Sparse Recovery from Saturated Measurements

Simon Foucart and Tom Needham — Texas A&M University and University of Georgia

Abstract

A novel theory of sparse recovery is presented in order to bridge the standard compressive sensing framework and the one-bit compressive sensing framework. In the former setting, sparse vectors observed via few linear measurements can be reconstructed exactly. In the latter setting, the linear measurements are only available through their signs, so exact reconstruction of sparse vectors is replaced by estimation of their directions. In the hybrid setting introduced here, a linear measurement is conventionally acquired if is not too large in absolute value, but otherwise it is seen as saturated to plus-or-minus a given threshold. Intuition suggests that sparse vectors of small magnitude should be exactly recoverable, since saturation would not occur, and that sparse vectors of larger magnitude should be accessible though more than just their directions. The purpose of the article is to confirm this intuition and to justify rigorously the following informal statement: measuring at random with Gaussian vectors and reconstructing via an ℓ_1-minimization scheme, it is highly likely that all sparse vectors are faithfully estimated from their saturated measurements as long as the number of saturated measurements marginally exceeds the sparsity level. Faithful estimation means exact reconstruction in a small-magnitude regime and control of the relative reconstruction error in a larger-magnitude regime.

Key words and phrases: Compressive sensing, saturation, linear programming, restricted isometry properties.

AMS classification: 94A12, 90C05, 60D05.

1 Introduction

This article considers a scenario where sparse vectors are to be recovered from few measurements of a novel type. After putting the scenario in the context of standard and one-bit compressive sensing, we motivate the relevance of saturated measurements, and we outline our contribution towards the solution of the problem. We close this introductory section with an account of specific notation used in the rest of the article.

1.1 Standard and one-bit compressive sensing

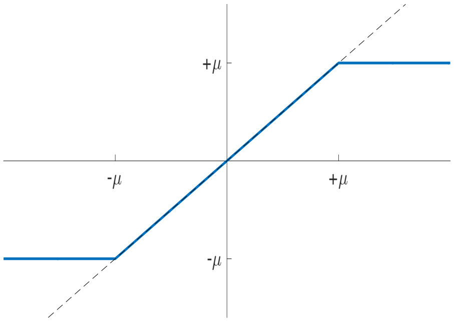
Suppose that high-dimensional vectors $\mathbf{x} \in \mathbb{R}^N$ are acquired ('sensed') via linear measurements $y_1 = \langle \mathbf{a}_1, \mathbf{x} \rangle, \ldots, y_m = \langle \mathbf{a}_m, \mathbf{x} \rangle$ given as inner products of $\mathbf{x}$ with prescribed vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^N$. In matrix form, this information reads $\mathbf{y} = \mathbf{A}\mathbf{x}$, where the rows of the matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ consist of $\mathbf{a}_1^\top, \ldots, \mathbf{a}_m^\top$. Without prior knowledge about the structure of the targeted vectors $\mathbf{x} \in \mathbb{R}^N$, the number m of measurements necessary for the recovery of $\mathbf{x}$ from $\mathbf{y}$ equals N, which is prohibitively large. But it is conceivable that recovering vectors $\mathbf{x} \in \mathbb{R}^N$ belonging to a certain structured class is possible from the 'compressed' information $\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{R}^m$ with $m \ll N$. As a matter of fact, the theory of compressive sensing initiated in the seminal works of Candès et al. [5] and of Donoho [6] made it apparent to the scientific community that a suitable choice of the measurement matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ enables the recovery of all sparse vectors $\mathbf{x} \in \mathbb{R}^N$ acquired as $\mathbf{y} = \mathbf{A}\mathbf{x}$ (the recovery step can be performed using a variety of efficient procedures). One only needs the number m of measurements to scale like the optimal order $s \ln(eN/s)$, where s denotes the sparsity level of the targeted vectors $\mathbf{x} \in \mathbb{R}^N$ (i.e., these vectors have at most s nonzero entries). A self-contained and comprehensive account of this theory of standard compressive sensing can be found in the textbook/monograph [7].

