Hindawi Publishing Corporation
Abstract and Applied Analysis
Volume 2013, Article ID 101974, 9 pages
http://dx.doi.org/10.1155/2013/101974

Research Article

s-Goodness for Low-Rank Matrix Recovery

Lingchen Kong,[1] Levent Tunçel,[2] and Naihua Xiu[1]

[1] *Department of Applied Mathematics, Beijing Jiaotong University, Beijing 100044, China*
[2] *Department of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Waterloo, ON, Canada N2L 3G1*

Correspondence should be addressed to Lingchen Kong; konglchen@126.com

Received 21 January 2013; Accepted 17 March 2013

Academic Editor: Jein-Shan Chen

Copyright © 2013 Lingchen Kong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Low-rank matrix recovery (LMR) is a rank minimization problem subject to linear equality constraints, and it arises in many fields such as signal and image processing, statistics, computer vision, and system identification and control. This class of optimization problems is generally $\mathcal{NP}$ hard. A popular approach replaces the rank function with the nuclear norm of the matrix variable. In this paper, we extend and characterize the concept of *s*-goodness for a sensing matrix in sparse signal recovery (proposed by Juditsky and Nemirovski (Math Program, 2011)) to linear transformations in LMR. Using the two characteristic *s*-goodness constants, γ_s and $\widehat{\gamma}_s$, of a linear transformation, we derive necessary and sufficient conditions for a linear transformation to be *s*-good. Moreover, we establish the equivalence of *s*-goodness and the null space properties. Therefore, *s*-goodness is a necessary and sufficient condition for exact *s*-rank matrix recovery via the nuclear norm minimization.

1. Introduction

Low-rank matrix recovery (LMR for short) is a rank minimization problem (*RMP*) with linear constraints or the affine matrix rank minimization problem which is defined as follows:

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

where $X \in \mathbb{R}^{m\times n}$ is the matrix variable, $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ is a linear transformation, and $b \in \mathbb{R}^p$. Although specific instances can often be solved by specialized algorithms, the LMR is $\mathcal{NP}$ hard. A popular approach for solving LMR in the systems and control community is to minimize the trace of a positive semidefinite matrix variable instead of its rank (see, e.g., [1, 2]). A generalization of this approach to nonsymmetric matrices introduced by Fazel et al. [3] is the famous convex relaxation of LMR (1), which is called *nuclear norm minimization* (*NNM*):

$$\begin{aligned} &\min && \|X\|_*, \\ &\text{s.t.} && \mathcal{A}X = b, \end{aligned} \tag{2}$$

where $\|X\|_*$ is the *nuclear norm* of X, that is, the sum of its singular values. When $m = n$ and the matrix $X := \operatorname{Diag}(x)$, $x \in \mathbb{R}^n$, is diagonal, the LMR (1) reduces to *sparse signal recovery* (*SSR*), which is the so-called *cardinality minimization problem* (*CMP*):

$$\begin{aligned} &\min && \|x\|_0 \\ &\text{s.t.} && \Phi x = b, \end{aligned} \tag{3}$$

where $\|x\|_0$ denotes the number of nonzero entries in the vector x and $\Phi \in \mathbb{R}^{m\times n}$ is a given sensing matrix. A well-known heuristic for SSR is the ℓ_1*-norm minimization* relaxation (basis pursuit problem):

$$\begin{aligned} &\min && \|x\|_1 \\ &\text{s.t.} && \Phi x = b, \end{aligned} \tag{4}$$

where $\|x\|_1$ is the ℓ_1*-norm* of x, that is, the sum of absolute values of its entries.

LMR problems have many applications and they appeared in the literature of a diverse set of fields including signal and image processing, statistics, computer vision, and system identification and control. For more details, see the recent

paper [4]. LMR and NNM have been the focus of some recent research in the optimization community, see; for example, [4–15]. Although there are many papers dealing with algorithms for NNM such as interior-point methods, fixed point and Bregman iterative methods, and proximal point methods, there are fewer papers dealing with the conditions that guarantee the success of the low-rank matrix recovery via NNM. For instance, following the program laid out in the work of Candès and Tao in compressed sensing (CS, see, e.g., [16–18]), Recht et al. [4] provided a certain *restricted isometry property* (RIP) condition on the linear transformation which guarantees that the minimum nuclear norm solution is the minimum rank solution. Recht et al. [14, 19] gave the *null space property* (NSP) which characterizes a particular property of the null space of the linear transformation, which is also discussed by Oymak et al. [20, 21]. Note that NSP states a necessary and sufficient condition for exactly recovering the low-rank matrix via nuclear norm minimization. Recently, Chandrasekaran et al. [22] proposed that a fixed s-rank matrix X_0 can be recovered if and only if the null space of $\mathscr{A}$ does not intersect the tangent cone of the nuclear norm ball at X_0.

In the setting of CS, there are other characterizations of the sensing matrix, under which ℓ_1-norm minimization can be guaranteed to yield an optimal solution to SSR, in addition to RIP and null-space properties, see; for example, [23–26]. In particular, Juditsky and Nemirovski [24] established necessary and sufficient conditions for a Sensing matrix to be "*s-good*" to allow for exact ℓ_1-recovery of sparse signals with s nonzero entries when no measurement noise is present. They also demonstrated that these characteristics, although difficult to evaluate, lead to verifiable sufficient conditions for exact SSR and to efficiently computable upper bounds on those s for which a given sensing matrix is s-good. Furthermore, they established instructive links between s-goodness and RIP in the CS context. One may wonder whether we can generalize the s-goodness concept to LMR and still maintain many of the nice properties as done in [24]. Here, we deal with this issue. Our approach is based on the singular value decomposition (SVD) of a matrix and the partition technique generalized from CS. In the next section, following Juditsky and Nemirovski's terminology, we propose definitions of s-goodness and G-numbers, γ_s and $\widehat{\gamma}_s$, of a linear transformation in LMR and then we provide some basic properties of G-numbers. In Section 3, we characterize s-goodness of a linear transformation in LMR via G-numbers. We consider the connections between the s-goodness, NSP, and RIP in Section 4. We eventually obtain that $\delta_{2s} < 0.472 \Rightarrow \mathscr{A}$ satisfying NSP $\Leftrightarrow \widehat{\gamma}_s(\mathscr{A}) < 1/2 \Leftrightarrow \gamma_s(\mathscr{A}) < 1 \Leftrightarrow \mathscr{A}$ is s-good.

Let $W \in \mathbb{R}^{m\times n}$, $r := \min\{m, n\}$, and let $W = U\,\mathrm{Diag}(\sigma(W))V^T$ be an SVD of W, where $U \in \mathbb{R}^{m\times r}$, $V \in \mathbb{R}^{n\times r}$, and $\mathrm{Diag}(\sigma(W))$ is the diagonal matrix of $\sigma(W) = (\sigma_1(W), \ldots, \sigma_r(W))^T$ which is the vector of the singular values of W. Also let $\Xi(W)$ denote the set of pairs of matrices (U, V) in the SVD of W; that is,

$$\Xi(W) := \left\{ (U,V) : U \in \mathbb{R}^{m\times r}, V \in \mathbb{R}^{n\times r}, \; W = U\,\mathrm{Diag}(\sigma(W))\,V^T \right\}. \tag{5}$$

For $s \in \{0, 1, 2, \ldots, r\}$, we say $W \in \mathbb{R}^{m\times n}$ is an *s-rank matrix* to mean that the rank of W is no more than s. For an s-rank matrix W, it is convenient to take $W = U_{m\times s} W_s V_{n\times s}^T$ as its SVD where $U_{m\times s} \in \mathbb{R}^{m\times s}$, $V_{n\times s} \in \mathbb{R}^{n\times s}$ are orthogonal matrices and $W_s = \mathrm{Diag}((\sigma_1(W), \ldots, \sigma_s(W))^T)$. For a vector $y \in \mathbb{R}^p$, let $\|\cdot\|_d$ be the *dual norm* of $\|\cdot\|$ specified by $\|y\|_d := \max_v\{\langle v, y\rangle : \|v\| \le 1\}$. In particular, $\|\cdot\|_\infty$ is the dual norm of $\|\cdot\|_1$ for a vector. Let $\|X\|$ denote the *spectral or the operator norm* of a matrix $X \in \mathbb{R}^{m\times n}$, that is, the largest singular value of X. In fact, $\|X\|$ is the dual norm of $\|X\|_*$. Let $\|X\|_F := \sqrt{\langle X, X\rangle} = \sqrt{\mathrm{Tr}(X^T X)}$ be the *Frobenius norm* of X, which is equal to the ℓ_2-norm of the vector of its singular values. We denote by X^T the *transpose* of X. For a linear transformation $\mathscr{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$, we denote by $\mathscr{A}^* : \mathbb{R}^p \to \mathbb{R}^{m\times n}$ the *adjoint* of $\mathscr{A}$.

