



# Stability of low-rank matrix recovery and its connections to Banach space geometry



Javier Alejandro Chávez-Domínguez<sup>a,\*</sup>, Denka Kutzarova<sup>b,2</sup>

<sup>a</sup> Department of Mathematics, University of Texas at Austin, 2515 Speedway Stop C1200, Austin, TX 78712-1202, United States

<sup>b</sup> Institute of Mathematics, Bulgarian Academy of Sciences, Sofia, Bulgaria

## ARTICLE INFO

### Article history:

Received 26 June 2014

Available online 17 February 2015

Submitted by P. Nevai

### Keywords:

Compressed sensing

Low-rank matrix recovery

Schatten  $p$ -minimization

Gelfand widths

## ABSTRACT

There are well-known relationships between compressed sensing and the geometry of the finite-dimensional  $\ell_p$  spaces. A result of Kashin and Temlyakov [20] can be described as a characterization of the stability of the recovery of sparse vectors via  $\ell_1$ -minimization in terms of the Gelfand widths of certain identity mappings between finite-dimensional  $\ell_1$  and  $\ell_2$  spaces, whereas a more recent result of Foucart, Pajor, Rauhut and Ullrich [16] proves an analogous relationship even for  $\ell_p$  spaces with  $p < 1$ . In this paper we prove what we call matrix or noncommutative versions of these results: we characterize the stability of low-rank matrix recovery via Schatten  $p$ -(quasi-)norm minimization in terms of the Gelfand widths of certain identity mappings between finite-dimensional Schatten  $p$ -spaces.

© 2015 Elsevier Inc. All rights reserved.

## 1. Introduction

A mathematical problem that appears often in real-world situations is the following: we wish to recover a high-dimensional vector  $x \in \mathbb{R}^N$  from a measurement  $Ax$  where  $A : \mathbb{R}^N \rightarrow \mathbb{R}^m$  is a linear map and  $m$  is smaller than  $N$ . As stated the problem of course cannot be solved, but that changes if we have the additional condition that the unknown vector  $x$  is *sparse*, i.e. it has a small number of non-zero coordinates. This is the subject matter of compressed sensing, a very active area of research with numerous applications; the book [17] is a recent comprehensive reference. Formally, this *sparse recovery problem* can be stated as

$$\min \|x\|_0 \quad \text{subject to } Ax = y, \tag{1.1}$$

\* Corresponding author.

E-mail addresses: jachavezd@math.utexas.edu (J.A. Chávez-Domínguez), denka@math.uiuc.edu (D. Kutzarova).

<sup>1</sup> Current address: Instituto de Ciencias Matemáticas, CSIC-UAM-UC3M-UCM, C/Nicolás Cabrera, n° 13-15, Campus de Cantoblanco, UAM, 28049, Madrid, Spain.

<sup>2</sup> Current address: Department of Mathematics, University of Illinois at Urbana-Champaign, 1409 W. Green Street Urbana, IL 61801, United States.

where  $\|\cdot\|_0$  represents the number of nonzero coordinates of a vector. This is an NP-hard [28] and non-convex problem, so we are interested in conditions (especially on the map  $A$ ) that would allow us to solve an easier problem and still arrive at the right solution. In that spirit, a basic technique in compressed sensing is that of  $\ell_1$ -minimization: if the vector  $x$  is sparse enough, then minimizing  $\|x'\|_{\ell_1}$  over all vectors  $x'$  so that  $Ax' = Ax$  actually allows us to recover  $x$ . Formally, instead of problem (1.1) we are considering its *convex relaxation*

$$\min \|x\|_{\ell_1} \quad \text{subject to } Ax = y. \tag{1.2}$$

Additionally, one can consider the analogous problem of  $\ell_p$ -minimization.

In practice the unknown vectors are not necessarily sparse, but are close to sparse ones. Thus for any method of recovery it is of utmost importance to investigate its stability, that is, having a control on the distance between the original vector and its reconstruction in terms of the distance from the original vector to the sparse vectors. It turns out that the stability of sparse vector recovery through  $\ell_p$ -minimization has connections to the Banach-space geometry of finite-dimensional  $\ell_p$ -spaces. More generally, it is known that there are connections between recovery – in particular the compressed sensing model – and geometric quantities called *Gelfand widths*, see e.g. [29,10,9,20].

In many practical situations, there is extra structure in the space of unknown vectors. A good example is the famous *matrix completion problem* (also known as the *Netflix problem*), where the unknown is a matrix and the measurement map gives us a subset of its entries. In this case sparsity gets replaced by the more natural condition of having low rank, and the last few years have witnessed an explosion of work in this area. In what follows,  $M_N$  will denote the space of  $N \times N$  real-valued matrices. We now consider a linear operator  $\mathcal{A} : M_N \rightarrow \mathbb{R}^m$ , and a fixed vector  $y \in \mathbb{R}^m$ . The low-rank recovery problem can thus be stated as the problem of finding the solution to

$$\min \text{rank}(X) \quad \text{subject to } \mathcal{A}X = y. \tag{1.3}$$

This is again an NP-hard problem, so once again we would like to replace it by another one which is simpler to solve but has the same solution.

In noncommutative functional analysis the Schatten  $p$ -spaces are usually considered to be the counterparts of the classical  $\ell_p$  spaces (recall that the Schatten  $p$ -norm of a matrix  $X$  is the  $\ell_p$ -norm of its vector of singular values), so from that point of view it is natural to wonder whether the Schatten  $p$ -norm minimization approach can work in the matrix context. We would like to consider operators  $\mathcal{A}$  for which the previous problem is equivalent to

$$\min \|X\|_{S_1} \quad \text{subject to } \mathcal{A}X = y, \tag{1.4}$$

where  $\|X\|_{S_p}$  denotes the Schatten  $p$ -norm of the matrix  $X \in M_N$ . This has already been studied in several situations of interest, with the idea going back to the Ph.D. thesis of M. Fazel [13]. Schatten 1-norm (also known as nuclear norm) minimization in the particular case of the matrix completion problem was studied by Candès and Recht [4] (and later on Candès and Tao [7] gave optimality results quantifying the minimum number of entries needed to recover a matrix of low rank exactly by any method whatsoever, and showed that nuclear norm minimization is nearly optimal). Plenty of concepts from the classical theory of compressed sensing have found matrix counterparts: Candès and Recht [4] use the idea of *coherence*; Recht, Fazel and Parrilo [33] used the matrix version of the *restricted isometry property*; whereas both Recht, Xu and Hassibi [34] and Fornasier, Rauhut and Ward [14] considered *null-space conditions*; the *spherical section property* was used by Dvijoatham and Fazel [11] and Oymak, Mohan, Fazel and Hassibi [30].

One thing that does not appear to have been explicitly studied in the matrix context is the aforementioned relationship to Gelfand widths. Recall that the Gelfand  $k$ -width of a subset  $K$  of a normed space  $E$  is defined as

$$d^k(K, E) := \inf \left\{ \sup_{x \in K \cap L} \|x\|_E : L \text{ subspace of } E \text{ with } \text{codim}(L) \leq k \right\}.$$

A closely related concept that is more commonly used in Banach space theory is that of a *Gelfand number*: if  $T : X \rightarrow Y$  is a linear operator between normed spaces, its  $k$ -Gelfand number is defined by

$$c_k(T) := \inf \left\{ \sup_{x \in L, \|x\| \leq 1} \|Tx\| : L \text{ subspace of } X \text{ with } \text{codim}(L) < k \right\}.$$

The speed of convergence to zero of the sequence of Gelfand numbers  $(c_k(T))_{k=1}^\infty$  is a measure of the compactness of the operator  $T$ , and is an example of a sequence of  $s$ -numbers; see [31,22] for more details. In the cases under consideration in this paper the concepts of Gelfand numbers and Gelfand widths actually coincide (up to a small shift in the index), so we will freely use them both depending on the particular context. It should be mentioned that there is a general concept of Gelfand width for a linear map that is not always the same as the corresponding Gelfand number (see [32, Sec. 6.2.6] for the details), but both concepts do coincide in nice situations (see [12]).

