative spectrum sensing from sparse observations in cognitive radio networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 2, pp. 327–337, 2011.

[11] N. Ito, E. Vincent, N. Ono, R. Gribonval, and S. Sagayama, "Crystal-music: Accurate localization of multiple sources in diffuse noise environments using crystal-shaped microphone arrays," *Latent Variable Analysis and Signal Separation*, pp. 81–88, 2010.

[12] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antenna and Prorogation*, vol. 34, no. 3, pp. 276–280, 1986.

[13] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[14] A. L. Chistov and Yu. Grigoriev, "Complexity of quantifier elimination in the theory of algebraically closed fields," in *Proceedings of the 11th Symposium on Mathematical Foundations of Computer Science*, 1984, vol. 176, pp. 17–31.

[15] M. Fazel, *Matrix Rank Minimization with Applications*, Ph.D. thesis, Stanford University, 2002.

[16] K. Mohan, M. Fazel, and B. Hassibi, "A simplified approach to recovery conditions for low rank matrices," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, July and August 2011, pp. 2318–2322.

[17] S. Oymak and B. Hassibi, "New null space results and recovery thresholds for matrix rank minimization," *arXiv preprint arXiv:1011.6326*, 2010.

[18] L. Kong, L. Tunçel, and N. Xiu, "Sufficient conditions for low-rank matrix recovery, translated from sparse signal recovery," *arXiv preprint arXiv:1106.3276*, 2011.

[19] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and bregman iterative methods for matrix rank minimization," *Mathematical Programming*, vol. 128, no. 1, pp. 321–353, 2011.

[20] K. C. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pacific Journal of Optimization*, vol. 6, pp. 615–640, 2010.

[21] J. F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[22] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Trans. on Information Theory*, vol. 56, no. 6, pp. 2980–2998, 2010.

[23] K. Lee and Y. Bresler, "Admira: Atomic decomposition for minimum rank approximation," *IEEE Trans. on Information Theory*, vol. 56, no. 9, pp. 4402–4416, 2010.

[24] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[25] H. Mohimani, M. Babaie-Zadeh, and Ch. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 289–301, January 2009.

[26] H. Ghasemi, M. Malek-Mohammadi, M. Babaie-Zadeh, and C. Jutten, "SRF: Matrix completion based on smoothed rank function," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, 2011, pp. 3672–3675.

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

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

[29] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[30] Y. Zhang, "Theory of compressive sensing via ℓ_1 minimization: A non-rip analysis and extensions," Technical report tr08-11 revised, Dept. of Computational and Applied Mathematics, Rice University, 2008, Available at http://www.caam.rice.edu/ zhang/reports/tr0811_revised.pdf.

[31] S. Oymak, M. A. Khajehnejad, and B. Hassibi, "Improved thresholds for rank minimizations," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 5988–5991.

[32] A. S. Lewis and H. S. Sendov, "Nonsmooth analysis of singular values. part i: Theory," *Set-Valued Analysis*, vol. 13, no. 3, pp. 213–241, 2005.

[33] A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, 1987.

[34] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, 1999.

[35] A. S. Lewis, "The convex analysis of unitarily invariant matrix norms," *Journal of Convex Analysis*, vol. 2, pp. 173–183, 1995.

[36] A. S. Lewis, "Convex analysis on the hermitian matrices," *SIAM J. Optimization*, vol. 6, pp. 164–177, 1996.

[37] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

[38] K. Dvijotham and M. Fazel, "A nullspace analysis of the nuclear norm heuristic for rank minimization," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2010, pp. 3586–3589.

[39] W. Fulton, "Eigenvalues of sums of Hermitian matrices," in *Séminaire N. Bourbaki*, 1997-1998, Exp. no 845, pp. 255–269.