2. Definitions and Basic Properties

2.1. Definitions. We first go over some concepts related to s-goodness of the linear transformation in LMR (RMP). These are extensions of those given for SSR (CMP) in [24].

Definition 1. Let $\mathscr{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ be a linear transformation and $s \in \{0, 1, 2, \ldots, r\}$. One says that $\mathscr{A}$ is s-good, if for every s-rank matrix $W \in \mathbb{R}^{m\times n}$, W is the unique optimal solution to the optimization problem

$$\min_{X\in\mathbb{R}^{m\times n}} \{\|X\|_* : \mathscr{A}X = \mathscr{A}W\}. \tag{6}$$

We denote by $s_*(\mathscr{A})$ the largest integer s for which $\mathscr{A}$ is s-good. Clearly, $s_*(\mathscr{A}) \in \{0, 1, \ldots, r\}$. To characterize s-goodness we introduce two useful *s-goodness constants*: γ_s and $\widehat{\gamma}_s$. We call γ_s and $\widehat{\gamma}_s$ G-numbers.

Definition 2. Let $\mathscr{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ be a linear transformation, $\beta \in [0, +\infty]$ and $s \in \{0, 1, 2, \ldots, r\}$. Then we have the following.

(i) G-number $\gamma_s(\mathscr{A}, \beta)$ is the infimum of $\gamma \ge 0$ such that for every matrix $X \in \mathbb{R}^{m\times n}$ with singular value decomposition $X = U_{m\times s} V_{n\times s}^T$ (i.e., s nonzero singular values, all equal to 1), there exists a vector $y \in \mathbb{R}^p$ such that

$$\|y\|_d \le \beta, \qquad \mathscr{A}^* y = U\,\mathrm{Diag}\left(\sigma\left(\mathscr{A}^* y\right)\right) V^T, \tag{7}$$

where $U = [U_{m\times s} \;\; U_{m\times(r-s)}]$, $V = [V_{n\times s} \;\; V_{n\times(r-s)}]$ are orthogonal matrices, and

$$\sigma_i\left(\mathscr{A}^* y\right) \begin{cases} = 1, & \text{if } \sigma_i(X) = 1, \\ \in [0, \gamma], & \text{if } \sigma_i(X) = 0, \end{cases} \quad i \in \{1, 2, \ldots, r\}. \tag{8}$$

If there does not exist such y for some X as above, we set $\gamma_s(\mathscr{A}, \beta) = +\infty$.

(ii) G-number $\widehat{\gamma}_s(\mathscr{A}, \beta)$ is the infimum of $\gamma \ge 0$ such that for every matrix $X \in \mathbb{R}^{m\times n}$ with s nonzero singular values, all equal to 1, there exists a vector $y \in \mathbb{R}^p$ such that $\mathscr{A}^* y$ and X share the same orthogonal row and column spaces:

$$\|y\|_d \le \beta, \qquad \|\mathscr{A}^* y - X\| \le \gamma. \tag{9}$$

If there does not exist such y for some X as above, we set $\gamma_s(\mathscr{A},\beta) = +\infty$ and to be compatible with the special case given by [24] we write $\gamma_s(\mathscr{A})$, $\widehat{\gamma}_s(\mathscr{A})$ instead of $\gamma_s(\mathscr{A},+\infty)$, $\widehat{\gamma}_s(\mathscr{A},+\infty)$, respectively.

From the above definition, we easily see that the set of values that γ takes is closed. Thus, when $\gamma_s(\mathscr{A},\beta) < +\infty$, for every matrix $X \in \mathbb{R}^{m\times n}$ with s nonzero singular values, all equal to 1, there exists a vector $y \in \mathbb{R}^p$ such that

$$\begin{aligned} &\|y\|_d \le \beta, \\ &\sigma_i(\mathscr{A}^* y) \begin{cases} = 1, & \text{if } \sigma_i(X) = 1, \\ \in [0, \gamma_s(\mathscr{A},\beta)], & \text{if } \sigma_i(X) = 0, \end{cases} \\ &\qquad i \in \{1,2,\ldots,r\}. \end{aligned} \tag{10}$$

Similarly, for every matrix $X \in \mathbb{R}^{m\times n}$ with s nonzero singular values, all equal to 1, there exists a vector $\widehat{y} \in \mathbb{R}^p$ such that $\mathscr{A}^*\widehat{y}$ and X share the same orthogonal row and column spaces:

$$\|\widehat{y}\|_d \le \beta, \qquad \|\mathscr{A}^*\widehat{y} - X\| \le \widehat{\gamma}_s(\mathscr{A},\beta). \tag{11}$$

Observing that the set $\{\mathscr{A}^* y : \|y\|_d \le \beta\}$ is convex, we obtain that if $\gamma_s(\mathscr{A},\beta) < +\infty$ then for every matrix X with at most s nonzero singular values and $\|X\| \le 1$ there exist vectors y satisfying (10) and there exist vectors $\widehat{y}$ satisfying (11).

2.2. Basic Properties of G-Numbers. In order to characterize the s-goodness of a linear transformation $\mathscr{A}$, we study the basic properties of G-numbers. We begin with the result that G-numbers $\gamma_s(\mathscr{A},\beta)$ and $\widehat{\gamma}_s(\mathscr{A},\beta)$ are convex nonincreasing functions of β.

Proposition 3. *For every linear transformation $\mathscr{A}$ and every $s \in \{0,1,\ldots,r\}$, G-numbers $\gamma_s(\mathscr{A},\beta)$ and $\widehat{\gamma}_s(\mathscr{A},\beta)$ are convex nonincreasing functions of $\beta \in [0,+\infty]$.*

Proof. We only need to demonstrate that the quantity $\gamma_s(\mathscr{A},\beta)$ is a convex nonincreasing function of $\beta \in [0,+\infty]$. It is evident from the definition that $\gamma_s(\mathscr{A},\beta)$ is nonincreasing for given $\mathscr{A}, s$. It remains to show that $\gamma_s(\mathscr{A},\beta)$ is a convex function of β. In other words, for every pair $\beta_1, \beta_2 \in [0,+\infty]$, we need to verify that

$$\begin{aligned} &\gamma_s(\mathscr{A}, \alpha\beta_1 + (1-\alpha)\beta_2) \\ &\quad \le \alpha\gamma_s(\mathscr{A},\beta_1) + (1-\alpha)\gamma_s(\mathscr{A},\beta_2), \quad \forall \alpha \in [0,1]. \end{aligned} \tag{12}$$