In some practical situations, however, the precise values of $\langle \mathbf{a}_i, \mathbf{x} \rangle$ are inaccessible, not only because of inevitable measurement errors but also because these values must be quantized. Putting aside sophisticated schemes such as $\Sigma\Delta$-quantization [9], we focus on the extreme scenario where one bit of information about $\langle \mathbf{a}_i, \mathbf{x} \rangle$ is retained, so that only $y_1 = \mathrm{sgn}(\langle \mathbf{a}_1, \mathbf{x} \rangle), \ldots, y_m = \mathrm{sgn}(\langle \mathbf{a}_m, \mathbf{x} \rangle)$ are available. In this setting, introduced in [3], one cannot aim at exact recovery of $\mathbf{x} \in \mathbb{R}^N$ but rather at estimates of the type $\|(\mathbf{x}/\|\mathbf{x}\|_2) - \mathbf{x}^\sharp\|_2 \le \delta$ where $\mathbf{x}^\sharp$ is the output of a recovery procedure utilizing $\mathbf{y} = \mathrm{sgn}(\mathbf{A}\mathbf{x}) \in \{\pm 1\}^m$ as an input (note that only the direction $\mathbf{x}/\|\mathbf{x}\|_2$ of $\mathbf{x} \in \mathbb{R}^N$ can be estimated, since $\mathrm{sgn}(\mathbf{A}\mathbf{x})$ is invariant under multiplication of $\mathbf{x}$ by a positive scalar). This objective can be attained for all s-sparse vectors $\mathbf{x} \in \mathbb{R}^N$ using optimization-based procedures, provided that the number m of measurement scales like the order $\delta^{-\gamma} s \ln(eN/s)$ (with $\gamma = 5$ and $\gamma = 12$ in [12] and [13], respectively, but these powers of δ^{-1} are improvable). The existence of suitable measurement matrices $\mathbf{A} \in \mathbb{R}^{m \times N}$ in this regime of parameters (s, m, N) is proved by probabilistic arguments, as in the case of standard compressive sensing. Contrary to the standard case, though, $\mathbf{A} \in \mathbb{R}^{m \times N}$ cannot be taken as arbitrary subgaussian random matrices (e.g. Bernoulli random matrices are not appropriate) and are taken to be Gaussian random matrices (or close, see [2]).

1.2 Saturated measurements

As a generalization of the standard and one-bit compressive sensing situations just described, we suppose that the vectors $\mathbf{x} \in \mathbb{R}^N$ are acquired via $y_1 = \mathcal{F}(\langle \mathbf{a}_i, \mathbf{x} \rangle), \ldots, y_m = \mathcal{F}(\langle \mathbf{a}_m, \mathbf{x} \rangle)$ for some prescribed vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^N$ and for some odd function $\mathcal{F}$ (for a general $\mathcal{F}$, we can reduce to this case by taking measurements in pairs $(\mathbf{a}_i, -\mathbf{a}_i)$ and potentially doubling m). If $\mathcal{F}$ was strictly increasing, then we would directly return to the standard case by preprocessing the measurement y_i with $\mathcal{F}^{-1}$, i.e., by forming $y_i' := \mathcal{F}^{-1}(y_i) = \langle \mathbf{a}_i, \mathbf{x} \rangle$. Hence, it is only relevant to consider odd functions that assign a common value to distinct elements. Since it is more than plausible that physical sensors saturate above some threshold, we shall concentrate our attention on the saturation function $\mathcal{S} = \mathcal{S}_\mu$ defined by

$$\mathcal{S}(t) = \begin{cases} -\mu, & t \in (-\infty, -\mu], \\ t, & t \in (-\mu, +\mu), \\ +\mu, & t \in [+\mu, +\infty). \end{cases}$$



Thus, the s-sparse high-dimensional vectors $\mathbf{x} \in \mathbb{R}^N$ are observed via the saturated measurements

$$y_i = \mathcal{S}(\langle \mathbf{a}_i, \mathbf{x} \rangle), \qquad i = 1, \ldots, m,$$

written in condensed form as $\mathbf{y} = \mathcal{S}(\mathbf{A}\mathbf{x})$. Intuitively, standard compressive sensing corresponds to the case $\mu \to \infty$, while one-bit compressive sensing corresponds to the case $\mu \to 0$ (after preprocessing the y_i with a division by μ).

