One-Bit Sensing of Low-Rank and Bisparse Matrices

Simon Foucart[*] and Laurent Jacques[†]
[*]Texas A&M University, USA. [†]UCLouvain, Belgium.

Abstract—This note studies the worst-case recovery error of low-rank and bisparse matrices as a function of the number of one-bit measurements used to acquire them. First, by way of the concept of consistency width, precise estimates are given on how fast the recovery error can in theory decay. Next, an idealized recovery method is proved to reach the fourth-root of the optimal decay rate for Gaussian sensing schemes. This idealized method being impractical, an implementable recovery algorithm is finally proposed in the context of factorized Gaussian sensing schemes. It is shown to provide a recovery error decaying as the sixth-root of the optimal rate.

I. INTRODUCTION

We are interested in matrices that are simultaneously low-rank and bisparse, *i.e.*, anticipating over some notation to be introduced momentarily, matrices in the set

$$\Sigma_{(s)}^{[r]} := \big\{ \boldsymbol{X} \in \mathbb{R}^{n \times n} : \mathrm{rank}(\boldsymbol{X}) \le r, \exists S, T \subseteq [\![n]\!],$$
$$|S| = |T| = s, \boldsymbol{X}_{\overline{S} \times [\![n]\!]} = \boldsymbol{0}, \boldsymbol{X}_{[\![n]\!] \times \overline{T}} = \boldsymbol{0} \big\}.$$

Determining the minimal number of linear measurements allowing for uniform stable recovery of matrices in $\Sigma_{(s)}^{[r]}$ is a difficult problem, *e.g.* [1] contains negative results for optimization-based recovery, while [2] showed, with some caveats, that $\Theta(rs \ln(n/s))$ measurement suffice. Here, we consider a similar problem for quantized linear measurements, hoping for more favorable conclusions in this case. Precisely, suppose that matrices $\boldsymbol{X} \in \Sigma_{(s)}^{[r]}$ are acquired via the one-bit measurements

$$y_i = \mathrm{sgn}(\langle \boldsymbol{A}_i, \boldsymbol{X} \rangle_F) = \mathrm{sgn}(\mathrm{tr}(\boldsymbol{A}_i^\top \boldsymbol{X})), \quad i \in [\![m]\!]. \quad (1)$$

We ask: how large should m be to make it possible to derive, from $\boldsymbol{y} \in \{\pm 1\}^m$, an approximant $\boldsymbol{X}'$ such that $\|\boldsymbol{X} - \boldsymbol{X}'\|_F \le \varepsilon$, or, turning things around, how fast can $\|\boldsymbol{X} - \boldsymbol{X}'\|_F$ decay when m increases? Phrased in terms of *consistency width*, Section II addresses this question with lower and upper estimates that almost match. A (weaker) upper estimate also appears in Section III, which introduces an idealized recovery scheme not enforcing consistency of the approximant. This idealized scheme being impractical (probably NP-hard), we exhibit in Section IV a multistep recovery scheme working with specific sensing schemes.

Notation: Given $\boldsymbol{M} \in \mathbb{R}^{n \times n}$, $\boldsymbol{M}_{\Omega_1 \times \Omega_2}$ represents the submatrix of $\boldsymbol{M}$ indexed by $\Omega_1 \times \Omega_2 \subseteq [\![n]\!] \times [\![n]\!]$, with $[\![n]\!] := \{1, \ldots, n\}$. The set $\overline{\Omega} \subseteq [\![n]\!]$ stands for $[\![n]\!] \setminus \Omega$. The Frobenius norm is denoted by $\|\cdot\|_F$ and the Frobenius inner product by $\langle \cdot, \cdot \rangle_F$. Given index sets $S, T \subseteq [\![n]\!]$, the set $\Sigma_{S \times T}^{[r]}$ contains all rank-r matrices supported on $S \times T$. A tilde over a set $\mathcal{K} \subseteq \mathbb{R}^{n \times n}$, *e.g.* $\widetilde{\Sigma}_{(s)}^{[r]}$ or $\widetilde{\Sigma}_{S \times T}^{[r]}$, means that this set is intersected with the sphere $\mathbb{S}_F^{n \times n}$ of $\mathbb{R}^{n \times n}$ with respect to the Frobenius norm. The notation for the unit ball is $\mathbb{B}_F^{n \times n}$.

We often rewrite (1) in more compact form by defining the map $\boldsymbol{\mathcal{A}} : \boldsymbol{M} \in \mathbb{R}^{n \times n} \mapsto (\langle \boldsymbol{A}_i, \boldsymbol{M} \rangle_F)_{i=1}^m \in \mathbb{R}^m$, so that (1) also reads $\boldsymbol{y} = \mathrm{sgn}\, \boldsymbol{\mathcal{A}}(\boldsymbol{X})$ with $\mathrm{sgn}\, \boldsymbol{\mathcal{A}} := \mathrm{sgn} \circ \boldsymbol{\mathcal{A}}$. The sign operator sgn acts entrywise when applied to a vector. We say that $\boldsymbol{\mathcal{A}}$ is a *Gaussian random* map when the matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$ are independent $n \times n$ Gaussian random matrices, *i.e.*, $(\boldsymbol{A}_i)_{k,l} \sim_{\text{i.i.d.}} \mathcal{N}(0,1)$. The adjoint of $\boldsymbol{\mathcal{A}}$ is the map $\boldsymbol{\mathcal{A}}^* : \boldsymbol{v} \in \mathbb{R}^m \to \sum_i v_i \boldsymbol{A}_i \in \mathbb{R}^{n \times n}$. All absolute constants are denoted by $C, C', c, c', \ldots$ and their values can change from line to line.