The above inequality follows immediately if one of β_1, β_2 is $+\infty$. Thus, we may assume $\beta_1, \beta_2 \in [0,+\infty)$. In fact, from the argument around (10) and the definition of $\gamma_s(\mathscr{A},\cdot)$, we know that for every matrix $X = U\,\mathrm{Diag}(\sigma(X))V^T$ with s nonzero singular values, all equal to 1, there exist vectors $y_1, y_2 \in \mathbb{R}^p$ such that for $k \in \{1,2\}$

$$\begin{aligned} &\|y_k\|_d \le \beta_k, \\ &\sigma_i(\mathscr{A}^* y_k) \begin{cases} = 1, & \text{if } \sigma_i(X) = 1, \\ \in [0, \gamma_s(\mathscr{A},\beta_k)], & \text{if } \sigma_i(X) = 0, \end{cases} \\ &\qquad i \in \{1,2,\ldots,r\}. \end{aligned} \tag{13}$$

It is immediate from (13) that $\|\alpha y_1 + (1-\alpha)y_2\|_d \le \alpha\beta_1 + (1-\alpha)\beta_2$. Moreover, from the above information on the singular values of $\mathscr{A}^* y_1$, $\mathscr{A}^* y_2$, we may set $\mathscr{A}^* y_k = X + Y_k$, $k \in \{1,2\}$ such that

$$\begin{aligned} X^T Y_k = 0, &\qquad X Y_k^T = 0, \\ \mathrm{rank}(Y_k) \le r - s, &\qquad \|Y_k\| \le \gamma_s(\mathscr{A},\beta_k). \end{aligned} \tag{14}$$

This implies that for every $\alpha \in [0,1]$

$$\begin{aligned} X^T[\alpha Y_1 + (1-\alpha)Y_2] = 0, \\ X[\alpha Y_1 + (1-\alpha)Y_2]^T = 0, \end{aligned} \tag{15}$$

and hence $\mathrm{rank}[\alpha Y_1 + (1-\alpha)Y_2] \le r - s$, X, and $[\alpha Y_1 + (1-\alpha)Y_2]$ have orthogonal row and column spaces. Thus, noting that $\mathscr{A}^*[\alpha y_1 + (1-\alpha)y_2] = X + \alpha Y_1 + (1-\alpha)Y_2$, we obtain that $\|\alpha y_1 + (1-\alpha)y_2\|_d \le \alpha\beta_1 + (1-\alpha)\beta_2$ and

$$\begin{aligned} &\sigma_i(\mathscr{A}^*(\alpha y_1 + (1-\alpha)y_2)) \\ &\quad = \begin{cases} 1, & \text{if } \sigma_i(X) = 1, \\ \sigma_i(\alpha Y_1 + (1-\alpha)Y_2), & \text{if } \sigma_i(X) = 0, \end{cases} \end{aligned} \tag{16}$$

for every $\alpha \in [0,1]$. Combining this with the fact

$$\begin{aligned} \|\alpha Y_1 + (1-\alpha)Y_2\| &\le \alpha\|Y_1\| + (1-\alpha)\|Y_2\| \\ &\le \alpha\gamma_s(\mathscr{A},\beta_1) + (1-\alpha)\gamma_s(\mathscr{A},\beta_2), \end{aligned} \tag{17}$$

we obtain the desired conclusion. □

The following observation that G-numbers $\gamma_s(\mathscr{A},\beta)$, $\widehat{\gamma}_s(\mathscr{A},\beta)$ are nondecreasing in s is immediate.

Proposition 4. *For every $s' \le s$, one has $\gamma_{s'}(\mathscr{A},\beta) \le \gamma_s(\mathscr{A},\beta)$, $\widehat{\gamma}_{s'}(\mathscr{A},\beta) \le \widehat{\gamma}_s(\mathscr{A},\beta)$.*

We further investigate the relationship between the G-numbers $\gamma_s(\mathscr{A},\beta)$ and $\widehat{\gamma}_s(\mathscr{A},\beta)$.

Proposition 5. *Let $\mathscr{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ be a linear transformation, $\beta \in [0,+\infty]$, and $s \in \{0,1,2,\ldots,r\}$. Then one has*

$$\begin{aligned} &\gamma := \gamma_s(\mathscr{A},\beta) < 1 \Longrightarrow \widehat{\gamma}_s\left(\mathscr{A}, \frac{1}{1+\gamma}\beta\right) \\ &\quad = \frac{\gamma}{1+\gamma} < \frac{1}{2}, \\ &\widehat{\gamma} := \widehat{\gamma}_s(\mathscr{A},\beta) < \frac{1}{2} \Longrightarrow \gamma_s\left(\mathscr{A}, \frac{1}{1-\widehat{\gamma}}\beta\right) \\ &\quad = \frac{\widehat{\gamma}}{1-\widehat{\gamma}} < 1. \end{aligned} \tag{18}$$

Proof. Let $\gamma := \gamma_s(\mathscr{A},\beta) < 1$. Then, for every matrix $Z \in \mathbb{R}^{m\times n}$ with s nonzero singular values, all equal to 1, there exists $y \in \mathbb{R}^p$, $\|y\|_d \le \beta$, such that $\mathscr{A}^* y = Z + W$, where $\|W\| \le \gamma$ and W and Z have orthogonal row and column spaces.

For a given pair Z, y as above, take $\tilde{y} := (1/(1+\gamma))y$. Then we have $\|\tilde{y}\|_d \le (1/(1+\gamma))\beta$ and

$$\left\|\mathscr{A}^*\tilde{y} - Z\right\| \le \max\left\{1 - \frac{1}{1+\gamma}, \frac{\gamma}{1+\gamma}\right\} = \frac{\gamma}{1+\gamma}, \tag{19}$$

where the first term under the maximum comes from the fact that $\mathscr{A}^* y$ and Z agree on the subspace corresponding to the nonzero singular values of Z. Therefore, we obtain

$$\hat{\gamma}_s\left(\mathscr{A}, \frac{1}{1+\gamma}\beta\right) \le \frac{\gamma}{1+\gamma} < \frac{1}{2}. \tag{20}$$

Now, we assume that $\hat{\gamma} := \hat{\gamma}_s(\mathscr{A}, \beta) < 1/2$. Fix orthogonal matrices $U \in \mathbb{R}^{m\times r}$, $V \in \mathbb{R}^{n\times r}$. For an s-element subset J of the index set $\{1, 2, \ldots, r\}$, we define a set S_J with respect to orthogonal matrices U, V as

$$S_J := \left\{x \in \mathbb{R}^r : \exists y \in \mathbb{R}^p, \|y\|_d \le \beta, \right. \\ \left. \mathscr{A}^* y = U\,\mathrm{Diag}(x)\,V^T \text{ with } |x_i| \le \hat{\gamma} \text{ for } i \in \overline{J}\right\}. \tag{21}$$

In the above, $\overline{J}$ denotes the complement of J. It is immediately seen that S_J is a closed convex set in $\mathbb{R}^r$. Moreover, we have the following

Claim 1. S_J contains the $\|\cdot\|_\infty$-ball of radius $(1-\hat{\gamma})$ centered at the origin in $\mathbb{R}^r$.