The work of Kashin and Temlyakov [20] made more precise the already-known connection between compressed sensing and the Kashin–Garnaev–Gluskin [19,18] result that calculates the  $m$ -Gelfand numbers of the identity map from  $\ell_1^N$  to  $\ell_2^N$ , namely

$$c_{m+1}(id : \ell_1^N \rightarrow \ell_2^N) \leq C \sqrt{\frac{1 + \log(N/m)}{m}}.$$

In a nutshell, the main result of Kashin and Temlyakov shows that the stability of sparse recovery via  $\ell_1$ -minimization is equivalent to the kernel of the measurement map being a “good” subspace where the Gelfand number of a certain order is achieved. This idea was taken further by Foucart, Pajor, Rauhut and Ullrich [16], who used compressed sensing ideas to calculate the Gelfand numbers of identity maps from  $\ell_p^N$  to  $\ell_q^N$  for  $0 < p \leq 1$ ,  $p < q \leq 2$ .

In this paper, we prove matrix versions of the aforementioned theorems, relating the stability of low-rank matrix recovery to the Gelfand numbers of identity maps between finite-dimensional Schatten  $p$ -spaces. As far as we know the only part of our results that is already written down in the literature is the following analogue of the Kashin–Garnaev–Gluskin result due to Carl and Defant [8, p. 252], namely the calculation of the  $m$ -Gelfand numbers of the identity map from  $S_1^N$  to  $S_2^N$ : for  $1 \leq m \leq N^2$ ,

$$c_m(id : S_1^N \rightarrow S_2^N) \asymp \min \left\{ 1, \frac{N}{m} \right\}^{1/2}.$$

Here and in the rest of the paper, the symbol  $\asymp$  means that the quantities on the left and the right are equivalent up to universal constants. If we want to emphasize the dependence of the constants on some parameters, those will appear as subindices of the equivalence symbol ( $\asymp_{p,q}$ , for example).

The rest of this paper is organized as follows. In Section 2 we introduce our notation and state several known results that will be needed in the sequel. In Section 3 we show the first relationships between the stability of low-rank matrix recovery and the geometry of Banach spaces, by proving a matrix version of the Kashin–Temlyakov theorem. Section 4 contains a technical result, a matrix version of the main theorem from [15] that gives conditions on the measurement map  $\mathcal{A}$  that guarantee the stability of the Schatten  $p$ -minimization scheme. A very similar theorem was recently obtained independently by Liu, Huang and

Chen [25], though our proof is different and we require a weaker hypothesis. In the final section, the technical result from Section 4 is used to calculate the Gelfand numbers of the identity maps from  $S_p^N$  to  $S_q^N$  for  $0 < p \leq 1, p < q \leq 2$  in the spirit of the work of Foucart, Pajor, Rauhut and Ullrich.

**2. Notation and preliminaries**

In this paper we will only consider square matrices, but all the results can be adapted to rectangular ones. For  $p > 0$  we will denote by  $S_p^N$  the space of  $N \times N$  matrices with the Schatten  $p$ -quasi-norm, given by

$$\|X\|_{S_p} = \left( \sum_{i=1}^N |\sigma_i|^p \right)^{1/p},$$

where  $(\sigma_i)_{i=1}^N$  is the vector of singular values of the matrix  $X$ . Similarly,  $S_{p,\infty}^N$  will denote the space of  $N \times N$  matrices with the weak-Schatten- $p$ -quasi-norm given by

$$\|X\|_{S_{p,\infty}} = \max_{1 \leq k \leq N} k^{1/p} |\sigma_k^*|$$

where  $(\sigma_i^*)_{i=1}^N$  is the non-increasing rearrangement of  $(\sigma_i)_{i=1}^N$ . For any quasi-normed space  $X$ ,  $B_X$  will denote its unit ball.

We will need to consider the best  $s$ -rank approximation error in the Schatten  $p$ -quasi-norm,

$$\rho_s(X)_{S_p} := \inf \{ \|X - Y\|_{S_p} : \text{rank}(Y) \leq s \}.$$

It is well known that the infimum is actually attained at the  $s$ -spectral truncation  $Y = X_{[s]}$  (that is, keeping only the  $s$  largest singular values in the singular value decomposition).

Given a linear map  $\mathcal{A} : M_N \rightarrow \mathbb{R}^m$  and a vector  $y \in \mathbb{R}^m$ , for  $0 < p \leq 1$  we will denote by  $\Delta_p(y)$  a solution to

$$\text{minimize } \|Z\|_{S_p} \quad \text{subject to } \mathcal{A}Z = y.$$

That is,  $\Delta_p$  is the Schatten  $p$ -quasi-norm minimization reconstruction map. The map  $\Delta_p$  of course depends on the measuring map  $\mathcal{A}$ , but for simplicity we do not make this dependence explicit in the notation.

*2.1. The Restricted Isometry Property*

The Restricted Isometry Property (RIP) for a linear map  $A : \mathbb{R}^N \rightarrow \mathbb{R}^m$  was introduced by Candès and Tao [5], and quickly became a key concept in the analysis of sparse recovery via  $\ell_p$ -norm minimization. The  $s$ -order restricted isometry constant of such a map is the smallest  $\delta > 0$  such that for every vector  $x \in \mathbb{R}^N$  of sparsity at most  $s$ ,

$$(1 - \delta) \|x\|_{\ell_2}^2 \leq \|Ax\|_{\ell_2}^2 \leq (1 + \delta) \|x\|_{\ell_2}^2.$$

The importance of the RIP stems from the fact that small restricted isometry constants imply exact recovery via  $\ell_p$ -quasi-norm minimization for  $0 < p \leq 1$ , and it should be noted that it is well known that random choices of the matrix  $A$  give small RIP constants of order  $s$ , as long as  $m$  is at least of the order of  $s \ln(eN/s)$  [6,1,27].

The version of the RIP for matrix recovery was introduced by Recht, Fazel and Parrilo [33], and is as follows: a linear map  $\mathcal{A} : M_N \rightarrow \mathbb{R}^m$  is said to have the Restricted Isometry Property of rank  $s$  with constant  $\delta > 0$  if for every matrix  $Z \in M_N$  of rank at most  $s$ ,

$$(1 - \delta) \|Z\|_{S_2}^2 \leq \|\mathcal{A}Z\|_{\ell_2}^2 \leq (1 + \delta) \|Z\|_{S_2}^2.$$

The best such constant is denoted by  $\delta_s(\mathcal{A})$ .

Just as in the vector case, random constructions give small RIP constants. The next result follows from [3, Thm. 2.3], and will be very important for us in the sequel.

**Theorem 2.1.** *Given a prescribed  $\delta \in (0, 1)$ , there is a constant  $C_\delta$  such that if the entries of the map  $\mathcal{A}$  (seen a matrix with respect to the canonical bases in  $M_N$  and  $\mathbb{R}^m$ ) are independent gaussians with mean zero and variance  $1/m$ , then with positive (even overwhelming) probability  $\delta_s(\mathcal{A}) \leq \delta$  holds provided that*

$$m \geq C_\delta s N. \quad (2.1)$$

### 3. A noncommutative Kashin–Temlyakov theorem

We will prove a matrix version of the Kashin–Temlyakov characterization of the stability of sparse recovery via  $\ell_1$ -norm minimization in terms of widths. To this end, we define three properties modeled after the ones studied in [20].

