Jointly Low-Rank and Bisparse Recovery: Questions and Partial Answers

Simon Foucart[*], Rémi Gribonval, Laurent Jacques, and Holger Rauhut

Abstract

This preprint is not a finished product. It is presently intended to gather community feedback.
We investigate the problem of recovering jointly r-rank and s-bisparse matrices from as few linear measurements as possible, considering arbitrary measurements as well as rank-one measurements. In both cases, we show that $m \asymp rs \ln(en/s)$ measurements make the recovery possible in theory, meaning via a nonpractical algorithm. For arbitrary measurements, we also show that practical recovery could be achieved when $m \asymp rs \ln(en/s)$ via an iterative-hard-thresholding algorithm provided one could answer positively a question about head projections for the jointly low-rank and bisparse structure. Some related questions are partially answered in passing. For the rank-one measurements, we suggest on arcane grounds an iterative-hard-thresholding algorithm modified to exploit the nonstandard restricted isometry property obeyed by this type of measurements.

Contents

[*]S. F. partially supported by NSF grants DMS-1622134 and DMS-1664803.

1 Introduction

This whole article is concerned with the inquiry below.

Main Question. *What is the minimal number of linear measurements needed to recover jointly r-rank and s-bisparse symmetric $n \times n$ matrices via an efficient algorithm*?

This minimal number of measurements will be called sample complexity. We will show that it is of the order $rs\ln(en/s)$. Nevertheless, we do not consider the question fully resolved because of the lack of efficient algorithms for realistic measurements and of the limitation of an efficient algorithm to factorized measurements which are not too realistic in applications. Settling the question by providing an efficient algorithm applicable to any type of measurements is therefore still open. Before diving into our investigations, let us start by clarifying a few points.

• What are 'jointly r-rank and s-bisparse symmetric $n \times n$ matrices'?
In this article, we consider exclusively matrices $\mathbf{X} \in \mathbb{R}^{n\times n}$ that are symmetric, i.e., $\mathbf{X}^\top = \mathbf{X}$. The set of r-rank (symmetric) matrices will be denoted as

$$\Sigma^{[r]} := \left\{ \mathbf{X} \in \mathbb{R}^{n\times n} : \mathbf{X}^\top = \mathbf{X},\ \mathrm{rank}(\mathbf{X}) \le r \right\} \tag{1}$$

and the set of s-bisparse (symmetric) matrices will be denoted as

$$\Sigma_{(s)} := \left\{ \mathbf{X} \in \mathbb{R}^{n\times n} : \mathbf{X}^\top = \mathbf{X},\ \mathbf{X}_{\overline{S\times S}} = \mathbf{0} \text{ for some } S \subseteq [\![1:n]\!] \text{ with } |S| = s \right\}. \tag{2}$$

Hence, the jointly r-rank and s-bisparse (symmetric) matrices we are interested in are elements of

$$\Sigma_{(s)}^{[r]} := \Sigma^{[r]} \cap \Sigma_{(s)}. \tag{3}$$

We will often use the fact that $\Sigma_{(s)}^{[r]} + \Sigma_{(s)}^{[r]} \subseteq \Sigma_{(2s)}^{[2r]}$.

• What are the 'linear measurements' considered?
They can be of the arbitrary type

$$y_i = \langle \mathbf{X}, \mathbf{A}_i \rangle_F = \mathrm{tr}(\mathbf{A}_i^\top \mathbf{X}), \qquad i \in [\![1:m]\!], \tag{4}$$

or of the specific (rank-one) type

$$y_i = \langle \mathbf{X}\mathbf{a}_i, \mathbf{a}_i \rangle = \mathrm{tr}(\mathbf{a}_i\mathbf{a}_i^\top \mathbf{X}), \qquad i \in [\![1:m]\!]. \tag{5}$$

Generically, we write $\mathbf{y} = \boldsymbol{\mathcal{A}}(\mathbf{X})$, where $\boldsymbol{\mathcal{A}} : \mathbb{R}^{n\times n} \to \mathbb{R}^m$ is a linear map.

• What is meant by 'recover'?
More than just finding a map $\Delta : \mathbb{R}^m \to \mathbb{R}^{n\times n}$ such that $\Delta(\boldsymbol{\mathcal{A}}(\mathbf{X})) = \mathbf{X}$ for all $\mathbf{X} \in \Sigma_{(s)}^{[r]}$. Indeed, we require the recovery procedure to be stable and robust, in the sense that we want

$$\|\mathbf{X} - \Delta(\boldsymbol{\mathcal{A}}(\mathbf{X}) + \mathbf{e})\| \le C \min_{\mathbf{Z}\in\Sigma_{(s)}^{[r]}} \|\mathbf{X} - \mathbf{Z}\| + D\|\mathbf{e}\| \tag{6}$$

to hold for all $\mathbf{X} \in \mathbb{R}^{n \times n}$ and all $\mathbf{e} \in \mathbb{R}^m$. We give ourselves some freedom on the choice of the three norms appearing in (6). We also require the recovery procedure to be implementable by a practical algorithm, that is, an efficient algorithm whose run-time is at most polynomial in n and m (ideally, a polynomial of low degree, of course).

In our study of the Main Question, we faced the following puzzle.

Question 1. *Is there a practical algorithm that constructs, for each symmetric matrix* $\mathbf{M} \in \mathbb{R}^{n \times n}$*, an index set* S' *of size* $s' = Cs$ *such that*

$$\sum_{i,j \in S'} M_{i,j}^2 \geq c \max_{|S|=s} \sum_{i,j \in S} M_{i,j}^2 \,, \tag{7}$$

where the constant c *may depend on* C*, but not on* s *and* n*?*

In reality, the relevant question for our goal is broader. It involves the projection $P^{[r]}$ onto $\Sigma^{[r]}$.

Question 2. *Is there a practical algorithm that constructs, for each symmetric matrix* $\mathbf{M} \in \mathbb{R}^{n \times n}$*, an index set* S' *of size* $s' = Cs$ *such that*

$$\|P^{[r']}(\mathbf{M}_{S' \times S'})\|_F^2 \geq c \max_{|S|=s} \|P^{[r]}(\mathbf{M}_{S \times S})\|_F^2, \qquad r' = Cr \,, \tag{8}$$

where again the constant C *is independent of* r, s, n*.*

An affirmative answer to Question 2 would provide an affirmative answer to the Main Question for measurements satisfying the so-called restricted isometry property (see below). This is shown in Section 4.

In principle, we are more interested in the measurements of type (5). Indeed, in the particular case $r = 1$, the measurements taken on a matrix of the type $\mathbf{X} = \mathbf{x}\mathbf{x}^\top \in \Sigma_{(s)}^{[1]}$ with an s-sparse $\mathbf{x} \in \mathbb{R}^n$ would read

$$y_i = |\langle \mathbf{a}_i, \mathbf{x} \rangle|^2, \qquad i \in [\![1 : m]\!]. \tag{9}$$

This is exactly the framework of sparse phaseless recovery (except that everything should be written in the complex setting). In this case, the sample complexity is known [10] to be of the order $m \asymp s \ln(en/s)$, although it is unclear if this can be achieved with independent Gaussian vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$.

Remark. Similar problems as studied here appear in the context of low-rank tensor recovery where one would like to project onto the intersection of two or more low rank structures defined by different matricizations. It is NP-hard to compute exact projections and efficiently computable approximate projections are not yet good enough to show low-rank tensor recovery results for corresponding iterative hard thresholding guarantees [16].

2 Theoretical Sample Complexity

Restricted isometry properties have been central in all sorts of structured recovery problems. It is no surprise that another instance of a restricted isometry property plays a key role here, too. The proof sketch is deferred to the appendix.

Theorem 1. Suppose $\mathbf{A}_1, \ldots, \mathbf{A}_m$ are independent random matrices with independent $\mathcal{N}(0, 1/m)$ entries. Then, with failure probability at most $2\exp(-c(\delta)m)$,

$$(1-\delta)\|\mathbf{Z}\|_F^2 \leq \|\boldsymbol{\mathcal{A}}(\mathbf{Z})\|_2^2 \leq (1+\delta)\|\mathbf{Z}\|_F^2 \qquad \text{for all } \mathbf{Z} \in \Sigma_{(s)}^{[r]} \tag{10}$$

provided $m \geq C(\delta) rs \ln(en/s)$.

For the rest of this section, we place ourselves in the situation where the measurement map $\boldsymbol{\mathcal{A}}$ satisfies the restricted isometry property (10), which can occur as soon as m is of the order $rs\ln(en/s)$. We can then propose several robust algorithms that recover $\mathbf{X} \in \Sigma_{(s)}^{[r]}$ from $\mathbf{y} = \boldsymbol{\mathcal{A}}(\mathbf{X}) + \mathbf{e}$. The first obvious candidate is

$$\Delta(\mathbf{y}) = \underset{\mathbf{Z} \in \Sigma_{(s)}^{[r]}}{\operatorname{argmin}} \|\mathbf{y} - \boldsymbol{\mathcal{A}}(\mathbf{Z})\|_2. \tag{11}$$

We immediately see that $\|\mathbf{y} - \boldsymbol{\mathcal{A}}(\Delta(\mathbf{y}))\|_2 \leq \|\mathbf{y} - \boldsymbol{\mathcal{A}}(\mathbf{X})\|_2 = \|\mathbf{e}\|_2$, from where it follows that

$$\|\boldsymbol{\mathcal{A}}(\mathbf{X}) - \boldsymbol{\mathcal{A}}(\Delta(\mathbf{y}))\|_2 \leq \|\mathbf{y} - \boldsymbol{\mathcal{A}}(\Delta(\mathbf{y}))\|_2 + \|\mathbf{e}\|_2 \leq 2\|\mathbf{e}\|_2, \tag{12}$$

and we finally derive that

$$\|\mathbf{X} - \Delta(\boldsymbol{\mathcal{A}}(\mathbf{X}) + \mathbf{e})\|_F \leq \frac{1}{\sqrt{1-\delta}} \|\boldsymbol{\mathcal{A}}(\mathbf{X}) - \boldsymbol{\mathcal{A}}(\Delta(\boldsymbol{\mathcal{A}}(\mathbf{X}) + \mathbf{e}))\|_2 \leq \frac{2}{\sqrt{1-\delta}} \|\mathbf{e}\|_2. \tag{13}$$

However, this scheme is not really an appropriate candidate, since producing $\Delta(\mathbf{y})$ is NP-hard in general (see below).