Proof. Note that S_J is closed and convex. Moreover, S_J is the direct sum of its projections onto the pair of subspaces

$$L_J := \left\{x \in \mathbb{R}^r : x_i = 0, i \in \overline{J}\right\} \\ \text{and its orthogonal complement} \\ L_J^{\perp} = \left\{x \in \mathbb{R}^r : x_i = 0, i \in J\right\}. \tag{22}$$

Let Q denote the projection of S_J onto L_J. Then, Q is closed and convex (because of the direct sum property above and the fact that S_J is closed and convex). Note that L_J can be naturally identified with $\mathbb{R}^s$, and our claim is the image $\overline{Q} \subset \mathbb{R}^s$ of Q under this identification that contains the $\|\cdot\|_\infty$-ball B_s of radius $(1-\hat{\gamma})$ centered at the origin in $\mathbb{R}^s$. For a contradiction, suppose B_s is not contained in $\overline{Q}$. Then there exists $v \in B_s \setminus \overline{Q}$. Since $\overline{Q}$ is closed and convex, by a separating hyperplane theorem, there exists a vector $u \in \mathbb{R}^s$, $\|u\|_1 = 1$ such that

$$u^T v > u^T v' \quad \text{for every } v' \in \overline{Q}. \tag{23}$$

Let $z \in \mathbb{R}^r$ be defined by

$$z_i := \begin{cases} 1, & i \in J, \\ 0, & \text{otherwise.} \end{cases} \tag{24}$$

By definition of $\hat{\gamma} = \hat{\gamma}_s(\mathscr{A}, \beta)$, for s-rank matrix $U\,\mathrm{Diag}(z)V^T$, there exists $y \in \mathbb{R}^p$ such that $\|y\|_d \le \beta$ and

$$\mathscr{A}^* y = U\,\mathrm{Diag}(z)\,V^T + W, \tag{25}$$

where W and $U\,\mathrm{Diag}(z)V^T$ have the same orthogonal row and column spaces, $\|\mathscr{A}^* y - U\,\mathrm{Diag}(z)V^T\| \le \hat{\gamma}$ and $\|\sigma(\mathscr{A}^* y) - z\|_\infty \le \hat{\gamma}$. Together with the definitions of S_J and $\overline{Q}$, this means that $\overline{Q}$ contains a vector $\overline{v}$ with $|\overline{v}_i - \mathrm{sign}(u_i)| \le \hat{\gamma}$, $\forall i \in \{1, 2, \ldots, s\}$. Therefore,

$$u^T\overline{v} \ge \sum_{i=1}^{s} |u_i|\,(1-\hat{\gamma}) = (1-\hat{\gamma})\,\|u\|_1 = 1 - \hat{\gamma}. \tag{26}$$

By $v \in B_s$ and the definition of u, we obtain

$$1 - \hat{\gamma} \ge \|v\|_\infty = \|u\|_1\|v\|_\infty \ge u^T v > u^T\overline{v} \ge 1 - \hat{\gamma}, \tag{27}$$

where the strict inequality follows from the facts that $\overline{v} \in \overline{Q}$ and u separates v from $\overline{Q}$. The above string of inequalities is a contradiction, and hence the desired claim holds.

Using the above claim, we conclude that for every $J \subseteq \{1, 2, \ldots, r\}$ with cardinality s, there exists an $x \in S_J$ such that $x_i = (1-\hat{\gamma})$, for all $i \in J$. From the definition of S_J, we obtain that there exists $y \in \mathbb{R}^p$ with $\|y\|_d \le (1-\hat{\gamma})^{-1}\beta$ such that

$$\mathscr{A}^* y = U\,\mathrm{Diag}\left(\sigma\left(\mathscr{A}^* y\right)\right)V^T, \tag{28}$$

where $\sigma_i(\mathscr{A}^* y) = (1-\hat{\gamma})^{-1}x_i = 1$ if $i \in J$, and $\sigma_i(\mathscr{A}^* y)_i \le (1-\hat{\gamma})^{-1}\hat{\gamma}$ if $i \in \overline{J}$. Thus, we obtain that

$$\hat{\gamma}_s := \hat{\gamma}_s(\mathscr{A}, \beta) < \frac{1}{2} \Longrightarrow \gamma_s\left(\mathscr{A}, \frac{1}{1-\hat{\gamma}}\beta\right) \le \frac{\hat{\gamma}}{1-\hat{\gamma}} < 1. \tag{29}$$

To conclude the proof, we need to prove that the inequalities we established

$$\hat{\gamma}_s\left(\mathscr{A}, \frac{1}{1+\hat{\gamma}}\beta\right) \le \frac{\gamma}{1+\gamma}, \qquad \gamma_s\left(\mathscr{A}, \frac{1}{1-\hat{\gamma}}\beta\right) \le \frac{\hat{\gamma}}{1+\hat{\gamma}} \tag{30}$$

are both equations. This is straightforward by an argument similar to the one in the proof of [24, Theorem 1]. We omit it for the sake of brevity. □

We end this section with a simple argument which illustrates that for a given pair $(\mathscr{A}, s)$, $\gamma_s(\mathscr{A}, \beta) = \gamma_s(\mathscr{A})$ and $\hat{\gamma}_s(\mathscr{A}, \beta) = \hat{\gamma}_s(\mathscr{A})$, for all β large enough.

Proposition 6. *Let $\mathscr{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ be a linear transformation and $\beta \in [0, +\infty]$. Assume that for some $\rho > 0$, the image of the unit $\|\cdot\|_*$-ball in $\mathbb{R}^{m\times n}$ under the mapping $X \mapsto \mathscr{A}X$ contains the ball $B = \{x \in \mathbb{R}^p : \|x\|_1 \le \rho\}$. Then for every $s \in \{1, 2, \ldots, r\}$*

$$\beta \ge \frac{1}{\rho}, \qquad \gamma_s(\mathscr{A}) < 1 \Longrightarrow \gamma_s(\mathscr{A}, \beta) = \gamma_s(\mathscr{A}), \\ \beta \ge \frac{1}{\rho}, \qquad \hat{\gamma}_s(\mathscr{A}) < \frac{1}{2} \Longrightarrow \hat{\gamma}_s(\mathscr{A}, \beta) = \hat{\gamma}_s(\mathscr{A}). \tag{31}$$

Proof. Fix $s \in \{1, 2, \ldots, r\}$. We only need to show the first implication. Let $\gamma := \gamma_s(\mathscr{A}) < 1$. Then for every matrix $W \in \mathbb{R}^{m\times n}$ with its SVD $W = U_{m\times s}V_{n\times s}^T$, there exists a vector $y \in \mathbb{R}^p$ such that

$$\|y\|_d \le \beta, \qquad \mathscr{A}^* y = U \operatorname{Diag}\left(\sigma\left(\mathscr{A}^* y\right)\right) V^T, \tag{32}$$

where $U = [U_{m\times s} \;\; U_{m\times(r-s)}]$, $V = [V_{n\times s} \;\; V_{n\times(r-s)}]$ are orthogonal matrices, and

$$\sigma_i\left(\mathscr{A}^* y\right) \begin{cases} = 1, & \text{if } \sigma_i(W) = 1, \\ \in [0, \gamma], & \text{if } \sigma_i(W) = 0, \end{cases} \qquad i \in \{1, 2, \ldots, r\}. \tag{33}$$

Clearly, $\|\mathscr{A}^* y\| \le 1$. That is,

$$\begin{aligned} 1 \ge \left\|\mathscr{A}^* y\right\| &= \max_{X\in\mathbb{R}^{m\times n}} \left\{\langle X, \mathscr{A}^* y\rangle : \|X\|_* \le 1\right\} \\ &= \max_{X\in\mathbb{R}^{m\times n}} \left\{\langle u, y\rangle : u = \mathscr{A}X, \|X\|_* \le 1\right\}. \end{aligned} \tag{34}$$

From the inclusion assumption, we obtain that

$$\begin{aligned} &\max_{X\in\mathbb{R}^{m\times n}} \left\{\langle u, y\rangle : u = \mathscr{A}X, \|X\|_* \le 1\right\} \\ &\quad \ge \max_{u\in\mathbb{R}^p} \left\{\langle u, y\rangle : \|u\|_1 \le \rho\right\} = \rho\|y\|_\infty = \rho\|y\|_d. \end{aligned} \tag{35}$$

Combining the above two strings of relations, we derive the desired conclusion. □

3. s-Goodness and G-Numbers

We first give the following characterization result of s-goodness of a linear transformation $\mathscr{A}$ via the G-number $\gamma_s(\mathscr{A})$, which explains the importance of $\gamma_s(\mathscr{A})$ in LMR.