II. CONSISTENCY WIDTH

The mth *consistency width* of a set $\mathcal{K} \subseteq \mathbb{S}_F^{n \times n}$ is defined as

$$\mathrm{CW}^m(\mathcal{K}) := \inf_{\boldsymbol{\mathcal{A}}: \mathbb{R}^{n \times n} \to \mathbb{R}^m} \sup\{\|\boldsymbol{X} - \boldsymbol{X}'\|_F : \boldsymbol{X}, \boldsymbol{X}' \in \mathcal{K},$$
$$\mathrm{sgn}\, \boldsymbol{\mathcal{A}}(\boldsymbol{X}) = \mathrm{sgn}\, \boldsymbol{\mathcal{A}}(\boldsymbol{X}')\}.$$

Essentially, $\mathrm{CW}^m(\mathcal{K})$ provides the best (with respect to $\boldsymbol{\mathcal{A}}$) worst-case error of any *consistent* recovery scheme, *i.e.*, one that outputs $\boldsymbol{X}' \in \mathcal{K}$ with the same one-bit measurements as the sensed matrix $\boldsymbol{X} \in \mathcal{K}$, see [3]. We shall show that

$$c \frac{rs}{m} \le \mathrm{CW}^m\big(\widetilde{\Sigma}_{(s)}^{[r]}\big) \le C \frac{rs}{m} \ln\left(\frac{nm}{rs}\right). \quad (2)$$

A. Lower estimate

Let us first establish the lower bound in (2). Given $\mathcal{K} \subseteq \mathbb{S}_F^{n \times n}$, we define the smallest achievable worst-case error of any combination of one-bit maps $\mathrm{sgn}\, \boldsymbol{\mathcal{A}}$ with recovery schemes $\boldsymbol{\Delta} : \{\pm 1\}^m \to \mathcal{K}$ by

$$\mathcal{E}^m(\mathcal{K}) := \inf_{(\boldsymbol{\mathcal{A}}, \boldsymbol{\Delta})} \sup_{\boldsymbol{X} \in \mathcal{K}} \|\boldsymbol{X} - \boldsymbol{\Delta}\big(\mathrm{sgn}\, \boldsymbol{\mathcal{A}}(\boldsymbol{X})\big)\|_F.$$

Note that $\boldsymbol{\Delta}$ is not required to output a consistent approximant. However, the smallest achievable worst-case errors using consistent or unrestricted schemes are in fact equivalent, as emphasized below.

Proposition 1. *For any $\mathcal{K} \subseteq \mathbb{S}_F^{n \times n}$, one has*

$$\mathcal{E}^m(\mathcal{K}) \le \mathrm{CW}^m(\mathcal{K}) \le 2\mathcal{E}^m(\mathcal{K}). \quad (3)$$

Proof. Given $(\boldsymbol{\mathcal{A}}, \boldsymbol{\Delta})$ and $\boldsymbol{X}, \boldsymbol{X}' \in \mathcal{K}$ sharing the same one-bit measurements, in view of $\boldsymbol{\Delta}(\mathrm{sgn}\, \boldsymbol{\mathcal{A}}(\boldsymbol{X})) = \boldsymbol{\Delta}(\mathrm{sgn}\, \boldsymbol{\mathcal{A}}(\boldsymbol{X}'))$, there holds $\|\boldsymbol{X} - \boldsymbol{X}'\|_F \le \|\boldsymbol{X} - \boldsymbol{\Delta}(\mathrm{sgn}\, \boldsymbol{\mathcal{A}}(\boldsymbol{X}))\|_F + \|\boldsymbol{X}' - \boldsymbol{\Delta}(\mathrm{sgn}\, \boldsymbol{\mathcal{A}}(\boldsymbol{X}'))\|_F$, which yields the rightmost inequality in (3).

Now, given $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{X} \in \mathcal{K}$, we choose $\boldsymbol{\Delta}(\mathrm{sgn}\, \boldsymbol{\mathcal{A}}(\boldsymbol{X}))$ as an arbitrary matrix $\boldsymbol{X}' \in \mathcal{K}$ consistent with $\boldsymbol{X}$. The leftmost inequality in (3) follows easily. $\square$

Proposition 1 implies in particular that the lower bound in (2) can be derived from the following lower estimation of $\mathcal{E}^m(\widetilde{\Sigma}_{(s)}^{[r]})$.

Theorem 1. *For an error level $\varepsilon > 0$, one has*

$$\mathcal{E}^m\big(\widetilde{\Sigma}_{(s)}^{[r]}\big) \le \varepsilon \implies m \ge C\varepsilon^{-1} rs. \quad (4)$$

Equivalently, this means that $\mathcal{E}^m\big(\widetilde{\Sigma}_{(s)}^{[r]}\big) \ge C\frac{rs}{m}$.

Proof. Suppose that $\mathcal{E}^m\big(\widetilde{\Sigma}_{(s)}^{[r]}\big) \le \varepsilon$. We fix subsets S, T of $[\![n]\!]$ with $|S| = |T| = s$, a matrix $\boldsymbol{M} \in \mathbb{R}^{n \times r}$ such that $\boldsymbol{M}_{\overline{S} \times [\![r]\!]} = \boldsymbol{0}$, and consider the rs-dimensional subspace $\mathcal{S}$ of $\mathbb{R}^{n \times n}$ defined by $\mathcal{S} := \{\boldsymbol{M}\boldsymbol{N} : \boldsymbol{N} \in \mathbb{R}^{r \times n}, \boldsymbol{N}_{[\![r]\!] \times \overline{T}} = \boldsymbol{0}\}$. Note that $\mathcal{S} \subseteq \Sigma_{(s)}^{[r]}$, hence $\mathcal{E}^m(\widetilde{\mathcal{S}}) \le \mathcal{E}(\widetilde{\Sigma}_{(s)}^{[r]}) \le \varepsilon$. Therefore, for any $\varepsilon' > \varepsilon$, say $\varepsilon' = 2\varepsilon$, we can find a pair $(\boldsymbol{\mathcal{A}}, \boldsymbol{\Delta})$ such that $\|\boldsymbol{X} - \boldsymbol{\Delta}\big(\mathrm{sgn}\, \boldsymbol{\mathcal{A}}(\boldsymbol{X})\big)\|_F \le \varepsilon'$ for all $\boldsymbol{X} \in \widetilde{\mathcal{S}}$. This provides a 2ε-covering of $\widetilde{\mathcal{S}}$ with $|\mathrm{sgn}\, \boldsymbol{\mathcal{A}}(\mathcal{S})|$ elements. Recall, however, that the latter is much smaller than the obvious bound 2^m. Indeed, [4, Lemma 1] guarantees that the cardinality of the image under $\mathrm{sgn}\, \boldsymbol{\mathcal{A}}$ of a d-dimensional subspace $\mathcal{S}'$ of $\mathbb{R}^{n \times n}$ satisfies $|\mathrm{sgn}\, \boldsymbol{\mathcal{A}}(\mathcal{S}')| \le (\frac{2m}{d})^d$. In our situation, we thus obtain a 2ε-covering of $\widetilde{\mathcal{S}}$ with log-cardinality at most $rs \ln(2m/(rs))$. But, $\widetilde{\mathcal{S}}$ being isomorphic to the sphere $\mathbb{S}_F^{r \times s}$, the log-cardinality of such a covering must be at least $rs \ln(C/\varepsilon)$. Putting these two facts together gives $m \ge C\varepsilon^{-1} rs$, so that (4) is proved. $\square$

B. Upper estimate

Let us now establish the upper bound in (2). We first show that, with high probability on a Gaussian random map $\boldsymbol{\mathcal{A}}$, all pairs of matrices in a set $\mathcal{K} \subseteq \mathbb{B}_F^{n \times n}$ that are consistent under $\operatorname{sgn} \boldsymbol{\mathcal{A}}$ must be at most ε apart provided m is large compared to ε and to the intrinsic set complexity. This complexity is measured by the *Kolmogorov entropy* of $\mathcal{K}$, *i.e.*,

$$\mathcal{H}(\mathcal{K}, \eta) := \ln \inf\{ |\mathcal{G}| : \ \mathcal{G} \subseteq \mathcal{K} \subseteq \mathcal{G} + \eta\mathbb{B}_F^{n \times n}\},$$

where the addition is the Minkowski sum between sets. Thus, the smallest number of translated Frobenius balls with radius η that cover $\mathcal{K}$ is $\exp \mathcal{H}(\mathcal{K}, \eta)$.