After a decade or so of ℓ_1-norm and nuclear norm minimizations, the next obvious candidate stands out as

$$\Delta(\mathbf{y}) = \underset{\mathbf{Z} \in \mathbb{R}^{n \times n}}{\operatorname{argmin}} F(\mathbf{Z}) \qquad \text{subject to} \quad \|\mathbf{y} - \boldsymbol{\mathcal{A}}(\mathbf{Z})\|_2 \leq \|\mathbf{e}\|_2, \tag{14}$$

where F is a convex function promoting the joint low-rank and bisparsity structure. The negative results of [15] indicate that reducing the sample complexity below $\min\{rn, s^2 \ln(n/s)\}$ is unattainable when F is a positive combination of the ℓ_1-norm and nuclear norm.

What about a variant of iterative hard thresholding? Consider the sequence $(\mathbf{X}_k)_{k \geq 0}$ defined by

$$\mathbf{X}_{k+1} = P_{(s)}^{[r]}(\mathbf{X}_k + \boldsymbol{\mathcal{A}}^*(\mathbf{y} - \boldsymbol{\mathcal{A}}(\mathbf{X}_k))), \tag{15}$$

where the adjoint of $\boldsymbol{\mathcal{A}}$ is given by

$$\boldsymbol{\mathcal{A}}^* : \mathbf{u} \in \mathbb{R}^m \mapsto \sum_{i=1}^{m} u_i \mathbf{A}_i \in \mathbb{R}^{n \times n}$$

and where $P_{(s)}^{[r]} : \mathbb{R}^{n \times n} \to \Sigma_{(s)}^{[r]}$ denotes the projection onto $\Sigma_{(s)}^{[r]}$, that is, the operator of best approximation from $\Sigma_{(s)}^{[r]}$. One can show (see Appendix or [1]) that if $\Delta(\mathbf{y})$ is defined as a cluster point of $(\mathbf{X}_k)_{k \geq 0}$, then

$$\|\mathbf{X} - \Delta(\boldsymbol{\mathcal{A}}(\mathbf{X}) + \mathbf{e})\|_F \leq C \|\mathbf{e}\|_2 \tag{16}$$

holds for all $\mathbf{X} \in \Sigma_{(s)}^{[r]}$ and all $\mathbf{e} \in \mathbb{R}^m$. Here also the issue is that computing $P_{(s)}^{[r]}$ is NP-hard (see Section 5), which incidentally justifies the NP-hardness of (11) (think of $\boldsymbol{\mathcal{A}} = \mathbf{I}$). What about replacing $P_{(s)}^{[r]}$ by an operator of near-best approximation from $\Sigma_{(s)}^{[r]}$, as in e.g. [8]? After all, if there is any chance for (6) to hold, then such an operator must exist (think again of $\boldsymbol{\mathcal{A}} = \mathbf{I}$). We will in fact construct such an operator in Subsection 5.3. But substituting $P_{(s)}^{[r]}$ by such an operator in the proof of Theorem 12 (see Appendix) is not enough to do the trick.

3 Optimal Sample Complexity with Factorized Measurements

In this section, we show that the optimal sample complexity can be achieved with a practical algorithm in a rather special measurement framework. This framework does not seem realistic for applications, though, so the Main Question remains of interest.

We suppose here that matrices $\mathbf{X} \in \Sigma_{(s)}^{[r]}$ are acquired via measurements in factorized form, namely

$$y_i = \langle \mathbf{X}, \mathbf{B}^\top \mathbf{A}_i \mathbf{B} \rangle, \qquad i \in [\![1 : m]\!], \tag{17}$$

where $\mathbf{A}_1, \ldots, \mathbf{A}_m \in \mathbb{R}^{p \times p}$ allow for low-rank recovery and $\mathbf{B} \in \mathbb{R}^{p \times n}$ allows for sparse recovery. The recovery algorithm proceeds in two steps, which are both practical, i.e., efficiently implementable.

1. Compute $\mathbf{Y}^\sharp \in \mathbb{R}^{p \times p}$ from $\mathbf{y} \in \mathbb{R}^m$ as a solution of the nuclear norm minimization

$$\underset{\mathbf{Y} \in \mathbb{R}^{p \times p}}{\text{minimize}} \ \|\mathbf{Y}\|_* \qquad \text{subject to } \ \langle \mathbf{Y}, \mathbf{A}_i \rangle_F = y_i, \quad i \in [\![1 : m]\!],$$

   or as the output of another low-rank recovery algorithm such as iterative hard thresholding.

2. Compute $\mathbf{X}^\sharp \in \mathbb{R}^{n \times n}$ from $\mathbf{Y}^\sharp$ as the output of the HiHTP algorithm with measurement map $\boldsymbol{\mathcal{B}} : \mathbf{Z} \in \mathbb{R}^{n \times n} \mapsto \mathbf{B}\mathbf{Z}\mathbf{B}^\top \in \mathbb{R}^{p \times p}$.

Although we refer to [17, 18] for the exact formulation of the HiHTP algorithm, a few words about the concept of hierarchical sparsity are in order before we state our result about the two-step recovery procedure above. A matrix is said to be (s,t)-hierachical sparse (or simply (s,t)-sparse) if at most s of its columns are nonzero and each of these columns possesses at most t nonzero entries. Thus, s-bisparse matrices are in particular (s,s)-sparse. The HiHTP algorithm essentially relies on the possibility to compute the projection (operator of best approximation) onto (s,t)-sparse matrices. In contrast to the projection onto s-bisparse matrices, this is indeed an easy task: first, select the t largest absolute entries in each column and calculate the resulting ℓ_2-norm, then select the s columns with the largest of these ℓ_2-norms.

Theorem 2. Let $\mathbf{A}_1, \ldots, \mathbf{A}_m \in \mathbb{R}^{p \times p}$ be independent standard Gaussian matrices and let $\mathbf{B} \in \mathbb{R}^{p \times n}$ be a standard Gaussian matrix independent of $\mathbf{A}_1, \ldots, \mathbf{A}_m$. If

$$p \asymp s \ln(en/s) \qquad \text{and} \qquad m \asymp rp, \tag{18}$$

so that $m \asymp rs \ln(en/s)$, then the probability that every $\mathbf{X} \in \Sigma_{(s)}^{[r]}$ is exactly recovered from $y_i = \langle \mathbf{X}, \mathbf{B}^\top \mathbf{A}_i \mathbf{B} \rangle$, $i \in [\![1:m]\!]$, via the above two-step procedure is at least $1 - 2\exp(-cp)$.

Proof. First, notice that the matrix $\mathbf{B}\mathbf{X}\mathbf{B}^\top \in \mathbb{R}^{p \times p}$ has rank at most r, since $\mathbf{X}$ has rank at most r, and that it satisfies

$$\langle \mathbf{B}\mathbf{X}\mathbf{B}^\top, \mathbf{A}_i \rangle_F = \mathrm{tr}(\mathbf{A}_i^\top \mathbf{B}\mathbf{X}\mathbf{B}^\top) = \mathrm{tr}(\mathbf{B}^\top \mathbf{A}_i^\top \mathbf{B}\mathbf{X}) = \langle \mathbf{X}, \mathbf{B}^\top \mathbf{A}_i \mathbf{B} \rangle_F = y_i, \qquad i \in [\![1:m]\!]. \tag{19}$$

Since $\mathbf{A}_1, \ldots, \mathbf{A}_m \in \mathbb{R}^{p \times p}$ are independent standard Gaussian matrices and $m \asymp rp$, it is by now well-known (see e.g. [3, 11]) that, with failure probability at most $\exp(-cm)$, the matrix $\mathbf{B}\mathbf{X}\mathbf{B}^\top$ is recovered via nuclear norm minimization (or another suitable algorithm), so that $\mathbf{Y}^\sharp = \mathbf{B}\mathbf{X}\mathbf{B}^\top$.

Second, since the matrix $\mathbf{X} \in \mathbb{R}^{n \times n}$ is (s,s)-sparse and satisfies $\boldsymbol{\mathcal{B}}(\mathbf{X}) = \mathbf{B}\mathbf{X}\mathbf{B}^\top = \mathbf{Y}^\sharp$, Theorem 1 of [17] implies that the matrix $\mathbf{X}$ will be exactly recovered via HiHTP as long as the so-called HiRIP of order $(3s, 2s)$ holds. According to Theorem 1 of [18], the latter is satisfied when $\mathbf{B}$ obeys a standard RIP, and the latter is indeed fulfilled with failure at most $\exp(-cp)$ by the matrix $\mathbf{B}$ (or rather by a renormalization of it), because $\mathbf{B} \in \mathbb{R}^{p \times n}$ is a standard Gaussian matrix with $m \asymp rp$.