Theorem 7. *Let $\mathscr{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ be a linear transformation, and s be an integer $s \in \{0, 1, 2, \ldots, r\}$. Then $\mathscr{A}$ is s-good if and only if $\gamma_s(\mathscr{A}) < 1$.*

Proof. Suppose $\mathscr{A}$ is s-good. Let $W \in \mathbb{R}^{m\times n}$ be a matrix of rank $s \in \{1, 2, \ldots, r\}$. Without loss of generality, let $W = U_{m\times s}W_sV_{n\times s}^T$ be its SVD where $U_{m\times s} \in \mathbb{R}^{m\times s}$, $V_{n\times s} \in \mathbb{R}^{n\times s}$ are orthogonal matrices and $W_s = \operatorname{Diag}((\sigma_1(W), \ldots, \sigma_s(W))^T)$. By the definition of s-goodness of $\mathscr{A}$, W is the unique solution to the optimization problem (6). Using the first-order optimality conditions, we obtain that there exists $y \in \mathbb{R}^p$ such that the function $f_y(x) = \|X\|_* - y^T[\mathscr{A}X - \mathscr{A}W]$ attains its minimum value over $X \in \mathbb{R}^{m\times n}$ at $X = W$. So $0 \in \partial f_y(W)$ or $\mathscr{A}^* y \in \partial\|W\|_*$. Using the fact (see, e.g., [27])

$$\begin{aligned} \partial\|W\|_* = \Big\{&U_{m\times s}V_{n\times s}^T + M : W \text{ and } M \text{ have orthogonal} \\ &\text{row and column spaces, and } \|M\| \le 1\Big\}, \end{aligned} \tag{36}$$

it follows that there exist matrices $U_{m\times(r-s)}$, $V_{n\times(r-s)}$ such that $\mathscr{A}^* y = U \operatorname{Diag}(\sigma_i(\mathscr{A}^* y))V^T$ where $U = [U_{m\times s} \;\; U_{m\times(r-s)}]$, $V = [V_{n\times s} \;\; V_{n\times(r-s)}]$ are orthogonal matrices and

$$\sigma_i\left(\mathscr{A}^* y\right) \begin{cases} = 1, & \text{if } i \in J, \\ \in [0, 1], & \text{if } i \in \overline{J}, \end{cases} \tag{37}$$

where $J := \{i : \sigma_i(W) \ne 0\}$ and $\overline{J} := \{1, 2, \ldots, r\} \setminus J$. Therefore, the optimal objective value of the optimization problem

$$\min_{y,\gamma} \left\{\gamma : \mathscr{A}^* y \in \partial\|W\|_*, \sigma_i\left(\mathscr{A}^* y\right) \begin{cases} = 1, & \text{if } i \in J, \\ \in [0, \gamma], & \text{if } i \in \overline{J}, \end{cases}\right\} \tag{38}$$

is at most one. For the given W with its SVD $W = U_{m\times s}W_sV_{n\times s}^T$, let

$$\begin{aligned} \Pi := \Big\{&\operatorname{conv}\{M \in \mathbb{R}^{m\times n} : \text{the SVD of } M \text{ is} \\ &M = \left[U_{m\times s} \;\; \overline{U}_{m\times(r-s)}\right] \begin{pmatrix} 0_s & 0 \\ 0 & \sigma(M) \end{pmatrix} \left[V_{n\times s} \;\; \overline{V}_{n\times(r-s)}\right]^T\Big\}. \end{aligned} \tag{39}$$

It is easy to see that Π is a subspace and its normal cone (in the sense of variational analysis, see, e.g., [28] for details) is specified by $\Pi^\perp$. Thus, the above problem (38) is equivalent to the following convex optimization problem with set constraint:

$$\min_{y,M} \left\{\|M\| : \mathscr{A}^* y - U_{m\times s}V_{n\times s}^T - M = 0,\ M \in \Pi\right\}. \tag{40}$$

We will show that the optimal value is less than 1. For a contradiction, suppose that the optimal value is one. Then, by [28, Theorem 10.1 and Exercise 10.52], there exists a Lagrange multiplier $D \in \mathbb{R}^{m\times n}$ such that the function

$$L(y, M) = \|M\| + \left\langle D, \mathscr{A}^* y - U_{m\times s}V_{n\times s}^T - M\right\rangle + \delta_\Pi(M) \tag{41}$$

has unconstrained minimum in (y, M) equal to 1, where $\delta_\Pi(\cdot)$ is the indicator function of Π. Let (y^*, M^*) be an optimal solution. Then, by the optimality condition $0 \in \partial L$, we obtain that

$$0 \in \partial_y L\left(y^*, M^*\right), \qquad 0 \in \partial_M L\left(y^*, M^*\right). \tag{42}$$

Direct calculation yields that

$$\mathscr{A}D = 0, \qquad 0 \in -D + \partial\left\|M^*\right\| + \Pi^\perp. \tag{43}$$

Then there exist $D_J \in \Pi^\perp$ and $D_{\overline{J}} \in \partial\|M^*\|$ such that $D = D_J + D_{\overline{J}}$. Notice that [29, Corollary 6.4] implies that for $D_{\overline{J}} \in \partial\|M^*\|$, $D_{\overline{J}} \in \Pi$ and $\|D_{\overline{J}}\|_* \le 1$. Therefore, $\langle D, U_{m\times s}V_{n\times s}^T\rangle = \langle D_J, U_{m\times s}V_{n\times s}^T\rangle$ and $\langle D, M^*\rangle = \langle D_{\overline{J}}, M^*\rangle$. Moreover, $\langle D_{\overline{J}}, M^*\rangle \le \|M^*\|$ by the definition of the dual

norm of $\|\cdot\|$. This together with the facts $\mathscr{A}D = 0$, $D_J \in \Pi^{\perp}$ and $D_{\overline{J}} \in \partial\|M^*\| \subseteq \Pi$ yields

$$\begin{aligned} L(y^*, M^*) &= \|M^*\| - \langle D_{\overline{J}}, M^*\rangle + \langle D, \mathscr{A}^* y^*\rangle \\ &\quad - \langle D_J, U_{m\times s} V_{n\times s}^T\rangle + \delta_{\Pi}(M^*) \\ &\geq -\langle D_J, U_{m\times s} V_{n\times s}^T\rangle + \delta_{\Pi}(M^*). \end{aligned} \tag{44}$$

Thus, the minimum value of $L(y, M)$ is attained, $L(y^*, M^*) = -\langle D_J, U_{m\times s} V_{n\times s}^T\rangle$, when $M^* \in \Pi$, $\langle D_{\overline{J}}, M^*\rangle = \|M^*\|$. We obtain that $\|D_{\overline{J}}\|_* = 1$. By assumption, $1 = L(y^*, M^*) = -\langle D_J, U_{m\times s} V_{n\times s}^T\rangle$. That is, $\sum_{i=1}^{s} (U_{m\times s}^T D V_{n\times s})_{ii} = -1$. Without loss of generality, let SVD of the optimal M^* be $M^* = \widetilde{U}\begin{pmatrix} 0_s & 0 \\ 0 & \sigma(M^*)\end{pmatrix}\widetilde{V}^T$, where $\widetilde{U} := [U_{m\times s}\ \ \widetilde{U}_{m\times(r-s)}]$ and $\widetilde{V} := [V_{n\times s}\ \ \widetilde{V}_{n\times(r-s)}]$. From the above arguments, we obtain that

(i) $\mathscr{A}D = 0$,

(ii) $\sum_{i=1}^{s} (U_{m\times s}^T D V_{n\times s})_{ii} = \sum_{i\in J} (\widetilde{U}^T D \widetilde{V})_{ii} = -1$,

(iii) $\sum_{i\in\overline{J}} (\widetilde{U}^T D \widetilde{V})_{ii} = 1$.