Theorem 2. *Let* $\boldsymbol{\mathcal{A}} : \mathbb{R}^{n \times n} \to \mathbb{R}^m$ *be a Gaussian random map. Given a subset* $\mathcal{K}$ *of* $\mathbb{S}_F^{n \times n}$ *and* $\varepsilon > 0$, *if*

$$m \geq \tfrac{C}{\varepsilon}\mathcal{H}(\mathcal{K}, c\varepsilon), \tag{5}$$

then it occurs with probability at least $1 - \exp(-c\varepsilon m)$ *that, for all* $\boldsymbol{X}, \boldsymbol{X}' \in \mathcal{K}$,

$$\operatorname{sgn}\boldsymbol{\mathcal{A}}(\boldsymbol{X}) = \operatorname{sgn}\boldsymbol{\mathcal{A}}(\boldsymbol{X}') \implies \|\boldsymbol{X} - \boldsymbol{X}'\|_F \leq \varepsilon. \tag{6}$$

This result is related, but not strictly equivalent, to Theorem 2.4 from [5]. Under a requirement similar to (5), it states that a Gaussian random map $\boldsymbol{\mathcal{A}}$ defines, with high probability, a *local ε-isometry* of $\mathcal{K}$. In particular, it implies that, for all $\boldsymbol{X}, \boldsymbol{X}' \in \mathcal{K}$ such that $\|\boldsymbol{X} - \boldsymbol{X}'\|_F \geq \varepsilon$, one has

$$|\tfrac{1}{m}\|\operatorname{sgn}\boldsymbol{\mathcal{A}}\boldsymbol{X}, \operatorname{sgn}\boldsymbol{\mathcal{A}}\boldsymbol{X}'\|_H - \tfrac{1}{\pi}\arccos\langle\boldsymbol{X}, \boldsymbol{X}'\rangle| \geq c\varepsilon, \tag{7}$$

with $\|\boldsymbol{s}, \boldsymbol{s}'\|_H := |\{i \in [\![m]\!] : s_i \neq s'_i\}|$ for $\boldsymbol{s}, \boldsymbol{s}' \in \{\pm 1\}^m$. Note that (7) does not yield the contraposition of (6), since it does not enforce that $\operatorname{sgn}\boldsymbol{\mathcal{A}}\boldsymbol{X} \neq \operatorname{sgn}\boldsymbol{\mathcal{A}}\boldsymbol{X}'$.

Proof of Theorem 2. The stated version appears in [6, Theorem 1.5]. For the sake of diversity of the arguments, we follow the idea from [4, Theorem 2] and prove a result that would still imply the upper bound in (2), but does require strengthening the assumption (5) to

$$m \geq \tfrac{C}{\varepsilon}\mathcal{H}(\mathcal{K}, c\tfrac{\varepsilon}{n}). \tag{8}$$

We shall prove that $\mathbb{P}_0 \leq \exp(-c\varepsilon m)$, where

$$\mathbb{P}_0 := \mathbb{P}\big(\exists \boldsymbol{X}, \boldsymbol{X}' \in \mathcal{K} : \|\boldsymbol{X} - \boldsymbol{X}'\|_F > \varepsilon, \ \operatorname{sgn}\boldsymbol{\mathcal{A}}(\boldsymbol{X}) = \operatorname{sgn}\boldsymbol{\mathcal{A}}(\boldsymbol{X}')\big).$$

For a number $\rho \in (0, \varepsilon/2)$ yet to be chosen, we consider a ρ-covering $\{\boldsymbol{X}_1, \cdots, \boldsymbol{X}_K\}$ of $\mathcal{K}$ with $K = \exp \mathcal{H}(\mathcal{K}, \rho)$. Given $\boldsymbol{X}, \boldsymbol{X}' \in \mathcal{K}$ satisfying $\|\boldsymbol{X} - \boldsymbol{X}'\|_F > \varepsilon$ and $\operatorname{sgn}\boldsymbol{\mathcal{A}}(\boldsymbol{X}) = \operatorname{sgn}\boldsymbol{\mathcal{A}}(\boldsymbol{X}')$, we select $k, k' \in [\![K]\!]$ such that $\|\boldsymbol{X} - \boldsymbol{X}_k\|_F \leq \rho$ and $\|\boldsymbol{X}' - \boldsymbol{X}_{k'}\|_F \leq \rho$, implying that $\|\boldsymbol{X}_k - \boldsymbol{X}_{k'}\|_F \geq \varepsilon - 2\rho$ and that $\operatorname{sgn}\boldsymbol{\mathcal{A}}(\boldsymbol{X}_k + \boldsymbol{Z}) = \operatorname{sgn}\boldsymbol{\mathcal{A}}(\boldsymbol{X}_{k'} + \boldsymbol{Z}')$ where $\boldsymbol{Z} := \boldsymbol{X} - \boldsymbol{X}_k$ and $\boldsymbol{Z}' := \boldsymbol{X}' - \boldsymbol{X}_{K'}$ both have norm at most ρ. It follows that

$$\mathbb{P}_0 \leq \mathbb{P}\big(\exists k, k' \in [\![K]\!] : \|\boldsymbol{X}_k - \boldsymbol{X}_{k'}\|_F \geq \varepsilon - 2\rho,$$
$$\exists \boldsymbol{Z}, \boldsymbol{Z}' \in \rho\mathbb{B}_F^{n \times n} : \operatorname{sgn}\boldsymbol{\mathcal{A}}(\boldsymbol{X}_k + \boldsymbol{Z}) = \operatorname{sgn}\boldsymbol{\mathcal{A}}(\boldsymbol{X}_{k'} + \boldsymbol{Z}')\big).$$