All in all, exact recovery of $\mathbf{X}$ is guaranteed after the two steps with failure probability bounded by $\exp(-cm) + \exp(-cp) \le 2\exp(-cp)$. □

Remark. It is possible to extend Theorem 2 beyond the strictly Gaussian setting. In particular, if $\mathbf{A}_1, \ldots, \mathbf{A}_m$ take the form $\mathbf{A}_i = \mathbf{a}_i \mathbf{a}_i^\top$ for some independent standard Gaussian vectors $\mathbf{a}_i \in \mathbb{R}^p$, then the first-step recovery of $\mathbf{B}\mathbf{X}\mathbf{B}^\top$ can still be achieved via nuclear norm minimization (see [2, 11, 12]) or by some modified iterative hard thresholding algorithm (see [7]). Note that the measurements made on $\mathbf{X} \in \mathbb{R}^{n \times n}$ are in this case rank-one measurements given by $y_i = \langle \mathbf{X}\mathbf{a}_i', \mathbf{a}_i' \rangle$, where $\mathbf{a}_i' := \mathbf{B}^\top \mathbf{a}_i$.

4 Towards Practical Sample Complexity

In most scenarios, the measurement map is not of the factorized type considered in the previous section, so the two-step procedure cannot even be executed. It is therefore still relevant to search for practical recovery algorithms that can be applied with arbitrary measurement schemes and study the sample complexity using e.g. Gaussian measurements. As mentioned at the end of Section 2, a difficulty occurs when one tries to use a near-best approximation operator instead of the best approximation operator $P_{(s)}^{[r]}$ in the iterative hard thresholding algorithm (15). Such a difficulty was also encountered in model-based compressive sensing. A workaround was found in [9]. As we will see below, our attempt to imitate it prompted Question 2.

Let us start with the observation that any of the structures $\Sigma_{(s)}$, $\Sigma^{[r]}$, or $\Sigma_{(s)}^{[r]}$ is a union of subspaces, which we generically write as

$$\Sigma = \bigcup_{V \in \mathcal{V}_\Sigma} V.$$

Then the projection onto Σ, i.e., the operator of best approximation from Σ with respect to the Frobenius norm, acts on any $\mathbf{M} \in \mathbb{R}^{n \times n}$ via

$$P_\Sigma(\mathbf{M}) = P_{V(\mathbf{M})}(\mathbf{M}) \tag{20}$$

where

$$V(\mathbf{M}) = \underset{V \in \mathcal{V}_\Sigma}{\operatorname{argmin}} \|\mathbf{M} - P_V(\mathbf{M})\|_F^2 \tag{21}$$

$$= \underset{V \in \mathcal{V}_\Sigma}{\operatorname{argmax}} \|P_V(\mathbf{M})\|_F^2, \tag{22}$$

and P_V evidently denotes the orthogonal projection onto the subspace V. By analogy with the vector case, we can think of (21) as a 'tail' property for the projection P_Σ and of (22) as a 'head' property. We keep this terminology introduced in [9] when relaxing the notion of projection. Precisely, we shall call an operator $T : \mathbb{R}^{n \times n} \to \Sigma$ a tail projection for Σ with constant $C_T \geq 1$ (or near best approximation from Σ with constant C_T) if

$$\|\mathbf{M} - T(\mathbf{M})\|_F \leq C_T \|\mathbf{M} - P_\Sigma(\mathbf{M})\|_F \qquad \text{for all } \mathbf{M} \in \mathbb{R}^{n \times n}. \tag{23}$$

We may have to relax this notion further by allowing the operator T to map into a bigger set $\Sigma' \supseteq \Sigma$. Thus, by tail projection for Σ into Σ' with constant C_T, we mean an operator $T : \mathbb{R}^{n \times n} \to \Sigma'$ which satisfies the tail condition (23). Similarly, an operator $H : \mathbb{R}^{n \times n} \to \Sigma$ is called a head projection for Σ with constant $c_H \leq 1$ if

$$\|H(\mathbf{M})\|_F \geq c_H \|P_\Sigma(\mathbf{M})\|_F \qquad \text{for all } \mathbf{M} \in \mathbb{R}^{n \times n}. \tag{24}$$

A head projection for Σ into $\Sigma' \supseteq \Sigma$ with constant c_H is an operator $H : \mathbb{R}^{n \times n} \to \Sigma'$ which satisfies the head condition (24).

At this point, it is worth mentioning (see Appendix) that the (genuine) projection onto $\Sigma_{(s)}^{[r]}$ acts on any $\mathbf{M} \in \mathbb{R}^{n\times n}$ via

$$P_{(s)}^{[r]}(\mathbf{M}) = P^{[r]}(\mathbf{M}_{S_\star \times S_\star}), \qquad \text{where } S_\star = \operatorname*{argmax}_{|S|=s} \|P^{[r]}(\mathbf{M}_{S\times S})\|_F. \tag{25}$$

In Section 5, we will see that we can produce a tail projection for $\Sigma_{(s)}^{[r]}$. Whether one can produce a head projection for $\Sigma_{(s)}^{[r]}$ into $\Sigma_{(s')}^{[r']}$ with r' and s' proportional to r and s is exactly Question 2. If it has an affirmative answer, then we can perform joint low-rank and bisparse recovery via a variant of iterative hard thresholding from only $m \asymp rs\ln(en/s)$ measurements, as a consequence of the result stated and proved below in the idealized setting where there is no measurement error.

Theorem 3. Let T be a tail projection for $\Sigma_{(s)}^{[r]}$ with constant $C_T \geq 1$ and let H be a head projection for $\Sigma_{(s)}^{[r]}$ into $\Sigma_{(2s)}^{[2r]}$ with constant $c_H \leq 1$ which additionally takes the form

$$H(\mathbf{M}) = P^{[2r]}(\mathbf{M}_{S'\times S'}) \qquad \text{for some index set } S' \text{ (depending on } \mathbf{M}) \text{ of size } 2s. \tag{26}$$

If $(1+C_T)^2(1-c_H^2) < 1$ and if the restricted isometry property (10) holds on $\Sigma_{(4s)}^{[4r]}$ with constant $\delta > 0$ small enough to have

$$\rho := (1+C_T)^2(1 - c_H^2(1-\delta)^2 + 2\delta(1+\delta)) < 1, \tag{27}$$

then any $\mathbf{X} \in \Sigma_{(s)}^{[r]}$ acquired from $\mathbf{y} = \boldsymbol{\mathcal{A}}(\mathbf{X})$ is recovered as the limit of the sequence $(\mathbf{X}_k)_{k\geq 0}$ defined by

$$\mathbf{X}_{k+1} = T[\mathbf{X}_k + H(\boldsymbol{\mathcal{A}}^*(\mathbf{y} - \boldsymbol{\mathcal{A}}(\mathbf{X}_k)))]. \tag{28}$$

Proof. We shall prove that, for any $k \geq 0$,

$$\|\mathbf{X} - \mathbf{X}_{k+1}\|_F^2 \leq \rho\, \|\mathbf{X} - \mathbf{X}_k\|_F^2. \tag{29}$$

The tail property guarantees that

$$\|[\mathbf{X}_k + H(\boldsymbol{\mathcal{A}}^*(\mathbf{y} - \boldsymbol{\mathcal{A}}(\mathbf{X}_k)))] - \mathbf{X}_{k+1}\|_F \leq C_T \|[\mathbf{X}_k + H(\boldsymbol{\mathcal{A}}^*(\mathbf{y} - \boldsymbol{\mathcal{A}}(\mathbf{X}_k)))] - \mathbf{X}\|_F \tag{30}$$

and the triangle inequality then yields[1]

$$\|\mathbf{X} - \mathbf{X}_{k+1}\|_F \leq (1+C_T)\|[\mathbf{X}_k + H(\boldsymbol{\mathcal{A}}^*(\mathbf{y} - \boldsymbol{\mathcal{A}}(\mathbf{X}_k)))] - \mathbf{X}\|_F. \tag{31}$$

We now concentrate on bounding $\|[\mathbf{X}_k + H(\boldsymbol{\mathcal{A}}^*(\mathbf{y} - \boldsymbol{\mathcal{A}}(\mathbf{X}_k)))] - \mathbf{X}\|_F = \|\mathbf{Z} - H(\boldsymbol{\mathcal{A}}^*\boldsymbol{\mathcal{A}}(\mathbf{Z}))\|_F$, where we have set $\mathbf{Z} := \mathbf{X} - \mathbf{X}_k \in \Sigma_{(2s)}^{[2r]}$. By expanding the square, we obtain

$$\begin{aligned} \|\mathbf{Z} - H(\boldsymbol{\mathcal{A}}^*\boldsymbol{\mathcal{A}}(\mathbf{Z}))\|_F^2 &= \|\mathbf{Z}\|_F^2 + \|H(\boldsymbol{\mathcal{A}}^*\boldsymbol{\mathcal{A}}(\mathbf{Z}))\|_F^2 - 2\langle \mathbf{Z}, H(\boldsymbol{\mathcal{A}}^*\boldsymbol{\mathcal{A}}(\mathbf{Z}))\rangle_F \\ &= \|\mathbf{Z}\|_F^2 + \|H(\boldsymbol{\mathcal{A}}^*\boldsymbol{\mathcal{A}}(\mathbf{Z}))\|_F^2 - 2\langle \boldsymbol{\mathcal{A}}^*\boldsymbol{\mathcal{A}}(\mathbf{Z}), H(\boldsymbol{\mathcal{A}}^*\boldsymbol{\mathcal{A}}(\mathbf{Z}))\rangle_F \\ &\qquad - 2\langle \mathbf{Z} - \boldsymbol{\mathcal{A}}^*\boldsymbol{\mathcal{A}}(\mathbf{Z}), H(\boldsymbol{\mathcal{A}}^*\boldsymbol{\mathcal{A}}(\mathbf{Z}))\rangle_F. \end{aligned} \tag{32}$$

[1] It is probably possible to replace $1 + C_T$ by a constant arbitrarily close to 1 if T mapped into $\Sigma_{(s')}^{[r']}$ with r' and s' proportional to r and s (with proportionality constant increasing when C_T decreases), as in [19] for the sparse vector case and in [7] for the low-rank matrix case. This would allow us to eliminate the condition $(1+C_T)^2(1-c_H^2) < 1$.