Clearly, for every $t \in \mathbb{R}$, the matrices $X_t := W + tD$ are feasible in (6). Note that

$$\begin{aligned} W &= U_{m\times s} W_s V_{n\times s}^T \\ &= \begin{bmatrix} U_{m\times s} & \widetilde{U}_{m\times(r-s)}\end{bmatrix}\begin{pmatrix} W_s & 0 \\ 0 & 0\end{pmatrix}\begin{bmatrix} V_{n\times s} & \widetilde{V}_{n\times(r-s)}\end{bmatrix}^T. \end{aligned} \tag{45}$$

Then, $\|W\|_* = \|\widetilde{U}^T W \widetilde{V}\|_* = \mathrm{Tr}(\widetilde{U}^T W \widetilde{V})$. From the above equations, we obtain that $\|X_t\|_* = \|W\|_*$ for all small enough $t > 0$ (since $\sigma_i(W) > 0$, $i \in \{1, 2, \ldots, s\}$). Noting that W is the unique optimal solution to (6), we have $X_t = W$, which means that $(\widetilde{U}^T D \widetilde{V})_{ii} = 0$ for $i \in J$. This is a contradiction, and hence the desired conclusion holds.

We next prove that $\mathscr{A}$ is s-good if $\gamma_s(\mathscr{A}) < 1$. That is, we let W be an s-rank matrix and we show that W is the unique optimal solution to (6). Without loss of generality, let W be a matrix of rank $s' \neq 0$ and $U_{m\times s'} W_{s'} V_{n\times s'}^T$ its SVD, where $U_{m\times s'} \in \mathbb{R}^{m\times s'}$, $V_{n\times s'} \in \mathbb{R}^{n\times s'}$ are orthogonal matrices and $W_{s'} = \mathrm{Diag}((\sigma_1(W), \ldots, \sigma_{s'}(W))^T)$. It follows from Proposition 4 that $\gamma_{s'}(\mathscr{A}) \leq \gamma_s(\mathscr{A}) < 1$. By the definition of $\gamma_s(\mathscr{A})$, there exists $y \in \mathbb{R}^p$ such that $\mathscr{A}^* y = U\,\mathrm{Diag}(\sigma(\mathscr{A}^* y))V^T$, where $U = [U_{m\times s'}\ \ U_{m\times(r-s')}]$, $V = [V_{n\times s'}\ \ V_{n\times(r-s')}]$, and

$$\sigma_i(\mathscr{A}^* y)\begin{cases} = 1, & \text{if } \sigma_i(W) \neq 0, \\ \in [0, 1), & \text{if } \sigma_i(W) = 0. \end{cases} \tag{46}$$

Now, we have the optimization problem of minimizing the function

$$\begin{aligned} f(X) &= \|X\|_* - y^T[\mathscr{A}X - \mathscr{A}W] \\ &= \|X\|_* - \langle \mathscr{A}^* y, X\rangle + \|W\|_* \end{aligned} \tag{47}$$

over all $X \in \mathbb{R}^{m\times n}$ such that $\mathscr{A}X = \mathscr{A}W$. Note that $\langle \mathscr{A}^* y, X\rangle \leq \|X\|_*$ by $\|\mathscr{A}^* y\| \leq 1$ and the definition of dual norm. So $f(X) \geq \|X\|_* - \|X\|_* + \|W\|_* = \|W\|_*$ and this function attains its unconstrained minimum in X at $X = W$. Hence $X = W$ is an optimal solution to (6). It remains to show that this optimal solution is unique. Let Z be another optimal solution to the problem. Then $f(Z) - f(W) = \|Z\|_* - y^T\mathscr{A}Z = \|Z\|_* - \langle \mathscr{A}^* y, Z\rangle = 0$. This together with the fact $\|\mathscr{A}^* y\| \leq 1$ implies that there exist SVDs for $\mathscr{A}^* y$ and Z such that

$$\begin{aligned} \mathscr{A}^* y &= \widetilde{U}\,\mathrm{Diag}(\sigma(\mathscr{A}^* y))\widetilde{V}^T, \\ Z &= \widetilde{U}\,\mathrm{Diag}(\sigma(Z))\widetilde{V}^T, \end{aligned} \tag{48}$$

where $\widetilde{U} \in \mathbb{R}^{m\times r}$ and $\widetilde{V} \in \mathbb{R}^{n\times r}$ are orthogonal matrices, and $\sigma_i(Z) = 0$ if $\sigma_i(\mathscr{A}^* y) \neq 1$. Thus, for $\sigma_i(\mathscr{A}^* y) = 0$, for all $i \in \{s' + 1, \ldots, r\}$, we must have $\sigma_i(Z) = \sigma_i(W) = 0$. By the two forms of SVDs of $\mathscr{A}^* y$ as above, $U_{m\times s'} V_{n\times s'}^T = \widetilde{U}_{m\times s'} \widetilde{V}_{n\times s'}^T$ where $\widetilde{U}_{m\times s'}, \widetilde{V}_{n\times s'}^T$ are the corresponding submatrices of $\widetilde{U}, \widetilde{V}$, respectively. Without loss of generality, let

$$\begin{aligned} U &= [u_1, u_2, \ldots, u_r], \qquad V = [v_1, v_2, \ldots, v_r], \\ \widetilde{U} &= [\widetilde{u}_1, \widetilde{u}_2, \ldots, \widetilde{u}_r], \qquad \widetilde{V} = [\widetilde{v}_1, \widetilde{v}_2, \ldots, \widetilde{v}_r], \end{aligned} \tag{49}$$

where $u_j = \widetilde{u}_j$ and $v_j = \widetilde{v}_j$ for the corresponding index $j \in \{i : \sigma_i(\mathscr{A}^* y) = 0,\ i \in \{s' + 1, \ldots, r\}\}$. Then we have

$$Z = \sum_{i=1}^{s'} \sigma_i(Z)\,\widetilde{u}_i \widetilde{v}_i^T, \qquad W = \sum_{i=1}^{s'} \sigma_i(W)\,u_i v_i^T. \tag{50}$$

From $U_{m\times s'} V_{n\times s'}^T = \widetilde{U}_{m\times s'} \widetilde{V}_{n\times s'}^T$, we obtain that

$$\sum_{i=s'+1}^{r} \sigma_i(\mathscr{A}^* y)\,\widetilde{u}_i \widetilde{v}_i^T = \sum_{i=s'+1}^{r} \sigma_i(\mathscr{A}^* y)\,u_i v_i^T. \tag{51}$$

Therefore, we deduce

$$\begin{aligned} &\sum_{i=s'+1, \sigma_i(\mathscr{A}^* y)\neq 0}^{r} \sigma_i(\mathscr{A}^* y)\,\widetilde{u}_i \widetilde{v}_i^T + \sum_{i=s'+1, \sigma_i(\mathscr{A}^* y)=0}^{r} \widetilde{u}_i \widetilde{v}_i^T \\ &= \sum_{i=s'+1, \sigma_i(\mathscr{A}^* y)\neq 0}^{r} \sigma_i(\mathscr{A}^* y)\,u_i v_i^T + \sum_{i=s'+1, \sigma_i(\mathscr{A}^* y)=0}^{r} u_i v_i^T \\ &=: \Omega. \end{aligned} \tag{52}$$

Clearly, the rank of Ω is no less than $r - s' \geq r - s$. From the orthogonality property of U, V and $\widetilde{U}, \widetilde{V}$, we easily derive that

$$\Omega^T \widetilde{u}_i \widetilde{v}_i^T = 0, \quad \Omega^T u_i v_i^T = 0, \quad \forall i \in \left\{1, 2, \ldots, s'\right\}. \tag{53}$$

Thus, we obtain $\Omega^T(Z - W) = 0$, which implies that the rank of the matrix $Z - W$ is no more than s. Since $\gamma_s(\mathscr{A}) < 1$, there exists $\widetilde{y}$ such that

$$\sigma_i(\mathscr{A}^* \widetilde{y})\begin{cases} = 1, & \text{if } \sigma_i(Z - W) \neq 0, \\ \in [0, 1) & \text{if } \sigma_i(Z - W) = 0. \end{cases} \tag{54}$$

Therefore, $0 = \widetilde{y}^T \mathscr{A}(Z - W) = \langle \mathscr{A}^* \widetilde{y}, Z - W\rangle = \|Z - W\|_*$. Then $Z = W$. □

For the G-number $\widehat{\gamma}_s(\mathscr{A})$, we directly obtain the following equivalent theorem of s-goodness from Proposition 5 and Theorem 7.