With $\Omega := \{(k, k') \in [\![K]\!]^2 : \|\boldsymbol{X}_k - \boldsymbol{X}_{k'}\|_F \geq \varepsilon - 2\rho\}$, a union bound then gives

$$\mathbb{P}_0 \leq \sum_{(k,k') \in \Omega} \mathbb{P}_{k,k'},$$

where the summands are

$$\mathbb{P}_{k,k'} := \mathbb{P}\big(\exists \boldsymbol{Z}, \boldsymbol{Z}' \in \rho\mathbb{B}_F^{n \times n} : \operatorname{sgn}\boldsymbol{\mathcal{A}}(\boldsymbol{X}_k + \boldsymbol{Z}) = \operatorname{sgn}\boldsymbol{\mathcal{A}}(\boldsymbol{X}_{k'} + \boldsymbol{Z}')\big)$$
$$= \mathbb{P}\big(\exists \boldsymbol{Z}, \boldsymbol{Z}' \in \rho\mathbb{B}_F^{n \times n} : \forall i \in [\![m]\!],$$
$$\operatorname{sgn}(\langle\boldsymbol{A}_i, \boldsymbol{X}_k + \boldsymbol{Z}\rangle_F) = \operatorname{sgn}(\langle\boldsymbol{A}_i, \boldsymbol{X}_{k'} + \boldsymbol{Z}'\rangle_F)\big).$$

Let us observe that, for $(k, k') \in \Omega$,

$$\mathbb{P}_{k,k'} \leq \mathbb{P}\big(\forall i \in [\![m]\!], \exists \boldsymbol{Z}, \boldsymbol{Z}' \in \rho\mathbb{B}_F^{n \times n} :$$
$$\operatorname{sgn}(\langle\boldsymbol{A}_i, \boldsymbol{X}_k + \boldsymbol{Z}\rangle_F) = \operatorname{sgn}(\langle\boldsymbol{A}_i, \boldsymbol{X}_{k'} + \boldsymbol{Z}'\rangle_F)\big)$$
$$= \prod_{i=1}^m \mathbb{P}\big(\exists \boldsymbol{Z}, \boldsymbol{Z}' \in \rho\mathbb{B}_F^{n \times n} :$$
$$\operatorname{sgn}(\langle\boldsymbol{A}_i, \boldsymbol{X}_k + \boldsymbol{Z}\rangle_F) = \operatorname{sgn}(\langle\boldsymbol{A}_i, \boldsymbol{X}_{k'} + \boldsymbol{Z}'\rangle_F)\big)$$
$$\leq \prod_{i=1}^m \big(1 - \tfrac{1}{\pi}\arccos(\langle\boldsymbol{X}_k, \boldsymbol{X}_{k'}\rangle_F) + \sqrt{\tfrac{\pi}{2}}n\rho\big),$$

where the last step was an application of [4, Lemma 9]. In view of $\arccos(\langle\boldsymbol{X}_k, \boldsymbol{X}_{k'}\rangle_F) \geq \|\boldsymbol{X}_k - \boldsymbol{X}_{k'}\|_F \geq \varepsilon - 2\rho$ and of the choice $\rho = c\varepsilon/n$ for an appropriate c, we obtain

$$\mathbb{P}_{k,k'} \leq \prod_{i=1}^m \big(1 - \tfrac{\varepsilon}{2\pi}\big) \leq \exp(-\tfrac{\varepsilon}{2\pi}m).$$

Altogether, we conclude that

$$\mathbb{P}_0 \leq K^2 \exp(-\tfrac{\varepsilon}{2\pi}m) = \exp(2\mathcal{H}(\mathcal{K}, c\tfrac{\varepsilon}{n}) - \tfrac{\varepsilon}{2\pi}m),$$

which readily implies the result under the assumption (8). $\qquad\square$

Interestingly, an explicit bound for the Kolmogorov entropy of the set of low-rank and bisparse matrices can be obtained as follows.

Proposition 2. *Given* $r, s \in [\![n]\!]$ *with* $r \leq s$ *and* $\eta > 0$, *one has*

$$\mathcal{H}(\widetilde{\Sigma}_{(s)}^{[r]}, \eta) \leq 2s \ln(en/s) + r(2s + 1)\ln(9/\eta)$$
$$\leq Crs \ln\big(n/(\eta s)\big).$$

Proof. Fixing subsets S, T of $[\![n]\!]$ with $|S| = |T| = s$, one can find an η-covering $\mathcal{G}_{S,T}$ of $\widetilde{\Sigma}_{S \times T}^{[r]}$ with cardinality $|\mathcal{G}_{S,T}| \leq (9/\eta)^{r(2s+1)}$, see [7, Lemma 3.1]. In view of $\widetilde{\Sigma}_{(s)}^{[r]} = \bigcup_{S,T} \widetilde{\Sigma}_S^{[r]}$, an η-covering of $\widetilde{\Sigma}_{(s)}^{[r]}$ is then given by $\mathcal{G} := \bigcup_{S,T} \mathcal{G}_{S,T}$. Using Stirling's bound, we derive that $\mathcal{H}(\widetilde{\Sigma}_{(s)}^{[r]}, \eta) \leq \ln|\mathcal{G}| \leq \ln\big[\binom{n}{s}^2 (9/\eta)^{r(2s+1)}\big] \leq 2s \ln(en/s) + r(2s + 1)\ln(9/\eta)$, which concludes the proof. $\qquad\square$

Theorem 2 and Proposition 2 combine into the following result.

Corollary 1. *Let* $\boldsymbol{\mathcal{A}} : \mathbb{R}^{n \times n} \to \mathbb{R}^m$ *be a Gaussian random map. Given* $\varepsilon > 0$, *if*

$$m \geq \tfrac{C}{\varepsilon}rs \ln(\tfrac{n}{\varepsilon s}), \tag{9}$$

then it occurs with probability at least $1 - \exp(-c\varepsilon m)$ *that* (6) *is fulfilled for all* $\boldsymbol{X}, \boldsymbol{X}' \in \mathcal{K}$.

With Corollary 1 at hand, we can justify the upper bound in (2). For this purpose, consider $\varepsilon > 0$ making (9) an equality. The assumption of Theorem 2 then holds, so there is a linear map $\boldsymbol{\mathcal{A}} : \mathbb{R}^{n \times n} \to \mathbb{R}^m$ such that

$$\sup_{\substack{\boldsymbol{X}, \boldsymbol{X}' \in \widetilde{\Sigma}_{(s)}^{[r]} \\ \operatorname{sgn}\boldsymbol{\mathcal{A}}(\boldsymbol{X}) = \operatorname{sgn}\boldsymbol{\mathcal{A}}(\boldsymbol{X}')}} \|\boldsymbol{X} - \boldsymbol{X}'\|_F \leq \varepsilon.$$

We conclude by remarking that

$$\varepsilon = \tfrac{C}{m}rs \ln(\tfrac{n}{\varepsilon s}) \geq C\tfrac{rs}{m}, \quad \text{and in turn} \quad \varepsilon \leq \tfrac{C}{m}rs \ln(\tfrac{cnm}{rs}),$$

which implies the desired upper bound on $\mathrm{CW}^m(\widetilde{\Sigma}_{(s)}^{[r]})$.