In view of the form (26) of the head projection, followed by the facts that $P^{[2r]}$ acts locally as an orthogonal projection and that it preserves the bisupport of a matrix, we observe that

$$(33)\quad \begin{aligned}\|H(\mathcal{A}^*\mathcal{A}(\mathbf{Z}))\|_F^2 &= \langle P^{[2r]}(\mathcal{A}^*\mathcal{A}(\mathbf{Z})_{S'\times S'}), P^{[2r]}(\mathcal{A}^*\mathcal{A}(\mathbf{Z})_{S'\times S'})\rangle_F \\ &= \langle \mathcal{A}^*\mathcal{A}(\mathbf{Z})_{S'\times S'}, P^{[2r]}(\mathcal{A}^*\mathcal{A}(\mathbf{Z})_{S'\times S'})\rangle_F = \langle \mathcal{A}^*\mathcal{A}(\mathbf{Z}), P^{[2r]}(\mathcal{A}^*\mathcal{A}(\mathbf{Z})_{S'\times S'})\rangle_F \\ &= \langle \mathcal{A}^*\mathcal{A}(\mathbf{Z}), H(\mathcal{A}^*\mathcal{A}(\mathbf{Z}))\rangle_F.\end{aligned}$$

Substituting the latter into (32) gives

$$(34)\qquad \|\mathbf{Z} - H(\mathcal{A}^*\mathcal{A}(\mathbf{Z}))\|_F^2 = \|\mathbf{Z}\|_F^2 - \|H(\mathcal{A}^*\mathcal{A}(\mathbf{Z}))\|_F^2 - 2\langle \mathbf{Z} - \mathcal{A}^*\mathcal{A}(\mathbf{Z}), H(\mathcal{A}^*\mathcal{A}(\mathbf{Z}))\rangle_F.$$

The inner product term is small in absolute value. Indeed, in view of Lemma 11 (see Appendix), we have

$$(35)\qquad |\langle \mathbf{Z} - \mathcal{A}^*\mathcal{A}(\mathbf{Z}), H(\mathcal{A}^*\mathcal{A}(\mathbf{Z}))\rangle_F| \le \delta\|\mathbf{Z}\|_F\|H(\mathcal{A}^*\mathcal{A}(\mathbf{Z}))\|_F \le \delta(1+\delta)\|\mathbf{Z}\|_F^2,$$

where the bound on $\|H(\mathcal{A}^*\mathcal{A}(\mathbf{Z}))\|_F$ followed from the observation (33) and the restricted isometry property (10), according to

$$(36)\qquad \begin{aligned}\|H(\mathcal{A}^*\mathcal{A}(\mathbf{Z}))\|_F^2 &= \langle \mathcal{A}^*\mathcal{A}(\mathbf{Z}), H(\mathcal{A}^*\mathcal{A}(\mathbf{Z}))\rangle_F = \langle \mathcal{A}(\mathbf{Z}), \mathcal{A}(H(\mathcal{A}^*\mathcal{A}(\mathbf{Z})))\rangle_F \\ &\le \|\mathcal{A}(\mathbf{Z})\|_F\|\mathcal{A}(H(\mathcal{A}^*\mathcal{A}(\mathbf{Z})))\|_F \le (1+\delta)\|\mathbf{Z}\|_F\|H(\mathcal{A}^*\mathcal{A}(\mathbf{Z}))\|_F.\end{aligned}$$

It now remains to prove that $\|H(\mathcal{A}^*\mathcal{A}(\mathbf{Z}))\|_F^2$ is large, and this is where the head condition comes into play. Precisely, assuming that $\mathbf{Z}$ is supported on $S''\times S''$ with $|S''| \le 2s$, we know on the one hand that

$$(37)\qquad \|H(\mathcal{A}^*\mathcal{A}(\mathbf{Z}))\|_F \ge c_H\|P^{[2r]}(\mathcal{A}^*\mathcal{A}(\mathbf{Z})_{S''\times S''})\|_F.$$

On the other hand, using in particular the restricted isometry property (10) and Von Neumann's trace inequality combined with the fact that $\mathbf{Z}$ has rank at most $2r$, we obtain

$$(38)\quad \begin{aligned}(1-\delta)\|\mathbf{Z}\|_F^2 &\le \|\mathcal{A}(\mathbf{Z})\|_2^2 = \langle \mathbf{Z}, \mathcal{A}^*\mathcal{A}(\mathbf{Z})\rangle_F = \langle \mathbf{Z}, \mathcal{A}^*\mathcal{A}(\mathbf{Z})_{S''\times S''}\rangle_F \\ &\le \sum_{i=1}^{2r}\sigma_i(\mathbf{Z})\sigma_i(\mathcal{A}^*\mathcal{A}(\mathbf{Z})_{S''\times S''}) \le \left[\sum_{i=1}^{2r}\sigma_i(\mathbf{Z})^2\right]^{1/2}\left[\sum_{i=1}^{2r}\sigma_i(\mathcal{A}^*\mathcal{A}(\mathbf{Z})_{S''\times S''})^2\right]^{1/2} \\ &= \|\mathbf{Z}\|_F\|P^{[2r]}(\mathcal{A}^*\mathcal{A}(\mathbf{Z})_{S''\times S''})\|_F.\end{aligned}$$

Combining (37) and (38) yields

$$(39)\qquad \|H(\mathcal{A}^*\mathcal{A}(\mathbf{Z}))\|_F \ge c_H(1-\delta)\|\mathbf{Z}\|_F.$$

Substituting (39) and (35) into (34), we deduce that

$$(40)\qquad \|\mathbf{Z} - H(\mathcal{A}^*\mathcal{A}(\mathbf{Z}))\|_F^2 \le \left(1 - c_H^2(1-\delta)^2 + 2\delta(1+\delta)\right)\|\mathbf{Z}\|_F^2.$$

Finally, using (31), we arrive that

$$(41)\qquad \|\mathbf{X} - \mathbf{X}_{k+1}\|_F^2 \le (1+C_T)^2\left(1 - c_H^2(1-\delta)^2 + 2\delta(1+\delta)\right)\|\mathbf{X} - \mathbf{X}_k\|_F^2,$$

which is the objective announced in (29). □

5 Tail and Head Projections

In this section, we gather some information about the construction of computable tail and head projections for each of the structures $\Sigma^{[r]}$, $\Sigma_{(s)}$, and $\Sigma^{[r]}_{(s)}$. We work under the implicit assumption that the domain of all these projections is the space of symmetric matrices, i.e., the projections are only applied to matrices $\mathbf{M} \in \mathbb{R}^{n \times n}$ satisfying $\mathbf{M}^\top = \mathbf{M}$.

5.1 Low-rank structure

There is no difficulty whatsoever here — even the exact projection $P^{[r]} : \mathbb{R}^{n \times n} \to \Sigma^{[r]}$ is accessible. Indeed, it is well known that if $\mathbf{X} \in \mathbb{R}^{n \times n}$ has singular value decomposition

$$\mathbf{X} = \sum_{i=1}^{n} \sigma_i(\mathbf{X}) \mathbf{u}_i \mathbf{v}_i^\top \tag{42}$$

where the singular values $\sigma_1(\mathbf{X}) \geq \cdots \geq \sigma_n(\mathbf{X}) \geq 0$ are arranged in nondecreasing order, then the projection of $\mathbf{X}$ onto the set of rank-r matrices is obtained by truncating this decomposition to include only the first r summands, i.e.,

$$P^{[r]}(\mathbf{X}) = \sum_{i=1}^{r} \sigma_i(\mathbf{X}) \mathbf{u}_i \mathbf{v}_i^\top. \tag{43}$$

Note that $P^{[r]}(\mathbf{M})$ is symmetric whenever $\mathbf{M}$ itself is symmetric.

5.2 Bisparsity structure

Quickly stated, exact projections for $\Sigma_{(s)}$ are NP-hard, but there are computable tail projections for $\Sigma_{(s)}$. Head projections for $\Sigma_{(s)}$ are still NP-hard if they are forced to map exactly into $\Sigma_{(s)}$. The situation is unclear if they are allowed to map into $\Sigma_{(s')}$ with $s' > s$ — this is in fact Question 1. We provide a few incomplete results related to this situation.

Exact projection. Finding the exact projection for $\Sigma_{(s)}$ amounts to solving the problem

$$\underset{|S|=s}{\text{maximize}} \, \|\mathbf{M}_{S \times S}\|_F^2. \tag{44}$$

This is NP-hard even with the restriction that $\mathbf{M}$ is an adjacency matrix of a graph because it then reduces to the densest k-subgraph problem, which is known to be NP-hard [14].

Tail projections. There is a simple procedure to obtain a practical tail projection for $\Sigma_{(s)}$, as described below.