**Definition 3.1.** Let  $N^2 > m$  and let  $\mathcal{A} : M_N \rightarrow \mathbb{R}^m$  be a linear operator. We say that  $\mathcal{A}$  has a:

(a) *Matrix Strong Compressed Sensing Property (MSCSP)* if for any  $X \in M_N$  we have

$$\|X - \Delta_1(\mathcal{A}X)\|_{S_2} \leq C s^{-1/2} \rho_s(X)_{S_1}$$

for  $s \asymp m/N$ .

(b) *Matrix Weak Compressed Sensing Property (MWCSPP)* if for any  $X \in M_N$  we have

$$\|X - \Delta_1(\mathcal{A}X)\|_{S_2} \leq C s^{-1/2} \|X\|_{S_1}$$

for  $s \asymp m/N$ .

(c) *Matrix Width Property (MWP)* if for any  $X \in \ker(\mathcal{A})$ ,

$$\|X\|_{S_2} \leq C(N/m)^{-1/2} \|X\|_{S_1}. \quad (3.1)$$

Notice that the MSCSP is a weakening of condition (i) in [30, Lemma 8], since we are only considering  $X' = \Delta_1(\mathcal{A}X)$ . Also, the name of the MWP comes from its clear relationship to the definition of the Gelfand numbers/widths. The following theorem is a matrix version of the Kashin–Temlyakov theorem [20, Thm. 2.2]:

**Theorem 3.2.** *For a linear operator  $\mathcal{A} : M_N \rightarrow \mathbb{R}^m$ , the MSCSP, MWCSPP and MWP are equivalent (up to a change in the constants).*

**Proof.** The MSCSP trivially implies the MWCSPP, since  $\rho_s(X)_{S_1} \leq \|X - 0\|_{S_1} = \|X\|$ . Assume that  $\mathcal{A}$  has the MWCSPP. Given  $X \in \ker(\mathcal{A})$ , note that  $\mathcal{A}X = 0 = \mathcal{A}0$ , so clearly  $0 = \Delta_1(0) = \Delta_1(\mathcal{A}X)$  and thus from the MWCSPP we have  $\|X\|_{S_2} \leq C s^{-1/2} \|X\|_{S_1}$ , giving the MWP. Assume now that we have the MWP, that

is, that Eq. (3.1) holds. If  $s < \frac{1}{4}C^{-2}N/m$ , from [30, Thm. 2] (which is a matrix version of [20, Thm. 2.1]) we obtain

$$\|X - \Delta_1(\mathcal{A}X)\|_{S_1} \leq C' \rho_s(X)_{S_1}$$

for  $C' = 2(1 - 2\sqrt{C^2 sm/N})^{-1}$ . Since  $X - \Delta_1(\mathcal{A}X) \in \ker \mathcal{A}$ , the previous equation together with (3.1) imply the MSCSP.  $\square$

The aforementioned Kashin–Temlyakov theorem says, in a nutshell, that the stability of sparse-vector recovery via  $\ell_1$ -minimization has limits imposed by the geometry of Banach spaces encoded in the appropriate Gelfand widths. In the previous proposition, we showed a similar relationship relating the stability of low-rank recovery via nuclear norm minimization with some other Gelfand widths. As in the vector case, following [17, Cor. 10.6], there is a relationship between the geometry of  $S_1^N$  and the stability of compressed sensing by any method. See Theorem 5.5 below for the precise statement.

#### 4. Stability of low-rank matrix recovery through Schatten $p$ quasi-norm minimization

In this technical section we prove a general result (a matrix version of the main theorem in [15]) that gives RIP-style conditions on the measuring map  $\mathcal{A}$  that guarantee the stability of the Schatten  $p$ -norm minimization scheme. For that we will need some notation: Let  $\alpha_s, \beta_s \geq 0$  be the best constants in the inequalities

$$\alpha_s \|Z\|_{S_2} \leq \|\mathcal{A}Z\|_{\ell_2} \leq \beta_s \|Z\|_{S_2}, \quad \text{rank}(Z) \leq s.$$

The results will be stated in terms of a quantity invariant under the change  $\mathcal{A} \leftarrow c\mathcal{A}$ , namely

$$\gamma_{2s} := \frac{\beta_{2s}^2}{\alpha_{2s}^2} \geq 1.$$

Note that this constant is related to the RIP constant, in fact

$$\gamma_{2s} = \frac{1 + \delta_{2s}}{1 - \delta_{2s}}.$$

Unlike in the rest of the paper, we will consider the more general situation of approximate recovery when measurements are moderately flawed, namely the problem

$$\text{minimize } \|Z\|_{S_p} \quad \text{subject to} \quad \|\mathcal{A}Z - y\|_{\ell_2} \leq \beta_{2s} \cdot \theta. \tag{P_{p,\theta}}$$

For simplicity, we will write  $(P_p)$  instead of  $(P_{p,0})$ . Note that by a compactness argument, a solution of  $(P_{p,\theta})$  exists for any  $0 < p \leq 1$  and any  $\theta \geq 0$ . The following theorem is a matrix version of [15, Thm. 3.1]. It gives conditions (in the spirit of the RIP) that guarantee not only the stability but also the robustness (that is, resistance to errors in the measurements) of the Schatten  $p$ -quasi-norm-minimization for low-rank matrix recovery.

**Theorem 4.1.** *Given  $0 < p \leq 1$ , if for some integer  $t \geq s$*

$$\gamma_{2t} - 1 < 4(\sqrt{2} - 1) \left(\frac{t}{s}\right)^{1/p-1/2} \tag{4.1}$$

*then a solution  $X^*$  of  $(P_{p,\theta})$  approximates the original matrix  $X$  with errors*

$$\|X - X^*\|_{S_p} \leq C_1 \rho_s(X)_{S_p} + D_1 \cdot s^{1/p-1/2} \cdot \theta, \tag{4.2}$$

$$\|X - X^*\|_{S_2} \leq C_2 \frac{\rho_s(X)_{S_p}}{t^{1/p-1/2}} + D_2 \cdot \theta, \tag{4.3}$$

where the constants  $C_1, C_2, D_1$  and  $D_2$  depend only on  $p, \gamma_{2t}$  and the ratio  $s/t$ .

**Proof.** We will need to recall some properties of the  $S_p$ -quasi-norm. Namely, for any matrices  $U$  and  $V$ ,

$$\|U\|_{S_1} \leq \|U\|_{S_p}, \quad \|U\|_{S_p} \leq N^{1/p-1/2} \|U\|_{S_2}, \quad \|U + V\|_{S_p}^p \leq \|U\|_{S_p}^p + \|V\|_{S_p}^p. \tag{4.4}$$

**STEP 1:** Consequence of the assumption on  $\gamma_{2t}$ .

We will consider certain matrix decompositions similar to the ones in [21]. Consider the singular value decomposition of  $X$ , given by

$$X = U \operatorname{diag}(\lambda_i(X)) V^T$$

where  $U, V$  are unitary matrices and  $\lambda(X) = (\lambda_1(X), \dots, \lambda_N(X))$  are the singular values of  $X$  arranged in decreasing order. For any matrix  $Z \in M_N$ , we will consider a block decomposition of  $Z$  with respect to  $X$  as follows: let  $U^T Z V$  have the block form

$$U^T Z V = \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix}$$

where  $Z_{11}, Z_{12}, Z_{21}, Z_{22}$  are of sizes  $s \times s, s \times (N - s), (N - s) \times s, (N - s) \times (N - s)$ , respectively. We then decompose  $Z$  as  $Z = Z_{(s)} + Z_{(s)}^c$  where

$$Z_{(s)} = U \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & 0 \end{pmatrix} V^T$$

$$Z_{(s)}^c = U \begin{pmatrix} 0 & 0 \\ 0 & Z_{22} \end{pmatrix} V^T$$

Furthermore, we now consider the singular value decomposition of  $Z_{22}$  given by

$$Z_{22} = P \operatorname{diag}(\lambda(Z_{22})) Q^T$$

with  $P$  and  $Q$  being  $(N - s) \times (N - s)$  unitary matrices, and  $\lambda(Z_{22})$  being the vector of the  $N - s$  singular values of  $Z_{22}$  arranged in decreasing order. We decompose  $\lambda(Z_{22})$  as a sum of vectors  $Z_{T_i}$ , each of sparsity at most  $t$ , where  $T_1$  corresponds to the locations of the  $t$  largest entries of  $\lambda(Z_{22})$ ,  $T_2$  to the locations of the next  $t$  largest entries, and so on. For  $i \geq 1$  we now define

$$Z_{T_i} = U \begin{pmatrix} 0 & 0 \\ 0 & P \operatorname{diag}(\lambda_{T_i}(Z_{22})) Q^T \end{pmatrix} V^T,$$

and denote  $Z_{T_0} := Z_{(s)}$ .