1.3 Overview of the results

For sparse vectors $\mathbf{x} \in \mathbb{R}^N$, if we know $\mathcal{S}(\langle \mathbf{a}_i, \mathbf{x} \rangle)$, $i = 1, \ldots, m$, then we also know $\mathrm{sgn}(\langle \mathbf{a}_i, \mathbf{x} \rangle)$, so that techniques from one-bit compressive sensing allow us to estimate the direction of $\mathbf{x}$. Can we achieve more than this? Certainly not when $\mathbf{x}$ is very large in magnitude, since all the measurements saturate, making $\mathbf{x}$ indistinguishable from any $\gamma \mathbf{x}$ with $\gamma > 1$. But when $\mathbf{x}$ is very small in magnitude, none of the measurements saturate, putting us in the standard compressive sensing framework, so we expect exact recovery. In the intermediate regime where $\mathbf{x}$ is not too small nor too large in magnitude, approximate recovery of $\mathbf{x}$ (i.e., of its direction and of its magnitude) should be possible. The purpose of this paper is to formalize this statement. The measurement matrices $\mathbf{A} \in \mathbb{R}^{m \times N}$ we shall consider are random matrices with independent Gaussian entries of mean zero and standard deviation σ (unlike standard compressive sensing, we do not impose $\sigma = 1/\sqrt{m}$, which is required for the restricted isometry property to hold). Instead, the 'amplification parameter' σ interacts with the 'saturation parameter' μ and with the magnitude $\|\mathbf{x}\|_2$. In fact, we anticipate the ratio $\sigma\|\mathbf{x}\|_2/\mu$ to be significant, according to the observation that

$$\mathbb{P}\left(\langle \mathbf{a}_i, \mathbf{x}\rangle \text{ is saturated}\right) = \mathbb{P}\left(|\langle \mathbf{a}_i, \mathbf{x}\rangle| \geq \mu\right) = \mathbb{P}\left(\sigma\|\mathbf{x}\|_2|g| \geq \mu\right) = \mathbb{P}\left(|g| \geq \frac{\mu}{\sigma\|\mathbf{x}\|_2}\right),$$

where g represents a standard Gaussian random variable. The recovery procedure we shall consider is a sparsity-promoting optimization program involving the ℓ_1-norm. Precisely, given a vector $\mathbf{y} \in [-\mu, +\mu]^m$ of saturated measurements, the recovery procedure simply reads

$$\text{(1)} \qquad \underset{\mathbf{z}\in\mathbb{R}^N}{\text{minimize}}\ \|\mathbf{z}\|_1 \qquad \text{subject to } \mathcal{S}(\mathbf{A}\mathbf{z}) = \mathbf{y}.$$

After introducing slack variables $\mathbf{c} \in \mathbb{R}^N$, this program can be recast as a linear program, namely

$$\underset{\mathbf{z},\mathbf{c}\in\mathbb{R}^N}{\text{minimize}}\ \sum_{j=1}^{N} c_j \qquad \text{subject to} \qquad -c_j \leq z_j \leq c_j, \quad \text{for all } j = 1, \ldots, N,$$

$$\text{and to} \qquad \begin{cases} \langle \mathbf{a}_i, \mathbf{z}\rangle \leq -\mu, & \text{if } y_i = -\mu, \\ \langle \mathbf{a}_i, \mathbf{z}\rangle = \ \ y_i, & \text{if } -\mu < y_i < \mu, \\ \langle \mathbf{a}_i, \mathbf{z}\rangle \geq +\mu, & \text{if } y_i = +\mu. \end{cases}$$

Solving this problem using CVX [1] for several realizations of Gaussian random matrices $\mathbf{A} \in \mathbb{R}^{m\times N}$ and of sparse Gaussian random vectors $\mathbf{x} \in \mathbb{R}^N$ suggests the experimental behavior depicted in Figure 1 for the reconstruction error as a function of the magnitude $\|\mathbf{x}\|_2$.