Theorem 8. *Let $\mathscr{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ be a linear transformation, and $s \in \{1, 2, \ldots, r\}$. Then $\mathscr{A}$ is s-good if and only if $\widehat{\gamma}_s(\mathscr{A}) < 1/2$.*

4. s-Goodness, NSP, and RIP

This section deals with the connections between s-goodness, the null space property (NSP), and the restricted isometry property (RIP). We start with establishing the equivalence of NSP and G-number $\widehat{\gamma}_s(\mathscr{A}) < 1/2$. Here, we say $\mathscr{A}$ *satisfies NSP* if for every nonzero matrix $X \in \mathrm{Null}(\mathscr{A})$ with the SVD $X = U\,\mathrm{Diag}(\sigma(X))V^T$, then we have

$$\sum_{i=1}^{s}\sigma_i(X) < \sum_{i=s+1}^{r}\sigma_i(X). \tag{55}$$

For further details, see, for example, [14, 19–21] and references therein.

Proposition 9. *For the linear transformation $\mathscr{A}$, $\widehat{\gamma}_s(\mathscr{A}) < 1/2$ if and only if $\mathscr{A}$ satisfies NSP.*

Proof. We first give an equivalent representation of the G-number $\widehat{\gamma}_s(\mathscr{A}, \beta)$. We define a compact convex set first:

$$P_s := \{Z \in \mathbb{R}^{m\times n} : \|Z\|_* \le s, \|Z\| \le 1\}. \tag{56}$$

Let $B_\beta := \{y \in \mathbb{R}^p : \|y\|_d \le \beta\}$ and $B := \{X \in \mathbb{R}^{m\times n} : \|X\| \le 1\}$. By definition, $\widehat{\gamma}_s(\mathscr{A}, \beta)$ is the smallest γ such that the closed convex set $C_{\gamma,\beta} := \mathscr{A}^* B_\beta + \gamma B$ contains all matrices with s nonzero singular values, all equal to 1. Equivalently, $C_{\gamma,\beta}$ contains the convex hull of these matrices, namely, P_s. Note that γ satisfies the inclusion $P_s \subseteq C_{\gamma,\beta}$ if and only if for every $X \in \mathbb{R}^{m\times n}$

$$\begin{aligned}\max_{Z\in P_s}\langle Z, X\rangle &\le \max_{Y\in C_{\gamma,\beta}}\langle Y, X\rangle \\ &= \max_{y\in\mathbb{R}^p, W\in\mathbb{R}^{m\times n}} \{\langle X, \mathscr{A}^* y\rangle + \gamma\langle X, W\rangle \\ &\qquad : \|y\|_d \le \beta, \|W\| \le 1\} \\ &= \beta\|\mathscr{A}X\| + \gamma\|X\|_*.\end{aligned} \tag{57}$$

For the above, we adopt the convention that whenever $\beta = +\infty$, $\beta\|\mathscr{A}X\|$ is defined to be $+\infty$ or 0 depending on whether $\|\mathscr{A}X\| > 0$ or $\|\mathscr{A}X\| = 0$. Thus, $P_s \subseteq C_{\gamma,\beta}$ if and only if $\max_{Z\in P_s}\{\langle Z, X\rangle - \beta\|\mathscr{A}X\|\} \le \gamma\|X\|_*$. Using the homogeneity of this last relation with respect to X, the above is equivalent to

$$\max_{Z,X}\{\langle Z, X\rangle - \beta\|\mathscr{A}X\| : Z \in P_s, \|X\|_* \le 1\} \le \gamma. \tag{58}$$

Therefore, we obtain $\widehat{\gamma}_s(\mathscr{A}, \beta) = \max_{Z,X}\{\langle Z, X\rangle - \beta\|\mathscr{A}X\| : Z \in P_s, \|X\|_* \le 1\}$. Furthermore,

$$\widehat{\gamma}_s(\mathscr{A}) = \max_{Z,X}\{\langle Z, X\rangle : Z \in P_s, \|X\|_* \le 1, \mathscr{A}X = 0\}. \tag{59}$$

For $X \in \mathbb{R}^{m\times n}$ with $\mathscr{A}X = 0$, let $X = U\,\mathrm{Diag}(\sigma(X))V^T$ be its SVD. Then, we obtain the sum of the s largest singular values of X as

$$\|X\|_{s,*} = \max_{Z\in P_s}\langle Z, X\rangle. \tag{60}$$

From (59), we immediately obtain that $\widehat{\gamma}_s(\mathscr{A})$ is the best upper bound on $\|X\|_{s,*}$ of matrices $X \in \mathrm{Null}(\mathscr{A})$ such that $\|X\|_* \le 1$. Therefore, $\widehat{\gamma}_s(\mathscr{A}) < 1/2$ implies that the maximum value of $\|\cdot\|_{s,*}$-norms of matrices $X \in \mathrm{Null}(\mathscr{A})$ with $\|X\|_* = 1$ is less than $1/2$. That is, $\sum_{i=1}^{s}\sigma_i(X) < 1/2\sum_{i=1}^{r}\sigma_i(X)$. Thus, $\sum_{i=1}^{s}\sigma_i(X) < \sum_{i=s+1}^{r}\sigma_i(X)$ and hence $\mathscr{A}$ satisfies NSP. Now, it is easy to see that $\mathscr{A}$ satisfies NSP if and only if $\widehat{\gamma}_s(\mathscr{A}) < 1/2$. □

Next, we consider the connection between restricted isometry constants and G-number of the linear transformation in LMR. It is well known that, for a nonsingular matrix (transformation) $T \in \mathbb{R}^{p\times p}$, the RIP constants of $\mathscr{A}$ and $T\mathscr{A}$ can be very different, as shown by Zhang [30] for the vector case. However, the s-goodness properties of $\mathscr{A}$ and $T\mathscr{A}$ are always the same for a nonsingular transformation $T \in \mathbb{R}^{p\times p}$ (i.e., s-goodness properties enjoy scale invariance in this sense). Recall that the *s-restricted isometry constant* δ_s of a linear transformation $\mathscr{A}$ is defined as the smallest constant such that the following holds for all s-rank matrices $X \in \mathbb{R}^{m\times n}$:

$$(1 - \delta_s)\|X\|_F^2 \le \|\mathscr{A}X\|_2^2 \le (1 + \delta_s)\|X\|_F^2. \tag{61}$$

In this case, we say $\mathscr{A}$ possesses the *RI (δ_s)-property (RIP)* as in the CS context. For details, see [4, 31–34] and the references therein.