We first observe that

$$\begin{aligned} \|Z_{T_0}\|_{S_2}^2 + \|Z_{T_1}\|_{S_2}^2 &= \|Z_{T_0} + Z_{T_1}\|_{S_2}^2 \leq \frac{1}{\alpha_{2t}^2} \|\mathcal{A}(Z_{T_0} + Z_{T_1})\|_{\ell_2}^2 \\ &= \frac{1}{\alpha_{2t}^2} \langle \mathcal{A}(Z - Z_{T_2} - Z_{T_3} - \dots), \mathcal{A}(Z_{T_0} + Z_{T_1}) \rangle \\ &= \frac{1}{\alpha_{2t}^2} \langle \mathcal{A}Z, \mathcal{A}(Z_{T_0} + Z_{T_1}) \rangle + \frac{1}{\alpha_{2t}^2} \sum_{k \geq 2} [\langle \mathcal{A}(-Z_{T_k}), \mathcal{A}Z_{T_0} \rangle + \langle \mathcal{A}(-Z_{T_k}), \mathcal{A}Z_{T_1} \rangle] \end{aligned} \tag{4.5}$$

Let us renormalize the vectors  $-Z_{T_k}$  and  $Z_{T_0}$  so that their  $S_2$ -norms equal one by setting  $Y_k := -Z_{T_k} / \|Z_{T_k}\|_{S_2}$  and  $Y_0 := Z_{T_0} / \|Z_{T_0}\|_{S_2}$ . We then obtain, using the polarization identity

$$\begin{aligned} \frac{\langle \mathcal{A}(-Z_{T_k}), \mathcal{A}Z_{T_0} \rangle}{\|Z_{T_k}\|_{S_2} \|Z_{T_0}\|_{S_2}} &= \langle \mathcal{A}Y_k, \mathcal{A}Y_0 \rangle = \frac{1}{4} [ \|\mathcal{A}(Y_k + Y_0)\|_{\ell_2}^2 - \|\mathcal{A}(Y_k - Y_0)\|_{\ell_2}^2 ] \\ &\leq \frac{1}{4} [ \beta_{2t}^2 \|Y_k + Y_0\|_{S_2}^2 - \alpha_{2t}^2 \|Y_k - Y_0\|_{S_2}^2 ] = \frac{1}{2} [ \beta_{2t}^2 - \alpha_{2t}^2 ]. \end{aligned}$$

An analogous argument with  $T_1$  in place of  $T_0$  allows us to conclude

$$\langle \mathcal{A}(-Z_{T_k}), \mathcal{A}Z_{T_0} \rangle + \langle \mathcal{A}(-Z_{T_k}), \mathcal{A}Z_{T_1} \rangle \leq \frac{\beta_{2t}^2 - \alpha_{2t}^2}{2} \|Z_{T_k}\|_{S_2} [ \|Z_{T_0}\|_{S_2} + \|Z_{T_1}\|_{S_2} ]. \tag{4.6}$$

On the other hand, we have

$$\langle \mathcal{A}Z, \mathcal{A}(Z_{T_0} + Z_{T_1}) \rangle \leq \|\mathcal{A}Z\|_{\ell_2} \cdot \|\mathcal{A}(Z_{T_0} + Z_{T_1})\|_{\ell_2} \leq \|\mathcal{A}Z\|_{\ell_2} \cdot \beta_{2t} [ \|Z_{T_0}\|_{S_2} + \|Z_{T_1}\|_{S_2} ]. \tag{4.7}$$

Substituting the inequalities (4.6) and (4.7) into (4.5) we have

$$\|Z_{T_0}\|_{S_2}^2 + \|Z_{T_1}\|_{S_2}^2 \leq \left( \frac{\gamma_{2t}}{\beta_{2t}} \|\mathcal{A}Z\|_{\ell_2} + \frac{\gamma_{2t} - 1}{2} \sum_{k \geq 2} \|Z_{T_k}\|_{S_2} \right) [ \|Z_{T_0}\|_{S_2} + \|Z_{T_1}\|_{S_2} ].$$

If we set  $c := \|\mathcal{A}Z\|_{\ell_2} \cdot \gamma_{2t} / \beta_{2t}$ ,  $d := (\gamma_{2t} - 1) / 2$  and  $\Sigma = \sum_{k \geq 2} \|Z_{T_k}\|_{S_2}$ , the previous inequality is

$$\|Z_{T_0}\|_{S_2}^2 - (c + d\Sigma) \|Z_{T_0}\|_{S_2} + \|Z_{T_1}\|_{S_2}^2 - (c + d\Sigma) \|Z_{T_1}\|_{S_2} \leq 0,$$

or equivalently,

$$\left[ \|Z_{T_0}\|_{S_2} - \frac{c + d\Sigma}{2} \right]^2 + \left[ \|Z_{T_1}\|_{S_2} - \frac{c + d\Sigma}{2} \right]^2 \leq \frac{(c + d\Sigma)^2}{2}.$$

By getting rid of the second squared term, this easily implies

$$\|Z_{T_0}\|_{S_2} \leq \frac{c + d\Sigma}{2} + \frac{c + d\Sigma}{\sqrt{2}} = \frac{1 + \sqrt{2}}{2} (c + d\Sigma). \tag{4.8}$$

By Hölder’s inequality (see (4.4)) we get

$$\|Z_{T_0}\|_{S_p} \leq s^{1/p-1/2} \|Z_{T_0}\|_{S_2} \leq s^{1/p-1/2} \frac{1 + \sqrt{2}}{2} (c + d\Sigma). \tag{4.9}$$

We now proceed to bound  $\Sigma$ . For  $k \geq 2$ , let  $\eta, \eta'$  be singular values of  $Z_{T_k}, Z_{T_{k-1}}$ , respectively. By definition, we must have  $\eta \leq \eta'$ . Raising to the  $p$ -th power and averaging over all singular values of  $Z_{T_{k-1}}$ ,  $\eta^p \leq t^{-1} \|Z_{T_{k-1}}\|_{S_p}^p$ , and hence  $\eta^2 \leq t^{-2/p} \|Z_{T_{k-1}}\|_{S_p}^2$ . Adding over all singular values of  $Z_{T_k}$  and taking the square root, this yields  $\|Z_{T_k}\|_{S_2} \leq t^{1/2-1/p} \|Z_{T_{k-1}}\|_{S_p}$ . Therefore,

$$\Sigma = \sum_{k \geq 2} \|Z_{T_k}\|_{S_2} \leq t^{1/2-1/p} \sum_{k \geq 1} \|Z_{T_k}\|_{S_p} \leq t^{1/2-1/p} \left[ \sum_{k \geq 1} \|Z_{T_k}\|_{S_p}^p \right]^{1/p} = t^{1/2-1/p} \|Z_{(s)}^c\|_{S_p}.$$

Combining the above inequality with (4.9), we obtain

$$\|Z(s)\|_{S_p} \leq \frac{\lambda}{2\beta_{2t}} \cdot \|AZ\|_{\ell_2} \cdot s^{1/p-1/2} + \mu \cdot \|Z(s)\|_{S_p}^c, \tag{4.10}$$

where the constants  $\lambda$  and  $\mu$  are given by

$$\lambda := (1 + \sqrt{2})\gamma_{2t} \text{ and } \mu := \frac{1}{4}(1 + \sqrt{2})(\gamma_{2t} - 1) \left(\frac{s}{t}\right)^{1/p-1/2}.$$

Note that the assumption on  $\gamma_{2t}$  translates into the inequality  $\mu < 1$ .