III. Idealized algorithm for Gaussian sensing

Corollary 1 — as well as Theorem 2 for more general sets — says that any $\boldsymbol{X} \in \widetilde{\Sigma}_{(s)}^{[r]}$ is well approximated by any other $\boldsymbol{X}' \in \widetilde{\Sigma}_{(s)}^{[r]}$ sharing the same one-bit measurements $\boldsymbol{y} \in \{\pm 1\}^m$. It does not, however, exhibit a scheme to create an approximant from $\boldsymbol{y}$. We now provide such a scheme by considering the *projected back projection*

$$\boldsymbol{X}' = \mathcal{P}_{\Sigma_{(s)}^{[r]}}(\boldsymbol{\mathcal{A}}^*\boldsymbol{y}). \tag{10}$$

2

Here, $\mathcal{P}_{\Sigma^{[r]}_{(s)}}$ denotes the projection onto the set $\Sigma^{[r]}_{(s)}$, but we stress from the onset that computing it is not a realistic task, see [2] for a related NP-hardness statement. We note also that X' need not be consistent, in the sense that $\operatorname{sgn} \mathcal{A}(X')$ is not necessarily equal to y, but that it has the same structure as X. The essential tool behind our argument is a restricted isometry property for the set $\Sigma^{[r]}_{(s)}$, established below for Gaussian measurements.

Theorem 3. *Let $\mathcal{A} : \mathbb{R}^{n \times n} \to \mathbb{R}^m$ be a Gaussian random map. Given $\delta \in (0, 1)$, if*

$$m \geq \frac{C}{\delta^2} rs \ln\left(\frac{n}{s}\right), \tag{11}$$

then it occurs with probability at least $1 - 2\exp(-c\delta^2 m)$ that, for all $Z \in \Sigma^{[r]}_{(s)}$,

$$(1 - \delta)\|Z\|_F \leq \frac{\sqrt{\pi/2}}{m}\|\mathcal{A}(Z)\|_1 \leq (1 + \delta)\|Z\|_F. \tag{12}$$

Proof. If we overlook the exact powers of δ, the classical proof consisting of a concentration inequality followed by a covering argument will go through — the key ingredient being the estimate of [7, Lemma 3.1] for the covering number of the ball of $\Sigma^{[r]}_{S \times T}$ combined with a union bound over all index sets S, T of size s.

Alternatively, in terms of Gaussian width, we can invoke the result from [8] (see also [9]) which guarantees that (12) holds provided

$$m \geq C\delta^{-2} w\big(\widetilde{\Sigma}^{[r]}_{(s)}\big)^2. \tag{13}$$

Then, in view of $w(\cup_{k=1}^K \mathcal{K}_k) \leq \max_k w(\mathcal{K}_k) + 3\sqrt{\ln K}$ for all subsets $\mathcal{K}_1, \ldots, \mathcal{K}_K$ of the sphere (see [10, Lemma 12(ii)]) and of the fact that $\widetilde{\Sigma}^{[r]}_{(s)} = \cup_{S,T} \widetilde{\Sigma}^{[r]}_{S \times T}$ with $w(\widetilde{\Sigma}^{[r]}_{S \times T})^2 \leq Crs$, we deduce

$$w\big(\widetilde{\Sigma}^{[r]}_{(s)}\big) \leq C\sqrt{rs} + 3\sqrt{2s \ln\left(\frac{en}{s}\right)} \leq C\sqrt{rs \ln\left(\frac{n}{s}\right)}.$$

It is clear that (13) holds under the assumption (11). $\square$

We now state and prove the main result of this section. The argument is an adaptation of the technique presented in [11, Section 8.4].

Theorem 4. *Assuming that (12) holds on $\widetilde{\Sigma}^{[2r]}_{(2s)}$, any $X \in \widetilde{\Sigma}^{[r]}_{(s)}$ acquired via $y = \operatorname{sgn} \mathcal{A}(X)$ is approximated by the matrix X' defined in (10) with error*

$$\|X - X'\|_F \leq C\sqrt{\delta}. \tag{14}$$

Proof. We write the singular value decompositions of X and X' as $X = \sum_{k=1}^r \sigma_k u_k v_k^\top$ and $X' = \sum_{k=1}^r \sigma_k u_k' v_k'^\top$, with the vectors u_k, v_k being supported on some S, T, $|S| = |T| = s$, and the vectors u_k', v_k' being supported on some S', T', $|S'| = |T'| = s$. Let also $\mathcal{P}_S$ denote the orthogonal projection onto the linear space $S = \operatorname{span}\{u_1 v_1^\top, \ldots, u_r v_r^\top, u_1' v_1'^\top, \ldots, u_r' v_r'^\top\} \subseteq \Sigma^{[2r]}_{(2s)}$. Notice that $\mathcal{P}_S(X) = X$ and $\mathcal{P}_S(X - X') = X - X'$ since $X \in S$ and $X - X' \in S$. Because the matrix X' is the best approximant to $\mathcal{A}^* \operatorname{sgn} \mathcal{A}(X)$ from $\Sigma^{[r]}_{(s)}$, it is a better approximant to $\mathcal{A}^* \operatorname{sgn} \mathcal{A}(X)$ than X is, i.e., $\|\mathcal{A}^* \operatorname{sgn} \mathcal{A}(X) - X'\|_F^2 \leq \|\mathcal{A}^* \operatorname{sgn} \mathcal{A}(X) - X\|_F^2$. Introducing X in the left-hand side and expanding the square yields

$$\begin{aligned} \|X - X'\|_F^2 &\leq 2\langle X - X', X - \mathcal{A}^* \operatorname{sgn} \mathcal{A}(X)\rangle_F \\ &= 2\langle \mathcal{P}_S(X - X'), X - \mathcal{A}^* \operatorname{sgn} \mathcal{A}(X)\rangle_F \\ &= 2\langle X - X', X - \mathcal{P}_S(\mathcal{A}^* \operatorname{sgn} \mathcal{A}(X))\rangle_F \\ &\leq 2\|X - X'\|_F \|X - \mathcal{P}_S(\mathcal{A}^* \operatorname{sgn} \mathcal{A}(X))\|_F, \end{aligned}$$

which implies that

$$\|X - X'\|_F \leq 2\|X - \mathcal{P}_S(\mathcal{A}^* \operatorname{sgn} \mathcal{A}(X))\|_F. \tag{15}$$