Proposition 4. Given a symmetric matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$, let $S_\star$ denote an index set corresponding to s columns of $\mathbf{M}$ with largest ℓ_2-norms, i.e.,

$$S_\star = \underset{|S|=s}{\operatorname{argmin}} \|\mathbf{M} - \mathbf{M}_{:\times S}\|_F. \tag{45}$$

Then

$$\|\mathbf{M} - \mathbf{M}_{S_\star \times S_\star}\|_F \le \sqrt{2} \min_{|S|=s} \|\mathbf{M} - \mathbf{M}_{S \times S}\|_F. \tag{46}$$

Proof. For any index set T, the symmetry of $\mathbf{M}$ imposes that $\|\mathbf{M}_{\overline{T}\times T}\|_F^2 = \|\mathbf{M}_{T \times \overline{T}}\|_F^2$, hence

$$\|\mathbf{M} - \mathbf{M}_{T\times T}\|_F^2 = \|\mathbf{M}_{T\times\overline{T}}\|_F^2 + \|\mathbf{M}_{\overline{T}\times T}\|_F^2 + \|\mathbf{M}_{\overline{T}\times\overline{T}}\|_F^2 = 2\|\mathbf{M}_{T\times\overline{T}}\|_F^2 + \|\mathbf{M}_{\overline{T}\times\overline{T}}\|_F^2. \tag{47}$$

In view of $\|\mathbf{M}_{T\times\overline{T}}\|_F^2 + \|\mathbf{M}_{\overline{T}\times\overline{T}}\|_F^2 = \|\mathbf{M}_{:\times\overline{T}}\|_F^2 = \|\mathbf{M} - \mathbf{M}_{:\times T}\|_F^2$, we deduce that

$$\|\mathbf{M} - \mathbf{M}_{:\times T}\|_F^2 \le \|\mathbf{M} - \mathbf{M}_{T\times T}\|_F^2 \le 2\|\mathbf{M} - \mathbf{M}_{:\times T}\|_F^2. \tag{48}$$

Applying the latter with T equal to $S_\star$ and with T equal to an arbitrary index set S of size s shows that

$$\|\mathbf{M} - \mathbf{M}_{S_\star\times S_\star}\|_F^2 \le 2\|\mathbf{M} - \mathbf{M}_{:\times S_\star}\|_F^2 \le 2\|\mathbf{M} - \mathbf{M}_{:\times S}\|_F^2 \le 2\|\mathbf{M} - \mathbf{M}_{S\times S}\|_F^2, \tag{49}$$

which yields the required result after taking the square root. □

Head projections. The literature on the densest k-subgraph problem informs us that finding a head projection for $\Sigma_{(s)}$ is also an NP-hard problem [14]. In our setting, though, there is no harm in relaxing the head projection to map into $\Sigma_{(s')}$ with $s' = Cs$, $C \ge 1$. In this regard, Question 1 asks if one can actually compute a head projection for $\Sigma_{(s)}$ into $\Sigma_{(s')}$. We do not have a definite answer for it, but we highlight a few observations featuring a nonabsolute constant c_H.

Proposition 5. There is a practical algorithm that yields a head projection for $\Sigma_{(s)}$ into $\Sigma_{(2s)}$ with constant $c_H = \sqrt{s/n}$.

Proof. For a symmetric matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$, we successively define index sets R and C (for rows and columns) of size s by

$$R = \underset{|S|=s}{\operatorname{argmax}} \|\mathbf{M}_{S\times :}\|_F^2, \tag{50}$$

$$C = \underset{|S|=s}{\operatorname{argmax}} \|\mathbf{M}_{R\times S}\|_F^2. \tag{51}$$

The algorithm consists in returning $\mathbf{M}_{T \times T}$ where $T := R \cup C$. It is painless to see that, for an arbitrary index set S of size s,

$$\|\mathbf{M}_{T\times T}\|_F^2 \geq \|\mathbf{M}_{R\times C}\|_F^2 \geq \frac{s}{n}\|\mathbf{M}_{R\times :}\|_F^2 \geq \frac{s}{n}\|\mathbf{M}_{S\times :}\|_F^2 \geq \frac{s}{n}\|\mathbf{M}_{S\times S}\|_F^2, \tag{52}$$

which concludes the proof. □

When $n > s^2$ (which is the most realistic situation from our perspective), the previous observation is superseded by the following one.

Proposition 6. There is a practical algorithm that yields a head projection for $\Sigma_{(s)}$ with constant $c_H = 1/\sqrt{s}$.

Proof. For a symmetric matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$, we define index sets $S_1, \ldots, S_n$ of size s by

$$S_j = \underset{|S|=s, S\ni j}{\operatorname{argmax}} \|\mathbf{M}_{S\times j}\|_2^2 \tag{53}$$

and we then consider an index $j_\star$ such that

$$j_\star = \underset{j\in[\![1:n]\!]}{\operatorname{argmax}} \|\mathbf{M}_{S_j\times j}\|_2^2. \tag{54}$$

The algorithm consists in returning $\mathbf{M}_{S_{j_\star}\times S_{j_\star}}$. It is painless to see that, for an arbitrary index set S of size s,

$$\|\mathbf{M}_{S\times S}\|_F^2 = \sum_{j\in S}\|\mathbf{M}_{S\times j}\|_2^2 \leq \sum_{j\in S}\|\mathbf{M}_{S_j\times j}\|_2^2 \leq s\|\mathbf{M}_{S_{j_\star}\times j_\star}\|_2^2 \leq s\|\mathbf{M}_{S_{j_\star}\times S_{j_\star}}\|_F^2, \tag{55}$$

which concludes the proof. □

As a final remark, we show that head projections can be computed for specific symmetric matrices, e.g. matrices of rank one.

Proposition 7. There is a practical algorithm that yields a head projection for $\Sigma_{(s)}$ into $\Sigma_{(rs)}$ with constant $c_H = 1/\sqrt{r}$ when applied to r-rank positive semidefinite matrices.

Proof. Consider a matrix $\mathbf{M} = \sum_{k=1}^r \mathbf{v}_k\mathbf{v}_k^\top$ and, for each $k \in [\![1:r]\!]$, let us denote by S_k an index set of s largest absolute entries of $\mathbf{v}_k$. With $S_\star := S_1 \cup \cdots \cup S_r$, we are going to show that, for any index set S of size s,

$$\|\mathbf{M}_{S_\star\times S_\star}\|_F \geq \frac{1}{\sqrt{r}}\|\mathbf{M}_{S\times S}\|_F. \tag{56}$$

To do so, we start by writing

$$M_{i,j}^2 = \left(\sum_{k=1}^r (\mathbf{v}_k)_i(\mathbf{v}_k)_j\right)^2 = \sum_{k,\ell=1}^r (\mathbf{v}_k)_i(\mathbf{v}_k)_j(\mathbf{v}_\ell)_i(\mathbf{v}_\ell)_j. \tag{57}$$

Then, for any index set T, in view of

$$\|\mathbf{M}_{T\times T}\|_F^2 = \sum_{i,j\in T}\sum_{k,\ell=1}^r (\mathbf{v}_k)_i(\mathbf{v}_k)_j(\mathbf{v}_\ell)_i(\mathbf{v}_\ell)_j = \sum_{k,\ell=1}^r\sum_{i,j\in T}(\mathbf{v}_k)_i(\mathbf{v}_\ell)_i(\mathbf{v}_k)_j(\mathbf{v}_\ell)_j = \sum_{k,\ell=1}^r\left(\sum_{i\in T}(\mathbf{v}_k)_i(\mathbf{v}_\ell)_i\right)^2, \tag{58}$$

we derive on the one hand that

$$\|\mathbf{M}_{T\times T}\|_F^2 \geq \sum_{k=1}^r\left(\sum_{i\in T}(\mathbf{v}_k)_i^2\right)^2 \tag{59}$$

and on the other hand, by the Cauchy–Schwarz inequality applied twice, that

$$\|\mathbf{M}_{T\times T}\|_F^2 \leq \sum_{k,\ell=1}^r\left(\sum_{i\in T}(\mathbf{v}_k)_i^2\right)\left(\sum_{i\in T}(\mathbf{v}_\ell)_i^2\right) = \left(\sum_{k=1}^r\sum_{i\in T}(\mathbf{v}_k)_i^2\right)^2 \leq r\sum_{k=1}^r\left(\sum_{i\in T}(\mathbf{v}_k)_i^2\right)^2. \tag{60}$$

Applying (60) with $T=S$ and using the defining property of each S_k and of $S_\star$, we obtain

$$\|\mathbf{M}_{S\times S}\|_F^2 \leq r\sum_{k=1}^r\left(\sum_{i\in S}(\mathbf{v}_k)_i^2\right)^2 \leq r\sum_{k=1}^r\left(\sum_{i\in S_k}(\mathbf{v}_k)_i^2\right)^2 \leq r\sum_{k=1}^r\left(\sum_{i\in S_\star}(\mathbf{v}_k)_i^2\right)^2 \leq r\|\mathbf{M}_{S_\star\times S_\star}\|_F^2, \tag{61}$$

the last inequality being (59) applied with $T=S_\star$. The prospective inequality (56) is proved. □

5.3 Joint low-rank and bisparsity structure

Quickly stated, exact projections for $\Sigma_{(s)}^{[r]}$ are NP-hard, but there are computable tail projections for $\Sigma_{(s)}^{[r]}$. Head projections for $\Sigma_{(s)}^{[r]}$ are still NP-hard if they are forced to map exactly into $\Sigma_{(s)}^{[r]}$. The situation is unclear if they are allowed to map into $\Sigma_{(s')}^{[r']}$ with $r'>r$ and $s'>s$ — this is in fact Question 2. We provide a few incomplete results related to this situation.