**STEP 2:** From now on let  $Z := X - X^*$ .

Because  $X^*$  is a minimizer of  $(P_{p,\theta})$ , we have

$$\|X^*\|_{S_p}^p \leq \|X\|_{S_p}^p. \tag{4.11}$$

From [21, Lemma 2.2], whenever  $B, C \in M_N$  satisfy  $B^T C = 0$  and  $BC^T = 0$  one has  $\|B + C\|_{S_p}^p = \|B\|_{S_p}^p + \|C\|_{S_p}^p$ . In particular, note that

$$\|X\|_{S_p}^p = \|X(s)\|_{S_p}^p + \|X(s)^c\|_{S_p}^p \text{ and } \|X(s) - Z(s)\|_{S_p}^p = \|X(s)\|_{S_p}^p + \|Z(s)\|_{S_p}^p. \tag{4.12}$$

From the  $p$ -triangle inequality (see (4.4)), since

$$X(s) - Z(s) = X - Z - X(s)^c + Z(s) = X^* - X(s)^c + Z(s),$$

we get

$$\|X(s) - Z(s)\|_{S_p}^p \leq \|X^*\|_{S_p}^p + \|X(s)^c\|_{S_p}^p + \|Z(s)\|_{S_p}^p.$$

Together with (4.11) and both equalities in (4.12), this yields

$$\|X(s)\|_{S_p}^p + \|Z(s)\|_{S_p}^p \leq \|X(s)\|_{S_p}^p + \|X(s)^c\|_{S_p}^p \|X(s)^c\|_{S_p}^p + \|Z(s)\|_{S_p}^p.$$

After a cancellation and noticing that  $\|X(s)^c\|_{S_p}^p = \rho_s(X)_{S_p}^p$ , we obtain

$$\|Z(s)\|_{S_p}^p \leq 2\rho_s(X)_{S_p}^p + \|Z(s)\|_{S_p}^p. \tag{4.13}$$

**STEP 3:** Error estimates.

We first note the bound

$$\|AZ\|_{\ell_2} = \|AX - AX^*\|_{\ell_2} \leq \|AX - y\|_{\ell_2} + \|y - AX^*\|_{\ell_2} \leq 2\beta_{2s} \cdot \theta.$$

For the  $S_p$ -error, we combine the estimates in (4.10) and (4.13) to obtain

$$\|Z(s)\|_{S_p}^p \leq 2\rho_s(X)_{S_p}^p + \lambda^p \cdot s^{1-p/2} \cdot \theta^p + \mu^p \cdot \|Z(s)\|_{S_p}^p.$$

As a consequence of  $\mu < 1$ , we have

$$\|Z(s)\|_{S_p}^p \leq \frac{2}{1 - \mu^p} \rho_s(X)_{S_p}^p + \frac{\lambda^p}{1 - \mu^p} \cdot s^{1-p/2} \cdot \theta^p.$$

Using the estimate (4.10) once again, we can derive that

$$\begin{aligned} \|Z\|_{S_p}^p &\leq \|Z_{(s)}\|_{S_p}^p + \|Z_{(s)}^c\|_{S_p}^p \leq (1 + \mu^q) \cdot \|Z_{(s)}^c\|_{S_p}^p + \lambda^p \cdot s^{1-p/2} \cdot \theta^p \\ &\leq \frac{2}{1 - \mu^p} (1 + \mu^p) \rho_s(X)_{S_p}^p + \frac{2\lambda^p}{1 - \mu^p} \cdot s^{1-p/2} \cdot \theta^p \\ &\leq 2^{1-1/p} \left[ \frac{2^{1/p}}{(1 - \mu^p)^{1/p}} (1 + \mu^p)^{1/p} \rho_s(X)_{S_p} + \frac{2^{1/p}\lambda}{(1 - \mu^p)^{1/p}} \cdot s^{1/p-1/2} \cdot \theta \right]^p \end{aligned}$$

where we have used the inequality  $(a^p + b^p)^{1/p} \leq 2^{1/p-1}(a + b)$  for  $a, b \geq 0$ . The desired estimate (4.2) follows with

$$C_1 := \frac{2^{2/p-1}(1 + \mu^p)^{1/p}}{(1 - \mu^p)^{1/p}}, \quad D_1 := \frac{2^{2/p-1}\lambda}{(1 - \mu^p)^{1/p}}.$$

For the  $S_2$ -error, let us observe that the bound in (4.8) also holds if we replace  $\|Z_{T_0}\|_{S_2}$  by  $\|Z_{T_1}\|_{S_2}$ , and hence

$$\|Z\|_{S_2} = \left[ \sum_{k \geq 0} \|Z_{T_k}\|_{S_2}^2 \right]^{1/2} \leq \sum_{k \geq 0} \|Z_{T_k}\|_{S_2} \leq (1 + \sqrt{2}) \cdot (c + d\Sigma) + \Sigma \leq \nu \cdot \Sigma + 2\lambda \cdot \theta,$$

where  $\nu = (\lambda + 1 - \sqrt{2})/2$ . We also have that

$$\begin{aligned} \Sigma \leq t^{1/2-1/p} \|Z_{(s)}^c\|_{S_p} &\leq t^{1/2-1/p} \left[ \frac{2}{1 - \mu^p} \rho_s(X)_{S_p}^p + \frac{\lambda^p}{1 - \mu^p} \cdot s^{1-p/2} \cdot \theta^p \right]^{1/p} \\ &\leq t^{1/2-1/p} 2^{1/p-1} \left[ \frac{2^{1/p}}{(1 - \mu^p)^{1/p}} \rho_s(X)_{S_p} + \frac{\lambda}{(1 - \mu^p)^{1/p}} \cdot s^{1/p-1/2} \cdot \theta \right], \end{aligned}$$

and hence we conclude that

$$\|Z\|_{S_2} \leq \nu t^{1/2-1/p} 2^{1/p-1} \left[ \frac{2^{1/p}}{(1 - \mu^p)^{1/p}} \rho_s(X)_{S_p} + \frac{\lambda}{(1 - \mu^p)^{1/p}} \cdot s^{1/p-1/2} \cdot \theta \right] + 2\lambda\theta.$$

This gives the estimate (4.3) with

$$C_2 = \frac{2^{2/p-2}(\lambda + 1 - \sqrt{2})}{(1 - \mu^p)^{1/p}}, \quad D_2 = \frac{2^{1/p-2}\lambda(\lambda + 1 - \sqrt{2})}{(1 - \mu^p)^{1/p}} + 2\lambda. \quad \square$$

As consequences of Theorem 4.1, we obtain two corollaries that are matrix versions of the ones in [15]. The first one corresponds to the case of exact recovery.

**Corollary 4.2.** *Given  $0 < p \leq 1$ , if*

$$\gamma_{2t} - 1 < 4(\sqrt{2} - 1) \left( \frac{t}{s} \right)^{1/p-1/2} \quad \text{for some integer } t \geq s,$$

*then every rank  $s$  matrix is exactly and stably recovered by solving  $(P_p)$ .*

The second one deals with the special case of nuclear norm minimization.