Taking $\|X\|_F^2 = 1$ into account, we remark that

$$\begin{aligned} \|X - \mathcal{P}_S(\mathcal{A}^* \operatorname{sgn} \mathcal{A}(X))\|_F^2 = 1 &+ \|\mathcal{P}_S(\mathcal{A}^* \operatorname{sgn} \mathcal{A}(X))\|_F^2 \\ &- 2\langle X, \mathcal{P}_S(\mathcal{A}^* \operatorname{sgn} \mathcal{A}(X))\rangle_F. \end{aligned} \tag{16}$$

Now, thanks to $\mathcal{P}_S(\mathcal{A}^* \operatorname{sgn} \mathcal{A}(X)) \in \Sigma^{[2r]}_{(2s)}$, we have

$$\begin{aligned} \|\mathcal{P}_S(\mathcal{A}^* \operatorname{sgn} \mathcal{A}(X))\|_F^2 &= \langle \mathcal{A}^* \operatorname{sgn} \mathcal{A}(X), \mathcal{P}_S(\mathcal{A}^* \operatorname{sgn} \mathcal{A}(X))\rangle_F \\ &= \langle \operatorname{sgn} \mathcal{A}(X), \mathcal{A}(\mathcal{P}_S(\mathcal{A}^* \operatorname{sgn} \mathcal{A}(X)))\rangle \\ &\leq \|\mathcal{A}(\mathcal{P}_S(\mathcal{A}^* \operatorname{sgn} \mathcal{A}(X)))\|_1 \\ &\leq (1 + \delta)\|\mathcal{P}_S(\mathcal{A}^* \operatorname{sgn}(\mathcal{A}X))\|_F, \end{aligned}$$

which simplifies to

$$\|\mathcal{P}_S(\mathcal{A}^* \operatorname{sgn}(\mathcal{A}X))\|_F \leq 1 + \delta. \tag{17}$$

Besides, thanks to $\mathcal{P}_S(X) = X$ and $X \in \Sigma^{[r]}_{(s)}$, we also have

$$\begin{aligned} \langle X, \mathcal{P}_S(\mathcal{A}^* \operatorname{sgn}(\mathcal{A}X))\rangle_F &= \langle X, \mathcal{A}^* \operatorname{sgn}(\mathcal{A}X)\rangle_F \\ &= \langle \mathcal{A}X, \operatorname{sgn}(\mathcal{A}X)\rangle = \|\mathcal{A}X\|_1 \geq (1 - \delta). \end{aligned} \tag{18}$$

Putting (17) and (18) together in (16) gives

$$\|X - \mathcal{P}_S(\mathcal{A}^* \operatorname{sgn} \mathcal{A}(X))\|_F^2 \leq 1 + (1 + \delta)^2 - 2(1 - \delta) \leq C\delta.$$

Finally, substituting in (15), we arrive at $\|X - X'\|_F \leq C\sqrt{\delta}$. $\square$

To close this section, we reformulate (14) by stating that, with high probability, the recovery error for projected back projection with Gaussian random maps achieves the decay rate

$$\|X - X'\|_F \leq C\left(\frac{rs \ln(n/s)}{m}\right)^{1/4},$$

since the restricted isometry property (12) holds in the regime (11).

IV. FACTORIZED GAUSSIAN SENSING

While the projected back projection scheme (10) is impractical, changing the sensing map $\mathcal{A} : \mathbb{R}^{n \times n} \to \mathbb{R}^m$ allows for the design of a practical algorithm to recover matrices in $\widetilde{\Sigma}^{[r]}_{(s)}$ with provable recovery error decaying as a root of $rs \ln(n/s)/m$. The strategy follows a multistep argument put forward in [12] for sparse phase retrieval and also imitated in [2] for low-rank and bisparse recovery from standard, *i.e.*, not quantized, measurements. It consists in defining a sensing map $\mathcal{A} : \mathbb{R}^{n \times n} \to \mathbb{R}^m$ associated with matrices

$$A_i := B^\top A_i' C \in \mathbb{R}^{n \times n}, \quad i \in [\![m]\!],$$

where $A_1', \ldots, A_m' \in \mathbb{R}^{p \times p}$ and $B, C \in \mathbb{R}^{p \times n}$ for some p chosen later. The one-bit measurements made on $X \in \mathbb{R}^{n \times n}$ through the map $\mathcal{A} : \mathbb{R}^{n \times n} \to \mathbb{R}^m$ can also be interpreted as one-bit measurements made on $BXC^\top \in \mathbb{R}^{p \times p}$ through the map $\mathcal{A}' : \mathbb{R}^{p \times p} \to \mathbb{R}^m$ associated with the A_i'. Namely, in view of $\langle B^\top A_i' C, X\rangle_F = \langle A_i', BXC^\top\rangle_F$, we have

$$y = \operatorname{sgn} \mathcal{A}(X) = \operatorname{sgn} \mathcal{A}'(BXC^\top). \tag{19}$$

We will see that, under appropriate assumptions on $\mathcal{A}'$, B, C, a good approximant to $X \in \widetilde{\Sigma}^{[r]}_{(s)}$ constructed from $y = \operatorname{sgn} \mathcal{A}(X)$ is

$$X' := \mathcal{H}^{\mathrm{col}}_{(s)}\{\mathcal{H}^{\mathrm{row}}_{(s)}\big[B^\top \mathcal{H}^{[r]}(\mathcal{A}'^* y)\big] C\}. \tag{20}$$

Here, $\mathcal{H}^{[r]}$ denotes the operator of best rank-r approximation (obtained by keeping r leading singular elements), while $\mathcal{H}^{\mathrm{row}}_{(s)}$ (respectively $\mathcal{H}^{\mathrm{col}}_{(s)}$) stands for the operator of best approximation by s-row-sparse (respectively s-column-sparse) matrices, *i.e.*, obtained by keeping s rows (respectively s columns) with largest ℓ_2-norms.