Exact projections. We already know from Subsection 5.2 that it is NP-hard to find the exact projection onto $\Sigma_{(s)}^{[r]}$ in general, since we are talking about exact projection onto $\Sigma_{(s)}$ when $r=n$. But we are more interested in the case where r is a small constant, say $r=1$ as a prototype. Then finding the exact projection onto $\Sigma_{(s)}^{[1]}$ amounts to solving the problem

$$\underset{|S|=s}{\text{maximize}}\ \|P^{[1]}(\mathbf{M}_{S\times S})\|_F = \sigma_{\max}(\mathbf{M}_{S\times S}). \tag{62}$$

Thus, when $\mathbf{M}$ is a positive semidefinite matrix, we consider the problem

$$\underset{\|\mathbf{x}\|_0\leq s,\|\mathbf{x}\|_2=1}{\text{maximize}}\ \langle\mathbf{M}\mathbf{x},\mathbf{x}\rangle. \tag{63}$$

This is the so-called sparse principal component analysis problem, which is NP-hard [13].

Tail projections. There is a fairly simple procedure to create a practical tail projection for $\Sigma_{(s)}^{[r]}$. It is based on the availability of tail projections for both $\Sigma^{[r]}$ and $\Sigma_{(s)}$. The argument is in fact valid for any two 'structures' Σ' and Σ'' such that Σ' is compatible with a tail projection T'' for Σ'', in the sense that

$$\mathbf{Z} \in \Sigma' \Longrightarrow T''(\mathbf{Z}) \in \Sigma'. \tag{64}$$

The compatibility applies to the low-rank and bisparsity structures in two different ways: firstly, $\Sigma^{[r]}$ is compatible with the tail projection for $\Sigma_{(s)}$ given in Proposition 4, by virtue of the fact that a matrix $\mathbf{Z}$ of rank at most r has all its submatrices $\mathbf{Z}_{S\times S}$ of rank at most r, too; secondly, $\Sigma_{(s)}$ is compatible with the exact projection for $\Sigma^{[r]}$, by virtue of the fact that a matrix $\mathbf{Z}$ supported on $S \times S$ has all its singular vectors supported on S, so that $P^{[r]}(\mathbf{Z})$ is supported on $S \times S$, too. Here is the abstract statement valid for arbitrary structures Σ' and Σ''.

Proposition 8. Let T' and T'' be tail projections for Σ' and Σ'' with constants $C_{T'}$ and $C_{T''}$. If Σ' is compatible with T'', then $T''\circ T'$ is a tail projection for $\Sigma' \cap \Sigma''$ with constant $C_{T'}+C_{T''}+C_{T'}C_{T''}$.

Proof. We first remark that the compatibility condition ensures that $T'' \circ T'$ maps into $\Sigma' \cap \Sigma''$. Let $\mathbf{M} \in \mathbb{R}^{n\times n}$ and let $P(\mathbf{M})$ denote its exact projection for $\Sigma' \cap \Sigma''$. The tail condition for T' implies that

$$\|\mathbf{M} - T'(\mathbf{M})\|_F \le C_{T'} \|\mathbf{M} - P(\mathbf{M})\|_F. \tag{65}$$

As a result, we obtain

$$\|T'(\mathbf{M}) - P(\mathbf{M})\|_F \le \|T'(\mathbf{M}) - \mathbf{M}\|_F + \|\mathbf{M} - P(\mathbf{M})\|_F \le (C_{T'} + 1)\|\mathbf{M} - P(\mathbf{M})\|_F. \tag{66}$$

The tail condition for T'' combined with (66) yields

$$\|T'(\mathbf{M}) - T''(T'(\mathbf{M}))\|_F \le C_{T''}\|T'(\mathbf{M}) - P(\mathbf{M})\|_F \le C_{T''}(C_{T'} + 1)\|\mathbf{M} - P(\mathbf{M})\|_F. \tag{67}$$

Using (65) and (67), we derive that

$$\begin{aligned} \|\mathbf{M} - T''(T'(\mathbf{M}))\|_F &\le \|\mathbf{M} - T'(\mathbf{M})\|_F + \|T'(\mathbf{M}) - T''(T'(\mathbf{M}))\|_F \\ &\le (C_{T'} + C_{T''}(C_{T'} + 1))\|\mathbf{M} - P(\mathbf{M})\|_F, \end{aligned} \tag{68}$$

which proves that $T'' \circ T'$ is a tail projection for $\Sigma' \cap \Sigma''$ with the desired constant. □

Head projections. The literature on the sparse principal component analysis problem informs us that finding a head projection for $\Sigma_{(s)}^{[r]}$ is still an NP-hard problem [13, Theorem 2]. In our setting, though, there is no harm in relaxing the head projection to map into $\Sigma_{(s')}^{[r']}$ with $r' = Cr$ and $s' = Cs$, $C \ge 1$. In this regard, Question 2 asks if one can actually compute a head projection for $\Sigma_{(s)}^{[r]}$ into $\Sigma_{(s')}^{[r']}$. We do not have a definite answer for it, but we highlight an observation featuring a nonabsolute constant c_H, based on what was done for the bisparsity structure.

Proposition 9. For $r \leq s$, there is a practical algorithm that yields a head projection for $\Sigma_{(s)}^{[r]}$ with constant $c_H = \sqrt{r}/s$.

Proof. For a symmetric matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$, we consider the index set $S_\star$ of size s constructed in Proposition 6. For any index set S of size s, we have

$$\|\mathbf{M}_{S_\star \times S_\star}\|_F^2 \geq \frac{1}{s}\|\mathbf{M}_{S \times S}\|_F^2. \tag{69}$$

Then, by noticing that the average of the r largest squared singular values of $\mathbf{M}_{S_\star \times S_\star}$ is larger than the average of all the squared singular values of $\mathbf{M}_{S_\star \times S_\star}$, we derive

$$\|P^{[r]}(\mathbf{M}_{S_\star \times S_\star})\|_F^2 \geq \frac{r}{s}\|\mathbf{M}_{S_\star \times S_\star}\|_F^2 \geq \frac{r}{s^2}\|\mathbf{M}_{S \times S}\|_F^2 \geq \frac{r}{s^2}\|P^{[r]}(\mathbf{M}_{S \times S})\|_F^2. \tag{70}$$

The desired result is now proved. □

A similar argument, based on Proposition 5 instead of Proposition 6, would yields a head projection for $\Sigma_{(s)}^{[r]}$ into $\Sigma_{(2s)}^{[r]}$ with constant $c_H = \sqrt{r/n}$.

6 Sample Complexity with Rank-One Measurements

The specific (rank-one) measurements (5) do not result in a measurement map $\boldsymbol{\mathcal{A}} : \mathbb{R}^{n \times n} \to \mathbb{R}^m$ obeying the standard restricted isometry property (10). However, it will satisfy the following version featuring the ℓ_1-norm as an inner norm. This was established in [2] when considering the low-rank structure alone. The proof sketch is deferred to the appendix. Note that the rank-one measurements (5) also satisfy a version of the null space property ensuring recovery via nuclear norm minimization, see [11, 12].

Theorem 10. Suppose $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^m$ are independent vectors with independent $\mathcal{N}(0, 1/m)$ entries. Then, with failure probability at most $2\exp(-cm)$,

$$\alpha\|\mathbf{Z}\|_F \leq \left\|(\mathbf{a}_i^\top \mathbf{Z}\mathbf{a}_i)_{i=1}^m\right\|_1 \leq \beta\|\mathbf{Z}\|_F \qquad \text{for all } \mathbf{Z} \in \Sigma_{(s)}^{[r]}, \tag{71}$$

provided $m \geq Crs\ln(en/s)$. The constants $\beta \geq \alpha > 0$ are absolute.

The restricted isometry property (71) already guarantees that the specific-sample complexity — the theoretical one — is $m \asymp rs\ln(en/s)$, as expected. Indeed, given $\mathbf{y} = \boldsymbol{\mathcal{A}}(\mathbf{X}) + \mathbf{e}$ for some $\mathbf{X} \in \Sigma_{(s)}^{[r]}$, consider the unpractical recovery scheme

$$\Delta(\mathbf{y}) = \underset{\mathbf{Z} \in \Sigma_{(s)}^{[r]}}{\operatorname{argmin}} \|\mathbf{y} - \boldsymbol{\mathcal{A}}(\mathbf{Z})\|_1. \tag{72}$$

In a similar spirit to (12)-(13), we can derive that

$$\|\mathbf{X} - \Delta(\boldsymbol{\mathcal{A}}(\mathbf{X}) + \mathbf{e})\|_F \leq \frac{2}{\alpha(1-\delta)}\|\mathbf{e}\|_1. \tag{73}$$

For a practical algorithm scheme, we have in mind an algorithm belonging to the iterative hard thresholding family. Namely, we can think of constructing a sequence $(\mathbf{X}_k)$ of matrices in $\Sigma_{(s')}^{[r']}$ by the recursion[2]

$$\mathbf{X}_{k+1} = T\left[\mathbf{X}_k + \nu_k H(\boldsymbol{\mathcal{A}}^* \mathrm{sgn}(\mathbf{y} - \boldsymbol{\mathcal{A}}\mathbf{X}_k))\right], \qquad \nu_k = \frac{\|\mathbf{y} - \boldsymbol{\mathcal{A}}\mathbf{X}_k\|_1}{\beta^2}. \tag{74}$$

Here, the operators $T : \mathbb{R}^{n\times n} \to \Sigma_{(s')}^{[r']}$ and $H : \mathbb{R}^{n\times n} \to \Sigma_{(s'')}^{[r'']}$, depending on parameters C' and C'' via $r' = C'r$, $s' = C's$ and $r'' = C''r$, $s'' = C''s$, may be tail and head projections. It could also be useful to require the operator T to satisfy the property[3] that, for all $\mathbf{X} \in \Sigma_{(s)}^{[r]}$ and all $\mathbf{Z} \in \mathbb{R}^{n\times n}$,

$$\|\mathbf{X} - T(\mathbf{Z})\|_F \leq \eta(C)\|\mathbf{X} - \mathbf{Z}\|_F \qquad \text{with} \quad \eta(C') \underset{C'\to\infty}{\longrightarrow} 1. \tag{75}$$

With $T = P_{(s')}^{[r']}$, this inequality seems rather intuitive, but it needs to be formalized — keep in mind, however, that $P_{(s')}^{[r']}$ is not accessible. When considering the low-rank structure alone, such an inequality has been established and exploited in [7] to prove that an iterative hard thresholding algorithm of the type (74) presents the same recovery guarantees as nuclear norm minimization for recovery from measurements of type (5). The type of inequality (75) was first put forward for the sparse vector case in [19] and it has been exploited in [5] to propose and analyze an iterative hard thresholding algorithm designed for the case when the standard restricted isometry property fails.