**Corollary 4.3.** *Under the assumption that  $\gamma_{2s} < 4\sqrt{2} - 3 \approx 2.6569$ , every rank  $s$  matrix is exactly and stably recovered by solving  $(P_1)$ .*

This last corollary is clearly related to existing results on the RIP, it corresponds to the condition  $\delta_{2s} < 2(3 - \sqrt{2})/7 \approx 0.4531$ . Note that for the specific case of  $p = 1$  this condition is not the best possible: a very recent result of Cai and Zhang [2] shows that the optimal condition to have exact recovery of rank  $s$  matrices via nuclear norm minimization is in fact  $\delta_{2s} < 1/\sqrt{2} \approx 0.7071$ . Another recent result similar to our Theorem 4.1 is [25, Thm. 6] (which in turn generalizes results of Lee and Bresler [24,23]), where they get a conclusion of the same form as (4.3) but requiring a stronger hypothesis. Finally, note also that Theorem 2.1 guarantees the existence of maps  $\mathcal{A}$  satisfying the hypothesis of Theorem 4.1.

**5. The Gelfand widths of  $S_p$ -balls for  $0 < p \leq 1$**

In this section we calculate the Gelfand numbers

$$c_m(id : S_p^N \rightarrow S_q^N)$$

for  $0 < p \leq 1, p < q \leq 2$ . This can be considered as a noncommutative version of the results from [16], where they use compressed sensing ideas to calculate the corresponding Gelfand numbers

$$c_m(id : \ell_p^N \rightarrow \ell_q^N).$$

Inspired by their approach, our proof is based on low-rank matrix recovery ideas.

Our main result is the following (compare to [16, Thm. 1.1]).

**Theorem 5.1.** *For  $0 < p \leq 1$  and  $p < q \leq 2$ , if  $1 \leq m < N^2$ , then*

$$d^m(B_{S_p^N}, S_q^N) \asymp_{p,q} \min \left\{ 1, \frac{N}{m} \right\}^{1/p-1/q}$$

and, if  $p < 1$ ,

$$d^m(B_{S_{p,\infty}^N}, S_q^N) \asymp_{p,q} \min \left\{ 1, \frac{N}{m} \right\}^{1/p-1/q}.$$

Before the proof, let us go through some preliminaries. Recall that it is classical to show that, for  $q > p$ ,

$$\rho_s(X)_{S_q} \leq \frac{1}{s^{1/p-1/q}} \|X\|_{S_p}, \tag{5.1}$$

$$\rho_s(X)_{S_q} \leq \frac{D_{p,q}}{s^{1/p-1/q}} \|X\|_{S_{p,\infty}}, \quad D_{p,q} := (q/p - 1)^{-1/q}. \tag{5.2}$$

*5.1. Lower bounds*

In this section we prove a result that will easily imply the desired lower bounds in Theorem 5.1. It is a matrix version of [16, Thm. 2.1] and, just like in their result, we note that the restriction  $q \leq 2$  is not imposed here.

**Proposition 5.2.** *For  $0 < p \leq 1$  and  $p < q \leq \infty$ , there exists a constant  $c_{p,q} > 0$  such that*

$$d^m(B_{S_p^N}, S_q^N) \geq c_{p,q} \min \left\{ 1, \frac{N}{m} \right\}^{1/p-1/q}$$

**Proof.** With  $c = (1/2)^{2/p-1/q}$  and  $\mu = \min\{1, N/2m\}$ , we are going to prove that

$$d^m(B_{S_p}^N, S_q^N) \geq c\mu^{1/p-1/q}.$$

We proceed by contradiction, assuming that  $d^m(B_{S_p}^N, S_q^N) < c\mu^{1/p-1/q}$ . This implies the existence of a linear map  $\mathcal{A} : M_N \rightarrow \mathbb{R}_m$  such that for all  $V \in \ker(\mathcal{A}) \setminus \{0\}$ ,

$$\|V\|_{S_q} < c\mu^{1/p-1/q} \|V\|_{S_p}.$$

For a fixed  $V \in \ker(\mathcal{A}) \setminus \{0\}$ , in view of the inequalities  $\|V\|_{S_p} \leq N^{1/p-1/q} \|V\|_{S_q}$  and  $c \leq (1/2)^{1/p-1/q}$ , we derive  $1 < (\mu N/2)^{1/p-1/q}$ , so  $1 \leq 1/\mu < N/2$ . We then define  $s := \lfloor 1/\mu \rfloor \geq 1$ , so  $2s < N$  and

$$\frac{1}{2\mu} < s \leq \frac{1}{\mu}.$$

Now for  $V \in \ker(\mathcal{A}) \setminus \{0\}$ ,

$$\|V_{[2s]}\|_{S_p} \leq (2s)^{1/p-1/q} \|V_{[2s]}\|_{S_q} \leq (2s)^{1/p-1/q} \|V\|_{S_q} < c(2s\mu)^{1/p-1/q} \|V\|_{S_p} \leq \frac{1}{2^{1/p}} \|V\|_{S_p}$$

and therefore, using that  $\|V\|_{S_p}^p = \|V_{[2s]}\|_{S_p}^p + \|V - V_{[2s]}\|_{S_p}^p$ , we conclude

$$\|V_{[2s]}\|_{S_p}^p \leq \|V - V_{[2s]}\|_{S_p}^p.$$

This means that  $\mathcal{A}$  satisfies the sufficient conditions in [30, Thm. 3], which implies that Schatten  $p$ -quasinorm minimization gives exact recovery of rank  $s$  matrices. By well-known arguments (see, for example, the discussion after the statement of Theorem 2.3 in [3]), this gives

$$m \geq Ns > N \frac{1}{2\mu} \geq \frac{N}{2} \frac{1}{\mu} \geq \frac{N}{2} \frac{2m}{N} = m,$$

a blatant contradiction.  $\square$

### 5.2. Upper bounds

In this subsection we establish a result from which the desired upper bounds in Theorem 5.1 will follow easily. The proof relies on low-rank matrix recovery methods, and the reader will notice similarities with the proof of Theorem 4.1. It should be mentioned that the bound for the case  $p \geq 1$  follows easily from the result of Carl and Defant [8] mentioned in the Introduction together with an interpolation argument, but the bound for the case  $p < 1$  is new (as far as the authors know). Our result is a matrix version of [16, Thm. 3.2], but the essence of the argument can be traced back to Donoho [10, Thm. 9]. As in the case of the result of [16], it is interesting to note that even when  $p < 1$ , an optimal reconstruction map  $\Delta$  for the realization of the number  $E_m(B_{S_{p,\infty}}^N, S_q^N)$  can be chosen to be the  $S_1$ -minimization mapping, at least when  $q \geq 1$ .

**Theorem 5.3.** *For  $0 < p < 1$  and  $p < q \leq 2$ , there exists a linear map  $\mathcal{A} : M_N \rightarrow \mathbb{R}^m$  such that, with  $r = \min\{1, q\}$ ,*

$$\sup_{X \in B_{S_{p,\infty}}^N} \|X - \Delta_r(\mathcal{A}X)\|_{S_q} \leq C_{p,q} \min\left\{1, \frac{N}{m}\right\}^{1/p-1/q},$$

where  $C_{p,q} > 0$  is a constant that depends only on  $p$  and  $q$ .

**Proof.** Let  $C$  be the constant in (2.1) relative to the RIP associated with  $\delta = 1/3$ , say.

*Case 1:  $m \geq CN$ .*

We define  $s \geq 1$  as the largest integer smaller than  $\frac{m}{CN}$ , so that

$$\frac{m}{2CN} < s \leq \frac{m}{CN}. \tag{5.3}$$

Let  $t = 2s$ . It is then possible to find a linear map  $\mathcal{A} : M_N \rightarrow \mathbb{R}^m$  with  $\delta_t(\mathcal{A}) \leq \delta$ . In particular, we have  $\delta_s(\mathcal{A}) \leq \delta$ . Now, given  $Z := X - \Delta_r(\mathcal{A}X) \in \ker \mathcal{A}$ , we decompose  $Z$  into matrices  $Z_{T_1}, Z_{T_2}, Z_{T_3}, \dots$  of rank at most  $s$  by taking the  $s$  largest singular values of  $Z$  for  $Z_{T_1}$ , then the next  $s$  largest ones for  $Z_{T_2}$  and so on.