The appropriate assumption on $\boldsymbol{A}'$ is a restricted isometry property of order $2r$ with constant $\delta' \in (0,1)$, reading

$$(1-\delta')\|\boldsymbol{Z}\|_F \leq \|\boldsymbol{A}'(\boldsymbol{Z})\|_1 \leq (1+\delta')\|\boldsymbol{Z}\|_F \qquad (21)$$

whenever $\boldsymbol{Z} \in \mathbb{R}^{p\times p}$ has rank $\leq 2r$. The appropriate assumptions on $\boldsymbol{D} = \boldsymbol{B}, \boldsymbol{C}$ are standard restricted isometry properties of order $2s$ with constant $\delta \in (0,1)$, reading

$$(1-\delta)\|\boldsymbol{z}\|_2^2 \leq \|\boldsymbol{D}\boldsymbol{z}\|_2^2 \leq (1+\delta)\|\boldsymbol{z}\|_2^2 \qquad (22)$$

whenever $\boldsymbol{z} \in \mathbb{R}^n$ is $2s$-sparse. It is routine to verify that the later condition guarantees that, for any $k \geq 1$,

$$\left|\langle (\boldsymbol{I} - \boldsymbol{D}^\top \boldsymbol{D})\boldsymbol{Z}, \boldsymbol{Z}'\rangle_F\right| \leq \delta\|\boldsymbol{Z}\|_F\|\boldsymbol{Z}'\|_F \qquad (23)$$

whenever the row-supports of $\boldsymbol{Z}, \boldsymbol{Z}' \in \mathbb{R}^{n\times k}$ have combined size at most $2s$. Under this condition, for any s-row-sparse $\boldsymbol{Z} \in \mathbb{R}^{n\times k}$ seen through the inexact measurements $\boldsymbol{Y} = \boldsymbol{D}\boldsymbol{Z} + \boldsymbol{E} \in \mathbb{R}^{p\times k}$, one has

$$\|\boldsymbol{Z} - \mathcal{H}_{(s)}^{\mathrm{row}}[\boldsymbol{D}^\top \boldsymbol{Y}]\|_F \leq 2\delta\|\boldsymbol{Z}\|_F + 2\sqrt{2}\|\boldsymbol{E}\|_F. \qquad (24)$$

Writing $\boldsymbol{Z}' = \mathcal{H}_{(s)}^{\mathrm{row}}[\boldsymbol{D}^\top \boldsymbol{Y}]$ and $\boldsymbol{M} = \boldsymbol{Z} - \boldsymbol{Z}'$, this comes from, using in turn the defining property of $\boldsymbol{Z}'$, (23), and (22),

$$\|\boldsymbol{M}\|_F^2 = \|\boldsymbol{D}^\top \boldsymbol{Y} - \boldsymbol{Z}'\|_F^2 - \|\boldsymbol{D}^\top \boldsymbol{Y} - \boldsymbol{Z}\|_F^2 + 2\langle \boldsymbol{Z} - \boldsymbol{D}^\top \boldsymbol{Y}, \boldsymbol{M}\rangle_F$$
$$\leq 2\langle \boldsymbol{Z} - \boldsymbol{D}^\top \boldsymbol{Y}, \boldsymbol{M}\rangle_F = 2\langle (\boldsymbol{I} - \boldsymbol{D}^\top \boldsymbol{D})\boldsymbol{Z}, \boldsymbol{M}\rangle_F - 2\langle \boldsymbol{E}, \boldsymbol{D}\boldsymbol{M}\rangle_F$$
$$\leq 2\delta\|\boldsymbol{Z}\|_F\|\boldsymbol{M}\|_F + 2\|\boldsymbol{E}\|_F(1+\delta)^{1/2}\|\boldsymbol{M}\|_F.$$

We now state and prove the main result of this section.

Theorem 5. *Given $\delta, \delta' \in (0,1)$, under assumptions (21) on $\boldsymbol{A}'$ and (22) on $\boldsymbol{B}$ and $\boldsymbol{C}$, any $\boldsymbol{X} \in \widetilde{\Sigma}_{(s)}^{[r]}$ acquired via $\boldsymbol{y} = \mathrm{sgn}\,\boldsymbol{A}(\boldsymbol{X})$ is approximated by the matrix $\boldsymbol{X}'$ given in (20) with error*

$$\|\boldsymbol{X} - \boldsymbol{X}'\|_F \leq C(\sqrt{\delta'} + \delta).$$

Proof. In view of (19), the vector $\boldsymbol{y} \in \{\pm 1\}^m$ provides one-bit measurements made on the rank-r matrix $\boldsymbol{B}\boldsymbol{X}\boldsymbol{C}^\top \in \mathbb{R}^{p\times p}$ through the sensing map $\boldsymbol{A}' : \mathbb{R}^{p\times p} \to \mathbb{R}^m$. Since this map satisfies the restricted isometry property (21), it is known (see [13, Theorem 2] or even Theorem 4 with $s = n = p$) that

$$\|c_X \boldsymbol{B}\boldsymbol{X}\boldsymbol{C}^\top - \mathcal{H}^{[r]}(\boldsymbol{A}'^*\boldsymbol{y})\|_F \leq C\sqrt{\delta'}, \qquad (25)$$

with $c_X := \|\boldsymbol{B}\boldsymbol{X}\boldsymbol{C}^\top\|_F^{-1}$. As an aside, we notice that both matrices $\boldsymbol{X}\boldsymbol{C}^\top \in \mathbb{R}^{n\times p}$ and $\boldsymbol{X}^\top \in \mathbb{R}^{n\times n}$ are s-row-sparse, so that

$$(1-\delta)\|\boldsymbol{X}\boldsymbol{C}^\top\|_F^2 \leq \|\boldsymbol{B}\boldsymbol{X}\boldsymbol{C}^\top\|_F^2 \leq (1+\delta)\|\boldsymbol{X}\boldsymbol{C}^\top\|_F^2, \qquad (26)$$
$$(1-\delta)\|\boldsymbol{X}^\top\|_F^2 \leq \|\boldsymbol{C}\boldsymbol{X}^\top\|_F^2 \leq (1+\delta)\|\boldsymbol{X}^\top\|_F^2, \qquad (27)$$

hence, since $\|\boldsymbol{X}\boldsymbol{C}^\top\|_F = \|\boldsymbol{C}\boldsymbol{X}^\top\|_F$ and $\|\boldsymbol{X}^\top\|_F = \|\boldsymbol{X}\|_F = 1$,

$$(1-\delta)^2 \leq \|\boldsymbol{B}\boldsymbol{X}\boldsymbol{C}^\top\|_F^2 \leq (1+\delta)^2. \qquad (28)$$