There is an additional property that we could require about the operator T. Namely, given $\mathbf{M} \in \mathbb{R}^{n\times n}$, if $T(\mathbf{M})$ is supported on $S \times S$, then

$$T(\mathbf{M}) = T(\mathbf{M}_{S'\times S'}) \qquad \text{whenever} \quad S' \supseteq S. \tag{76}$$

This property is true (see Appendix) for $T = P_{(s')}^{[r']}$, which again is inaccessible.

7 Appendix: Proofs of Auxiliary Results

This section collects the detailed arguments for some facts that have been stated but not proved in the narrative.

[2]It is 'natural' to include the sgn operator in order to exploit the restricted isometry property with ℓ_1 inner norm.

[3]The inequality of (75) implies that T is a tail projection with $C_T = 1 + \eta(C')$, since

$$\|\mathbf{M}-T(\mathbf{M})\|_F \leq \|\mathbf{M}-P_{(s)}^{[r]}(\mathbf{M})\|_F + \|P_{(s)}^{[r]}(\mathbf{M})-T(\mathbf{M})\|_F \leq \|\mathbf{M}-P_{(s)}^{[r]}(\mathbf{M})\|_F + \eta(C')\|P_{(s)}^{[r]}(\mathbf{M})-\mathbf{M}\|_F = C_T\|\mathbf{M}-P_{(s)}^{[r]}(\mathbf{M})\|_F.$$

Restricted isometry properties. First, let us concentrate on Theorem 1 and briefly justify that Gaussian measurements of type (4) satisfy the standard restricted isometry property (10). Without going into details, we simply mention that the classical proof consisting of a concentration inequality followed by a covering argument works — the key being to estimate the covering number of the 'ball' of $\Sigma_{(s)}^{[r]}$ essentially as in [3, Lemma 3.1] with the addition of a union bound.

Next, let us concentrate on Theorem 10 and briefly justify that Gaussian rank-one measurements of type (5) satisfy the modified restricted isometry property (71). Again, without going into details, we point out that the proof is in the spirit of [4]: for a fixed $\mathbf{Z} \in \mathbb{R}^{n\times n}$, establish a concentration inequality for $\left\|(\mathbf{a}_i^\top \mathbf{Z}\mathbf{a}_i)_{i=1}^m\right\|_1$ around its expectation $/\!\!/\mathbf{Z}/\!\!/$, prove that this slanted norm is equivalent to the Frobenius norm, and conclude with a covering argument.

Convergence of the idealized iterative hard thresholding. We now establish that the naive (and impractical) iterative hard thresholding algorithm (15) allows for stable and robust recovery of jointly low-rank and bisparse matrices under the standard restricted isometry property. The precise statement appears after the important observation below.

Lemma 11. Suppose that $\boldsymbol{\mathcal{A}} : \mathbb{R}^{n\times n} \to \mathbb{R}^m$ satisfies the restricted isometry property (10) on $\Sigma_{(2s)}^{[2r]}$ with constant $\delta \in (0,1)$. Then, for all $\mathbf{Z}, \mathbf{Z}' \in \Sigma_{(s)}^{[r]}$, one has

$$\left|\langle \mathbf{Z}, (\boldsymbol{\mathcal{A}}^*\boldsymbol{\mathcal{A}} - \mathbf{I})(\mathbf{Z}')\rangle\right| \le \delta \|\mathbf{Z}\|_F \|\mathbf{Z}'\|_F. \tag{77}$$

Proof. Assuming without loss of generality that $\|\mathbf{Z}\|_F = \|\mathbf{Z}'\|_F = 1$, we use in particular the parallelogram identity to write

$$\begin{aligned} \left|\langle \mathbf{Z}, (\boldsymbol{\mathcal{A}}^*\boldsymbol{\mathcal{A}} - \mathbf{I})(\mathbf{Z}')\rangle\right| &= \left|\langle \boldsymbol{\mathcal{A}}(\mathbf{Z}), \boldsymbol{\mathcal{A}}(\mathbf{Z}')\rangle - \langle \mathbf{Z}, \mathbf{Z}'\rangle\right| \\ &= \left|\frac{1}{4}\left(\|\boldsymbol{\mathcal{A}}(\mathbf{Z}+\mathbf{Z}')\|_2^2 - \|\boldsymbol{\mathcal{A}}(\mathbf{Z}-\mathbf{Z}')\|_2^2\right) - \frac{1}{4}\left(\|\mathbf{Z}+\mathbf{Z}'\|_F^2 - \|\mathbf{Z}-\mathbf{Z}'\|_F^2\right)\right| \\ &\le \frac{1}{4}\left|\|\boldsymbol{\mathcal{A}}(\mathbf{Z}+\mathbf{Z}')\|_2^2 - \|\mathbf{Z}+\mathbf{Z}'\|_F^2\right| + \frac{1}{4}\left|\|\boldsymbol{\mathcal{A}}(\mathbf{Z}-\mathbf{Z}')\|_2^2 - \|\mathbf{Z}-\mathbf{Z}'\|_F^2\right| \\ &\le \frac{1}{4}\delta\|\mathbf{Z}+\mathbf{Z}'\|_F^2 + \frac{1}{4}\delta\|\mathbf{Z}-\mathbf{Z}'\|_F^2 = \frac{1}{4}\delta\left(2\|\mathbf{Z}\|_F^2 + 2\|\mathbf{Z}'\|_F^2\right) = \delta, \end{aligned} \tag{78}$$

which is the required result. □

Theorem 12. If the restricted isometry property (10) holds on $\Sigma_{(4s)}^{[4r]}$ with constant $\delta \in (0, 1/2)$, then any $\mathbf{X} \in \Sigma_{(s)}^{[r]}$ is approximated from $\mathbf{y} = \boldsymbol{\mathcal{A}}\mathbf{X} + \mathbf{e} \in \mathbb{R}^m$ as a cluster point $\mathbf{X}_\infty$ of the sequence $(\mathbf{X}_k)_{k\ge 0}$ defined by

$$\mathbf{X}_{k+1} = P_{(s)}^{[r]}\left(\mathbf{X}_k + \boldsymbol{\mathcal{A}}^*(\mathbf{y} - \boldsymbol{\mathcal{A}}\mathbf{X}_k)\right) \tag{79}$$

with error

$$\|\mathbf{X} - \mathbf{X}_\infty\|_F \le C\|\mathbf{e}\|_2. \tag{80}$$

Proof. It is enough to prove that, for all $k \geq 0$,

(81) $$\|\mathbf{X} - \mathbf{X}_{k+1}\|_F \leq \rho \|\mathbf{X} - \mathbf{X}_k\|_F + \tau \|\mathbf{e}\|_2, \qquad \text{with} \quad \rho := 2\delta < 1 \text{ and } \tau > 0.$$

To start, notice that $\mathbf{X}_{k+1}$ better approximates $\mathbf{X}_k + \boldsymbol{\mathcal{A}}^*(\mathbf{y} - \boldsymbol{\mathcal{A}}\mathbf{X}_k) = \mathbf{X}_k + \boldsymbol{\mathcal{A}}^*\boldsymbol{\mathcal{A}}(\mathbf{X} - \mathbf{X}_k) + \boldsymbol{\mathcal{A}}^*\mathbf{e}$ as an element from $\Sigma_{(s)}^{[r]}$ than $\mathbf{X}$ does, so that

(82) $$\|\mathbf{X}_k + \boldsymbol{\mathcal{A}}^*\boldsymbol{\mathcal{A}}(\mathbf{X} - \mathbf{X}_k) + \boldsymbol{\mathcal{A}}^*\mathbf{e} - \mathbf{X}_{k+1}\|_F^2 \leq \|\mathbf{X}_k + \boldsymbol{\mathcal{A}}^*\boldsymbol{\mathcal{A}}(\mathbf{X} - \mathbf{X}_k) + \boldsymbol{\mathcal{A}}^*\mathbf{e} - \mathbf{X}\|_F^2.$$

Introducing $\mathbf{X}$ in the left-hand side, expanding the squares, and simplifying leads to

(83) $$\|\mathbf{X} - \mathbf{X}_{k+1}\|_F^2 \leq -2\langle \mathbf{X} - \mathbf{X}_{k+1}, (\boldsymbol{\mathcal{A}}^*\boldsymbol{\mathcal{A}} - \mathbf{I})(\mathbf{X} - \mathbf{X}_k) + \boldsymbol{\mathcal{A}}^*\mathbf{e}\rangle.$$