This easily implies  $(\|Z_{T_k}\|_{S_2}^2 / s)^{1/2} \leq (\|Z_{T_{k-1}}\|_{S_r}^r / s)^{1/r}$ , i.e.,

$$\|Z_{T_k}\|_{S_2} \leq \frac{1}{s^{1/r-1/2}} \|Z_{T_{k-1}}\|_{S_r}, \quad k \geq 2. \tag{5.4}$$

Using the  $r$ -triangle inequality, we have

$$\|Z\|_{S_q}^r = \left\| \sum_{k \geq 1} Z_{T_k} \right\|_{S_q}^r \leq \sum_{k \geq 1} \|Z_{T_k}\|_{S_q}^r \leq \sum_{k \geq 1} (s^{1/q-1/2} \|Z_{T_k}\|_{S_2})^r \leq \sum_{k \geq 1} \left( \frac{s^{1/q-1/2}}{\sqrt{1-\delta}} \|\mathcal{A}Z_{T_k}\|_{\ell_2} \right)^r.$$

The fact that  $Z \in \ker \mathcal{A}$  implies  $\mathcal{A}Z_{T_1} = -\sum_{k \geq 2} \mathcal{A}Z_{T_k}$ . It follows that

$$\begin{aligned} \|Z\|_{S_q}^r &\leq \left( \frac{s^{1/q-1/2}}{\sqrt{1-\delta}} \right)^r \left( \sum_{k \geq 2} \|\mathcal{A}Z_{T_k}\|_{\ell_2} \right)^r + \left( \frac{s^{1/q-1/2}}{\sqrt{1-\delta}} \right)^r \sum_{k \geq 2} \|\mathcal{A}Z_{T_k}\|_{\ell_2}^r \\ &\leq 2 \left( \frac{s^{1/q-1/2}}{\sqrt{1-\delta}} \right)^r \sum_{k \geq 2} \|\mathcal{A}Z_{T_k}\|_{\ell_2}^r \leq 2 \left( \sqrt{\frac{1+\delta}{1-\delta}} s^{1/q-1/2} \right)^r \sum_{k \geq 2} \|Z_{T_k}\|_{S_2}^r. \end{aligned}$$

We then derive, using the inequality (5.4),

$$\|Z\|_{S_q}^r \leq 2 \left( \sqrt{\frac{1+\delta}{1-\delta}} \frac{1}{s^{1/r-1/q}} \right)^r \sum_{k \geq 1} \|Z_{T_k}\|_{S_r}^r.$$

In view of the choice  $\delta = 1/3$  and of (5.3), we deduce

$$\|X - \Delta_r(\mathcal{A}X)\|_{S_q} \leq 2^{1/r} \sqrt{2} \left( \frac{2CN}{m} \right)^{1/r-1/q} \|X - \Delta_r(\mathcal{A}X)\|_{S_r}. \tag{5.5}$$

Moreover, in view of  $\delta_{2s} \leq 1/3$  and of Theorems 4.1 and 2.1, there exists a constant  $C_1 > 0$  such that

$$\|X - \Delta_r(\mathcal{A}X)\|_{S_r} \leq (C_1)^{1/r} \rho_s(x)_{S_r}. \tag{5.6}$$

Finally, using (5.2) and (5.3), we have

$$\rho_s(X)_{S_r} \leq \frac{D_{p,r}}{s^{1/p-1/r}} \leq D_{p,r} \left( \frac{2CN}{m} \right)^{1/p-1/r}. \tag{5.7}$$

Putting (5.5), (5.6), and (5.7) together, we obtain, for any  $x \in B_{p,\infty}^N$ ,

$$\|X - \Delta_r(\mathcal{A}X)\|_{S_q} \leq 2^{1/r} \sqrt{2} C_1^{1/r} D_{p,r} \left( \frac{2CN}{m} \right)^{1/p-1/q}.$$

Case 2:  $m \leq CN$ .

We simply choose the map  $\mathcal{A}$  as the zero map. Then, for any  $X \in B_{S_p, \infty}^N$ , we have

$$\|X - \Delta_r(\mathcal{A}X)\|_{S_q} = \|X\|_{S_q} \leq D_{p,q} \|X\|_{p, \infty} \leq D_{p,q},$$

for some constant  $D_{p,q} > 0$ .

This completes the proof.  $\square$

**Remark 5.4.** When  $p = 1$ , the same proof but using inequality (5.1) instead of (5.2) gives the following: for  $1 < q \leq 2$ , there exists a linear map  $\mathcal{A} : M_N \rightarrow \mathbb{R}^m$  such that,

$$\sup_{X \in B_1^N} \|X - \Delta_1(\mathcal{A}X)\|_{S_q} \leq C_q \min\left\{1, \frac{N}{m}\right\}^{1-1/q},$$

where  $C_q > 0$  is a constant that depends only on  $q$ .

### 5.3. Proof of Theorem 5.1

First, an observation. As in the vector case, the simple inclusion  $B_{S_p}^N \subseteq B_{S_p, \infty}^N$  implies

$$d^m(B_{S_p}^N, S_q^N) \leq d^m(B_{S_p, \infty}^N, S_q^N),$$

hence it suffices to show lower bounds for  $d^m(B_{S_p}^N, S_q^N)$  and upper bounds for  $d^m(B_{S_p, \infty}^N, S_q^N)$ .

The lower bounds follow immediately from Proposition 5.2. When  $0 < p < 1$ , the upper bounds follow from Theorem 5.3. For  $p = 1$ , the upper bound when  $1 \leq m \leq N$  follows from the trivial inequality  $\|X\|_{S_q} \leq \|X\|_{S_1}$ , whereas when  $N \leq m \leq N^2$  it follows from Remark 5.4.  $\square$

### 5.4. Relation to compressive widths

As promised after the proof of our matrix version of the Kashin–Temlyakov theorem, the relationship between the Banach space geometry of the finite-dimensional Schatten  $p$ -classes and matrix recovery goes beyond the norm minimization scheme. Below we use the notation from [17, Sec. 10.1]: the quantities  $E^m$  and  $E_{\text{ada}}^m$  measure the worst-case reconstruction errors of optimal measurement/reconstruction schemes in the nonadaptive and adaptive settings, respectively.

**Theorem 5.5.** *For  $0 < p \leq 1$  and  $p < q \leq 2$ , if  $1 \leq m < N^2$  then the adaptive and nonadaptive compressive widths satisfy*

$$E_{\text{ada}}^m(B_{S_p}^N, S_q^N) \asymp_{p,q} E^m(B_{S_p}^N, S_q^N) \asymp_{p,q} \min\left\{1, \frac{N}{m}\right\}^{1/p-1/q}.$$

**Proof.** Since  $-B_{S_p}^N = B_{S_p}^N$  and  $B_{S_p}^N + B_{S_p}^N \subseteq 2^{1/p}B_{S_p}^N$ , [17, Thm. 10.4] implies

$$d^m(B_{S_p}^N, S_q^N) \leq E_{\text{ada}}^m(B_{S_p}^N, S_q^N) \leq E^m(B_{S_p}^N, S_q^N) \leq 2^{1/p}d^m(B_{S_p}^N, S_q^N)$$

But now, since  $d^m(B_{S_p}^N, S_q^N) = c_{m+1}(id : S_p^N \rightarrow S_q^N)$ , an appeal to Theorem 5.1 finishes the proof.  $\square$

In the  $\ell_p$  case the lower estimate is of particular importance in compressed sensing, since it allows one to prove lower bounds for the number of measurements required to stably recover  $s$ -sparse vectors in  $\mathbb{R}^N$ . In the matrix case, that is no longer the case. Trying it only gives that (under certain conditions), the

minimum number of measurements  $m$  required to stably recover rank  $s$  matrices in  $M_N$  is  $\geq CNs$ , which is not an improvement over the informational-theoretical limit. The reason behind this is that, unlike in the  $\ell_p$  case, there are compressed sensing algorithms (including norm minimization) that give stability with a number of measurements of that order [3].