Now, viewing $\mathcal{H}^{[r]}(\boldsymbol{A}'^*\boldsymbol{y})$ as inexact measurements made on the s-row-sparse matrix $\boldsymbol{Z} := c_X \boldsymbol{X}\boldsymbol{C}^\top \in \mathbb{R}^{n\times p}$ via $\boldsymbol{B}$, namely $\mathcal{H}^{[r]}(\boldsymbol{A}'^*\boldsymbol{y}) = \boldsymbol{B}\boldsymbol{Z} + \boldsymbol{E}$ with $\boldsymbol{E} := \mathcal{H}^{[r]}(\boldsymbol{A}'^*\boldsymbol{y}) - c_X \boldsymbol{B}\boldsymbol{X}\boldsymbol{C}^\top$, (24) implies that

$$\left\| c_X \boldsymbol{X}\boldsymbol{C}^\top - \mathcal{H}_{(s)}^{\mathrm{row}}\big[\boldsymbol{B}^\top \mathcal{H}^{[r]}(\boldsymbol{A}'^*\boldsymbol{y})\big]\right\|_F$$
$$\leq 2\delta c_X \|\boldsymbol{X}\boldsymbol{C}^\top\|_F + 2\sqrt{2}\|\boldsymbol{E}\|_F \leq C\big(\delta + \sqrt{\delta'}\big), \quad (29)$$

where the last inequality used (26) and (25). Similarly, viewing $\big(\mathcal{H}_{(s)}^{\mathrm{row}}\big[\boldsymbol{B}^\top \mathcal{H}^{[r]}(\boldsymbol{A}'^*\boldsymbol{y})\big]\big)^\top = \boldsymbol{C}\boldsymbol{Z}' + \boldsymbol{E}'$ as inexact measurements

made on the s-row-sparse matrix $\boldsymbol{Z}' := c_X \boldsymbol{X}^\top \in \mathbb{R}^{n\times n}$ via $\boldsymbol{C}$ with $\boldsymbol{E}' := \big(\mathcal{H}_{(s)}^{\mathrm{row}}\big[\boldsymbol{B}^\top \mathcal{H}^{[r]}(\boldsymbol{A}'^*\boldsymbol{y})\big]\big)^\top - c_X \boldsymbol{C}\boldsymbol{X}^\top$, (24) implies that

$$\left\| c_X \boldsymbol{X}^\top - \mathcal{H}_{(s)}^{\mathrm{row}}\big\{ \boldsymbol{C}^\top \big(\mathcal{H}_{(s)}^{\mathrm{row}}\big[\boldsymbol{B}^\top \mathcal{H}^{[r]}(\boldsymbol{A}'^*\boldsymbol{y})\big]\big)^\top \big\}\right\|_F$$
$$\leq 2\delta c_X \|\boldsymbol{X}^\top\|_F + 2\sqrt{2}\|\boldsymbol{E}'\|_F \leq C\big(\delta + \sqrt{\delta'}\big), \quad (30)$$

where the last inequality used (28) and (29). By noticing that $\mathcal{H}_{(s)}^{\mathrm{row}}(\boldsymbol{M}^\top) = \big(\mathcal{H}_{(s)}^{\mathrm{col}}(\boldsymbol{M})\big)^\top$, (30) actually reads

$$\|c_X \boldsymbol{X} - \boldsymbol{X}'\|_F \leq C\big(\delta + \sqrt{\delta'}\big).$$

The final result follows from a triangle inequality and the fact that $\|\boldsymbol{X} - c_X \boldsymbol{X}\|_F = |\|\boldsymbol{B}\boldsymbol{X}\boldsymbol{C}^\top\|_F - 1|/\|\boldsymbol{B}\boldsymbol{X}\boldsymbol{C}^\top\|_F \leq C\delta$, which is a consequence of (28). $\square$

We conclude by remarking that assumption (21) is valid when $\boldsymbol{A}' : \mathbb{R}^{p\times p} \to \mathbb{R}^m$ is a properly normalized Gaussian random map with $m \asymp \delta'^{-2}rp$ (see [13, Theorem 1] as an instantiation of [8]) and that assumption (22) is valid when $\boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{p\times n}$ are a properly normalized Gaussian random matrices with $p \asymp \delta^{-2}s\ln(n/s)$. Choosing $\delta' = \delta^2$, we therefore arrive, with high probability, at $\|\boldsymbol{X} - \boldsymbol{X}'\|_F \leq C\delta$ with $m \asymp \delta^{-6}rs\ln(n/s)$, *i.e.*,

$$\|\boldsymbol{X} - \boldsymbol{X}'\|_F \leq C\left(\frac{rs\ln(n/s)}{m}\right)^{1/6}.$$

We note that the power $1/6$ is certainly not optimal — it can *e.g.* be increased by replacing $\mathcal{H}^{[r]}$ by a scheme improving on $\sqrt{\delta'}$ in (25).

References

[1] S. Oymak, A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi, "Simultaneously structured models with application to sparse and low-rank matrices." *IEEE Trans. Information Theory*, **61**(5):2886–2908, 2015.

[2] S. Foucart, R. Gribonval, L. Jacques, and H. Rauhut, "Jointly low-rank and bisparse recovery: questions and partial answers," *preprint*, 2019.

[3] L. Jacques, "Error decay of (almost) consistent signal estimations from quantized gaussian random projections," *IEEE Trans. Information Theory*, **62**(8):4696–4709, 2016.

[4] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *IEEE Trans. Information Theory*, **59**(4):2082–2102, 2013.

[5] S. Oymak and B. Recht, "Near-optimal bounds for binary embeddings of arbitrary sets," *arXiv:1512.04433*, 2015.

[6] D. Bilyk and M. T. Lacey, "Random tessellations, restricted isometric embeddings, and one bit sensing," *arXiv:1512.06697*, 2015.

[7] E. J. Candès and Y. Plan, "Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements," *IEEE Trans. Information Theory*, **57**(4):2342–2359, 2011.

[8] G. Schechtman, "Two observations regarding embedding subsets of euclidean spaces in normed spaces," *Advances in Mathematics*, **200**(1):125–135, 2006.

[9] Y. Plan and R. Vershynin, "Dimension reduction by random hyperplane tessellations," *Discrete & Computational Geometry*, **51**(2):438–461, 2014.

[10] R. G. Baraniuk, S. Foucart, D. Needell, Y. Plan, and M. Wootters, "One-bit compressive sensing of dictionary-sparse signals," *Information and Inference: A Journal of the IMA*, **7**(1):83–104, 2017.

[11] S. Foucart, "Flavors of compressive sensing," in *International Conference on Approximation Theory*. Springer, 2016, pp. 61–104.

[12] M. Iwen, A. Viswanathan, and Y. Wang, "Robust sparse phase retrieval made easy," *Applied and Computational Harmonic Analysis*, **42**(1):135–142, 2017.

[13] S. Foucart and R. G. Lynch, "Recovering low-rank matrices from binary measurements," *preprint*, 2017.