Thanks to Lemma 11, we have

(84) $$|\langle \mathbf{X} - \mathbf{X}_{k+1}, (\boldsymbol{\mathcal{A}}^*\boldsymbol{\mathcal{A}} - \mathbf{I})(\mathbf{X} - \mathbf{X}_k)\rangle| \leq 2\delta \|\mathbf{X} - \mathbf{X}_{k+1}\|_F \|\mathbf{X} - \mathbf{X}_k\|_F,$$

while the restricted isometry property (10) also guarantees that

(85) $$|\langle \mathbf{X} - \mathbf{X}_{k+1}, \boldsymbol{\mathcal{A}}^*\mathbf{e}\rangle| = |\langle \boldsymbol{\mathcal{A}}(\mathbf{X} - \mathbf{X}_{k+1}), \mathbf{e}\rangle| \leq \|\boldsymbol{\mathcal{A}}(\mathbf{X} - \mathbf{X}_{k+1})\|_2 \|\mathbf{e}\|_2 \leq \sqrt{1+\delta}\|\mathbf{X} - \mathbf{X}_{k+1}\|_F \|\mathbf{e}\|_2.$$

Therefore, using (84) and (85) in (83), we obtain

(86) $$\|\mathbf{X} - \mathbf{X}_{k+1}\|_F^2 \leq 2\delta \|\mathbf{X} - \mathbf{X}_{k+1}\|_F \|\mathbf{X} - \mathbf{X}_k\|_F + \sqrt{1+\delta}\|\mathbf{X} - \mathbf{X}_{k+1}\|_F \|\mathbf{e}\|_2,$$

which clearly implies the required estimates (81) with $\tau = \sqrt{1+\delta}$ and (80) with $C = \tau/(1-\rho)$. ☐

The exact projection for $\Sigma_{(s)}^{[r]}$. Here, we prove the statement (25) about the form of $P_{(s)}^{[r]}$ before justifying that property (76) holds for $T = P_{(s)}^{[r]}$.

Proposition 13. For $\mathbf{M} \in \mathbb{R}^{n \times n}$, the projection $P_{(s)}^{[r]}(\mathbf{M})$ of $\mathbf{M}$ onto $\Sigma_{(s)}^{[r]}$ has the form $P^{[r]}(\mathbf{M}_{S_\star \times S_\star})$, where $S_\star$ maximizes $\|P^{[r]}(\mathbf{M}_{S \times S})\|_F$ over all index sets S of size s.

Proof. Let us remark that, for any index set T,

(87) $$\begin{aligned} \|\mathbf{M} - P^{[r]}(\mathbf{M}_{T \times T})\|_F^2 &= \|\mathbf{M}_{\overline{T \times T}} + \mathbf{M}_{T \times T} - P^{[r]}(\mathbf{M}_{T \times T})\|_F^2 \\ &= \|\mathbf{M}_{\overline{T \times T}}\|_F^2 + \|\mathbf{M}_{T \times T} - P^{[r]}(\mathbf{M}_{T \times T})\|_F^2 \\ &= \|\mathbf{M}_{\overline{T \times T}}\|_F^2 + \|\mathbf{M}_{T \times T}\|_F^2 - \|P^{[r]}(\mathbf{M}_{T \times T})\|_F^2 \\ &= \|\mathbf{M}\|_F^2 - \|P^{[r]}(\mathbf{M}_{T \times T})\|_F^2. \end{aligned}$$

Now let $\mathbf{Z} \in \Sigma_{(s)}^{[r]}$ and consider an index set S of size s such that $\mathbf{Z}$ is supported on $S \times S$. The defining property of $S_\star$, together with (87), implies that

(88) $$\begin{aligned} \|\mathbf{M} - P^{[r]}(\mathbf{M}_{S_\star \times S_\star})\|_F^2 \leq \|\mathbf{M}\|_F^2 - \|P^{[r]}(\mathbf{M}_{S \times S})\|_F^2 &= \|\mathbf{M}_{\overline{S \times S}}\|_F^2 + \|\mathbf{M}_{S \times S} - P^{[r]}(\mathbf{M}_{S \times S})\|_F^2 \\ \leq \|\mathbf{M}_{\overline{S \times S}}\|_F^2 + \|\mathbf{M}_{S \times S} - \mathbf{Z}\|_F^2 &= \|\mathbf{M} - \mathbf{Z}\|_F^2, \end{aligned}$$

where we have taken into account the facts that $P^{[r]}(\mathbf{M}_{S \times S})$ is the best r-rank approximation to $\mathbf{M}_{S \times S}$ and that $\mathbf{M}_{\overline{S \times S}}$ and $\mathbf{M}_{S \times S} - \mathbf{Z}$ are disjointly supported. Thus, we have proved that $\|\mathbf{M} - P^{[r]}(\mathbf{M}_{S_\star \times S_\star})\|_F \leq \|\mathbf{M} - \mathbf{Z}\|_F$ for all $\mathbf{Z} \in \Sigma_{(s)}^{[r]}$, which is the desired result. ☐

Proposition 14. For $\mathbf{M} \in \mathbb{R}^{n \times n}$, considering an index set $S_\star$ of size s with $P_{(s)}^{[r]}(\mathbf{M}) = P^{[r]}(\mathbf{M}_{S_\star \times S_\star})$, one has

$$P_{(s)}^{[r]}(\mathbf{M}) = P_{(s)}^{[r]}(\mathbf{M}_{S' \times S'}) \qquad \text{whenever } S' \supseteq S_\star. \tag{89}$$

Proof. According to Proposition 13, it is enough to verify that, for any index set S of size s,

$$\left\| P^{[r]}((\mathbf{M}_{S' \times S'})_{S_\star \times S_\star}) \right\|_F \geq \left\| P^{[r]}((\mathbf{M}_{S' \times S'})_{S \times S}) \right\|_F. \tag{90}$$

But this is true because $(\mathbf{M}_{S' \times S'})_{S_\star \times S_\star} = \mathbf{M}_{S_\star \times S_\star}$ and $(\mathbf{M}_{S' \times S'})_{S \times S} = (\mathbf{M}_{S \times S})_{S' \times S'}$, so that

$$\left\| P^{[r]}((\mathbf{M}_{S' \times S'})_{S \times S}) \right\|_F \leq \| P^{[r]}(\mathbf{M}_{S \times S}) \|_F \leq \| P^{[r]}(\mathbf{M}_{S_\star \times S_\star}) \|_F, \tag{91}$$

where the last inequality follows from the defining property of $S_\star$. □

References

[1] T. Blumensath. *Sampling and reconstructing signals from a union of linear subspaces.* IEEE Transactions on Information Theory 57.7 (2011): 4660–4671.

[2] T. Cai and A. Zhang. *ROP: Matrix recovery via rank-one projections.* The Annals of Statistics 43.1 (2015): 102–138.

[3] 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 57.4 (2011): 2342–2359.

[4] S. Foucart and M.-J. Lai. *Sparse recovery with pre-Gaussian random matrices.* Studia Math 200.1 (2010): 91–102.

[5] S. Foucart and G. Lecué. *An IHT algorithm for sparse recovery from sub-exponential measurements.* IEEE Signal Processing Letters 24.9 (2017): 1280–1283.

[6] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser, 2013.

[7] S. Foucart and S. Subramanian. *Iterative hard thresholding for low-rank recovery from rank-one projections.* arXiv preprint arXiv:1810.11749 (2018).

[8] M. Golbabaee and M. E. Davies. *Inexact gradient projection and fast data driven compressed sensing.* IEEE Transactions on Information Theory (2018).

[9] C. Hegde, P. Indyk, and L. Schmidt. *Approximation algorithms for model-based compressive sensing.* IEEE Transactions on Information Theory 61.9 (2015): 5129–5147.

[10] M. Iwen, A. Viswanathan, and Y. Wang. *Robust sparse phase retrieval made easy.* Applied and Computational Harmonic Analysis 42.1 (2017): 135–142.

[11] M. Kabanava, R. Kueng, H. Rauhut, and U. Terstiege. *Stable low-rank matrix recovery via null space properties.* Information and Inference 5.4 (2016): 405–441.

[12] R. Kueng, H. Rauhut, and U. Terstiege. *Low rank matrix recovery from rank one measurements.* Applied and Computational Harmonic Analysis 42.1 (2017): 88–116.

[13] M. Magdon-Ismail, *NP-hardness and inapproximability of sparse PCA.* Information Processing Letters 126 (2017): 35–38.

[14] P. Manurangsi. *Almost-polynomial ratio ETH-hardness of approximating densest k-subgraph.* STOC'17–Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, ACM, pp. 954–961, 2017.

[15] S. Oymak, A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi. *Simultaneously structured models with application to sparse and low-rank matrices.* IEEE Transactions on Information Theory 61.5 (2015): 2886–2908.

[16] H. Rauhut, R. Schneider, and Z. Stojanac. *Low rank tensor recovery via iterative hard thresholding.* Linear Algebra and Its Applications 523 (2018): 220–262.

[17] I. Roth, M. Kliesch, G. Wunder, and J. Eisert. *Reliable recovery of hierarchically sparse signals and application in machine-type communications.* arXiv preprint arXiv:1612.07806 (2016).

[18] I. Roth, A. Flinth, R. Kueng, J. Eisert, and G. Wunder. *Hierarchical restricted isometry property for Kronecker product measurements.* arXiv preprint arXiv:1801.10433 (2018).

[19] J. Shen and P. Li. *A tight bound of hard thresholding.* arXiv preprint arXiv:1605.01656 (2016).