Proposition 10. *Let $\mathscr{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ be a linear transformation and $s \in \{0, 1, 2, \ldots, r\}$. For any nonsingular transformation $T \in \mathbb{R}^{p\times p}$, $\widehat{\gamma}_s(\mathscr{A}) = \widehat{\gamma}_s(T\mathscr{A})$.*

Proof. It follows from the nonsingularity of T that $\{X : \mathscr{A}X = 0\} = \{X : T\mathscr{A}X = 0\}$. Then, by the equivalent representation of the G-number $\widehat{\gamma}_s(\mathscr{A}, \beta)$ in (59),

$$\begin{aligned}\widehat{\gamma}_s(\mathscr{A}) &= \max_{Z,X}\{\langle Z, X\rangle : Z \in P_s, \|X\|_* \le 1, \mathscr{A}X = 0\} \\ &= \max_{Z,X}\{\langle Z, X\rangle : Z \in P_s, \|X\|_* \le 1, T\mathscr{A}X = 0\} \\ &= \widehat{\gamma}_s(T\mathscr{A}).\end{aligned} \tag{62}$$

□

For the RIP constant δ_{2s}, Oymak et al. [21] gave the current best bound on the restricted isometry constant $\delta_{2s} < 0.472$, where they proposed a general technique for translating results from SSR to LMR. Together with the above arguments, we immediately obtain the following theorem.

Theorem 11. *$\delta_{2s} < 0.472 \Rightarrow \mathscr{A}$ satisfying NSP $\Leftrightarrow \widehat{\gamma}_s(\mathscr{A}) < 1/2 \Leftrightarrow \gamma_s(\mathscr{A}) < 1 \Leftrightarrow \mathscr{A}$ is s-good.*

Proof. It follows from [21, Theorem 1], Proposition 9, and Theorems 7 and 8. □

The above theorem says that s-goodness is a necessary and sufficient condition for recovering the low-rank solution exactly via nuclear norm minimization.

5. Conclusion

In this paper, we have shown that s-goodness of the linear transformation in LMR is a necessary and sufficient conditions for exact s-rank matrix recovery via the nuclear norm minimization, which is equivalent to the null space property. Our analysis is based on the two characteristic s-goodness constants, γ_s and $\widehat{\gamma}_s$, and the variational property of matrix norm in convex optimization. This shows that s-goodness is an elegant concept for low-rank matrix recovery, although γ_s and $\widehat{\gamma}_s$ may not be easy to compute. Development of efficiently computable bounds on these quantities is left to future work. Even though we develop and use techniques based on optimization, convex analysis, and geometry, we do not provide explicit analogues to the results of Donoho [35] where necessary and sufficient conditions for vector recovery special case were derived based on the geometric notions of face preservation and neighborliness. The corresponding generalization to low-rank recovery is not known, currently the closest one being [22]. Moreover, it is also important to consider the semidefinite relaxation (SDR) for the rank minimization with the positive semidefinite constraint since the SDR convexifies nonconvex or discrete optimization problems by removing the rank-one constraint. Another future research topic is to extend the main results and the techniques in this paper to the SDR.

Acknowledgments

The authors thank two anonymous referees for their very useful comments. The work was supported in part by the National Basic Research Program of China (2010CB732501), the National Natural Science Foundation of China (11171018), a Discovery Grant from NSERC, a research grant from the University of Waterloo, ONR research Grant N00014-12-10049, and the Fundamental Research Funds for the Central Universities (2011JBM306, 2013JBM095).

References

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

[2] 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.

[3] M. Fazel, H. Hindi, and S. P. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proceedings of American Control Conference*, pp. 4734–4739, June 2001.

[4] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[5] B. P. W. Ames and S. A. Vavasis, "Nuclear norm minimization for the planted clique and biclique problems," *Mathematical Programming*, vol. 129, no. 1, pp. 69–89, 2011.

[6] 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.

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

[8] C. Ding, D. Sun, and K. C. Toh, "An introduction to a class of matrix cone programming," Tech. Rep., 2010.

[9] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of a corrupted low-rank matrices," http://arxiv.org/abs/1009.5055.

[10] Y. Liu, D. Sun, and K. C. Toh, "An implementable proximal point algorithmic framework for nuclear norm minimization," *Mathematical Programming*, vol. 133, no. 1-2, pp. 399–436, 2012.

[11] Z. Liu and L. Vandenberghe, "Interior-point method for nuclear norm approximation with application to system identification," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 3, pp. 1235–1256, 2009.

[12] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and Bregman iterative methods for matrix rank minimization," *Mathematical Programming*, vol. 128, no. 1-2, pp. 321–353, 2011.

[13] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "RASL: robust alignment by sparse and low-rank decomposition for linearly correlated images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 2233–2246, 2010.

[14] B. Recht, W. Xu, and B. Hassibi, "Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization," in *Proceedings of the 47th IEEE Conference on Decision and Control (CDC '08)*, pp. 3065–3070, Cancun, Mexico, December 2008.

[15] M. Tao and X. Yuan, "Recovering low-rank and sparse components of matrices from incomplete and noisy observations," *SIAM Journal on Optimization*, vol. 21, no. 1, pp. 57–81, 2011.

[16] 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, 2006.

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

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

[19] 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.

[20] S. Oymak and B. Hassibi, "New null space results and recovery thresholds for matrix rank minimization," http://arxiv.org/abs/1011.6326.

[21] S. Oymak, K. Mohan, M. Fazel, and B. Hassibi, "A simplified approach to recovery conditions for low-rank matrices," in *Proceedings of International Symposium on Information Theory (ISIT '11)*, pp. 2318–2322, August 2011.

[22] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," in *Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computing*, 2011.

[23] A. d'Aspremont and L. El Ghaoui, "Testing the nullspace property using semidefinite programming," *Mathematical Programming*, vol. 127, no. 1, pp. 123–144, 2011.

[24] A. Juditsky and A. Nemirovski, "On verifiable sufficient conditions for sparse signal recovery via ℓ_1 minimization," *Mathematical Programming*, vol. 127, no. 1, pp. 57–88, 2011.

[25] A. Juditsky, F. Karzan, and A. Nemirovski, "Verifiable conditions of ℓ_1-recovery for sparse signals with sign restrictions," *Mathematical Programming*, vol. 127, no. 1, pp. 89–122, 2011.

[26] A. Juditsky, F. Karzan, and A. S. Nemirovski, "Accuracy guarantees for ℓ_1-recovery," *IEEE Transactions on Information Theory*, vol. 57, no. 12, pp. 7818–7839, 2011.

[27] G. A. Watson, "Characterization of the subdifferential of some matrix norms," *Linear Algebra and Its Applications*, vol. 170, pp. 33–45, 1992.

[28] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer, New York, NY, USA, 2nd edition, 2004.

[29] A. S. Lewis and H. S. Sendov, "Nonsmooth analysis of singular values. II. Applications," *Set-Valued Analysis*, vol. 13, no. 3, pp. 243–264, 2005.

[30] Y. Zhang, "Theory of Compressive Sensing via L_1-Minimization: A Non-RIP Analysis and Extensions," 2008.

[31] E. J. Candès and Y. Plan, "Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2342–2359, 2011.

[32] K. Lee and Y. Bresler, "Guaranteed minimum rank approximation from linear observations by nuclear norm minimization with an ellipsoidal constraint," http://arxiv.org/abs/0903.4742.

[33] R. Meka, P. Jain, and I. S. Dhillon, "Guaranteed rank minimization via singular value projection," http://arxiv.org/abs/0909.5457.

[34] K. Mohan and M. Fazel, "New restricted isometry results for noisy low-rank matrix recovery," in *Proceedings of IEEE International Symposium on Information Theory (ISIT '10)*, Austin, Tex, USA, June 2010.

[35] D. L. Donoho, "Neighborly polytopes and sparse solution of underdetermined linear equations," Tech. Rep., 2004.