### Note added in proof

During the publication process of this work we became aware of the closely related one [26], which was uploaded to [arxiv.org](http://arxiv.org) after the present paper was submitted for publication. Their result [26, Prop. 2] is an independently obtained, essentially equivalent version of our [Theorem 4.1](#).

### Acknowledgments

We would like to thank Rachel Ward for suggesting the reference [15], and also thank the Workshop in Analysis in Probability at Texas A&M University. The first author was partially supported by NSF grant DMS-1400588.

### References

- [1] Richard Baraniuk, Mark Davenport, Ronald DeVore, Michael Wakin, A simple proof of the restricted isometry property for random matrices, *Constr. Approx.* 28 (3) (2008) 253–263.
- [2] T. Tony Cai, Anru Zhang, Sparse representation of a polytope and recovery of sparse signals and low-rank matrices, *IEEE Trans. Inform. Theory* 60 (1) (2014) 122–132.
- [3] Emmanuel J. Candès, Yaniv Plan, Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements, *IEEE Trans. Inform. Theory* 57 (4) (2011) 2342–2359.
- [4] Emmanuel J. Candès, Benjamin Recht, Exact matrix completion via convex optimization, *Found. Comput. Math.* 9 (6) (2009) 717–772.
- [5] Emmanuel J. Candès, Terence Tao, Decoding by linear programming, *IEEE Trans. Inform. Theory* 51 (12) (2005) 4203–4215.
- [6] Emmanuel J. Candès, Terence Tao, Near-optimal signal recovery from random projections: universal encoding strategies?, *IEEE Trans. Inform. Theory* 52 (12) (2006) 5406–5425.
- [7] Emmanuel J. Candès, Terence Tao, The power of convex relaxation: near-optimal matrix completion, *IEEE Trans. Inform. Theory* 56 (5) (2010) 2053–2080.
- [8] Bernd Carl, Andreas Defant, Asymptotic estimates for approximation quantities of tensor product identities, *J. Approx. Theory* 88 (2) (1997) 228–256.
- [9] Albert Cohen, Wolfgang Dahmen, Ronald DeVore, Compressed sensing and best  $k$ -term approximation, *J. Amer. Math. Soc.* 22 (1) (2009) 211–231.
- [10] David L. Donoho, Compressed sensing, *IEEE Trans. Inform. Theory* 52 (4) (2006) 1289–1306.
- [11] K. Dvijotham, Maryam Fazel, A nullspace analysis of the nuclear norm heuristic for rank minimization, in: *Proc. of ICASSP 2010*, Dallas, TX, March 2010.
- [12] David E. Edmunds, Jan Lang, Gelfand numbers and widths, *J. Approx. Theory* 166 (2013) 78–84.
- [13] Maryam Fazel, Matrix rank minimization with applications, PhD thesis, Stanford University, 2002.
- [14] Massimo Fornasier, Holger Rauhut, Rachel Ward, Low-rank matrix recovery via iteratively reweighted least squares minimization, *SIAM J. Optim.* 21 (4) (2011) 1614–1640.
- [15] Simon Foucart, Ming-Jun Lai, Sparsest solutions of underdetermined linear systems via  $l_q$ -minimization for  $0 < q \leq 1$ , *Appl. Comput. Harmon. Anal.* 26 (3) (2009) 395–407.
- [16] Simon Foucart, Alain Pajor, Holger Rauhut, Tino Ullrich, The Gelfand widths of  $\ell_p$ -balls for  $0 < p \leq 1$ , *J. Complexity* 26 (6) (2010) 629–640.
- [17] Simon Foucart, Holger Rauhut, *A Mathematical Introduction to Compressive Sensing*, Appl. Numer. Harmon. Anal., Birkhäuser/Springer, New York, 2013.
- [18] A.Yu. Garnaev, E.D. Gluskin, The widths of a Euclidean ball, *Dokl. Akad. Nauk SSSR* 277 (5) (1984) 1048–1052.
- [19] B.S. Kashin, The widths of certain finite-dimensional sets and classes of smooth functions, *Izv. Akad. Nauk SSSR Ser. Mat.* 41 (2) (1977) 334–351, 478.
- [20] B.S. Kashin, V.N. Temlyakov, A remark on the problem of compressed sensing, *Mat. Zametki* 82 (6) (2007) 829–837.
- [21] Lingchen Kong, Naihua Xiu, Exact low-rank matrix recovery via nonconvex Schatten  $p$ -minimization, *Asia-Pac. J. Oper. Res.* (2013), Article ID 1340010, 13 pp.
- [22] Hermann König, Eigenvalues of operators and applications, in: *Handbook of the Geometry of Banach Spaces*, vol. I, North-Holland, Amsterdam, 2001, pp. 941–974.
- [23] Kiryung Lee, Yoram Bresler, Guaranteed minimum rank approximation from linear observations by nuclear norm minimization with an ellipsoidal constraint, arXiv preprint, arXiv:0903.4742.

- [24] Kiryung Lee, Yoram Bresler, ADMiRA: atomic decomposition for minimum rank approximation, *IEEE Trans. Inform. Theory* 56 (9) (2010) 4402–4416.
- [25] Lu Liu, Wei Huang, Di-Rong Chen, Exact minimum rank approximation via Schatten  $p$ -norm minimization, *J. Comput. Appl. Math.* 267 (2014) 218–227.
- [26] Mohammadreza Malek-Mohammadi, Massoud Babaie-Zadeh, Mikael Skoglund, Performance guarantees for Schatten- $p$  quasi-norm minimization in recovery of low-rank matrices, arXiv preprint, arXiv:1407.3716.
- [27] Shahar Mendelson, Alain Pajor, Nicole Tomczak-Jaegermann, Uniform uncertainty principle for Bernoulli and subgaussian ensembles, *Constr. Approx.* 28 (3) (2008) 277–289.
- [28] B.K. Natarajan, Sparse approximate solutions to linear systems, *SIAM J. Comput.* 24 (2) (1995) 227–234.
- [29] Erich Novak, Optimal recovery and  $n$ -widths for convex classes of functions, *J. Approx. Theory* 80 (3) (1995) 390–408.
- [30] Samet Oymak, Karthik Mohan, Maryam Fazel, Babak Hassibi, A simplified approach to recovery conditions for low rank matrices, in: 2011 IEEE International Symposium on Information Theory Proceedings, IEEE, Piscataway, NJ, 2011, pp. 2318–2322.
- [31] Albrecht Pietsch, Eigenvalues and  $s$ -Numbers, *Cambridge Stud. Adv. Math.*, vol. 13, Cambridge University Press, Cambridge, 1987.
- [32] Albrecht Pietsch, *History of Banach Spaces and Linear Operators*, Birkhäuser Boston, Inc., Boston, MA, 2007.
- [33] Benjamin Recht, Maryam Fazel, Pablo A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, *SIAM Rev.* 52 (3) (2010) 471–501.
- [34] Benjamin Recht, Weiyu Xu, Babak Hassibi, Null space conditions and thresholds for rank minimization, *Math. Program., Ser. B* 127 (1) (2011) 175–202.