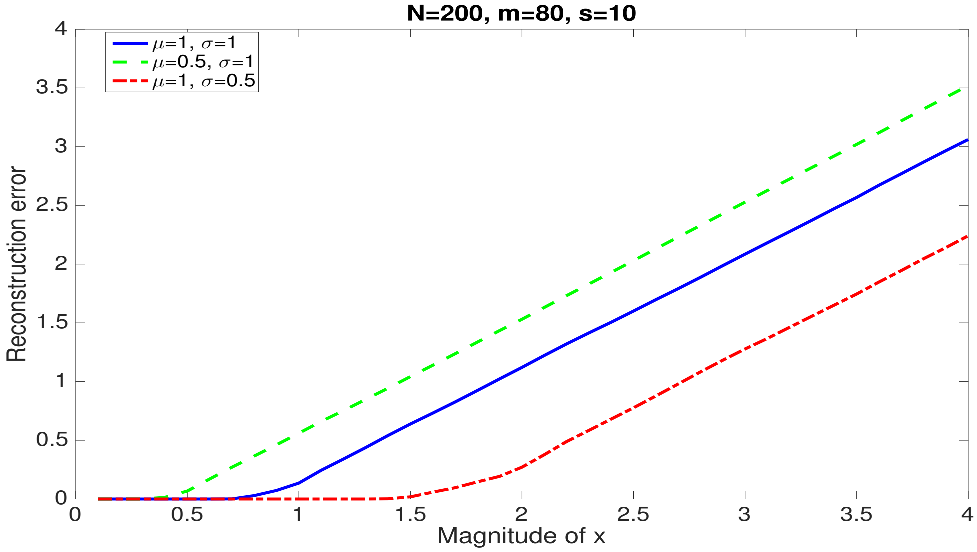


Figure 1: ℓ_1-reconstruction error for sparse vectors of varying magnitude acquired from saturated measurements

We can distinctly perceive the small-magnitude regime where exact recovery occurs. We also detect another regime where the recovery error appears proportional to the magnitude $\|\mathbf{x}\|_2$. The theorem stated below rigorously confirms this observation.

Theorem 1. Suppose that the random matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ has independent $\mathcal{N}(0, \sigma^2)$ entries.

(a) There exists a constant $\alpha > 0$ such that, with probability at least $1 - \gamma \exp(-cm)$, every s-sparse vector $\mathbf{x} \in \mathbb{R}^N$ satisfying $\sigma \|\mathbf{x}\|_2 \leq \alpha \mu$ is exactly recovered from $\mathbf{y} = \mathcal{S}(\mathbf{A}\mathbf{x})$ as a solution of (1), provided $m \geq C s \ln(eN/s)$;

(b) Given $\delta \in (0, 1)$ and any $\beta \geq \alpha$, with probability at least $1 - \gamma \exp(-c_\beta \delta^2 m)$, every s-sparse vector $\mathbf{x} \in \mathbb{R}^N$ satisfying $\alpha\mu \leq \sigma \|\mathbf{x}\|_2 \leq \beta\mu$ is approximately recovered from $\mathbf{y} = \mathcal{S}(\mathbf{A}\mathbf{x})$ as a solution of (1) with relative error at most δ, provided $m \geq C_\beta \delta^{-4} s \ln(eN/s)$.

The constants $C, c, \gamma > 0$ are universal, while the constants $C_\beta, c_\beta > 0$ depend only on β.

We point out that the large-magnitude regime, manifest for a fixed matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$, is only transitioned into when considering random matrices $\mathbf{A} \in \mathbb{R}^{m \times N}$. Indeed, as more and more random measurements are taken, there are bound to be nonsaturated measurements, and these enhance the reconstruction. However, while arbitrarily large β are allowed, the result becomes meaningless in the sense that the requirement on m becomes too demanding and the success probability becomes negligible. The remaining of this article is now dedicated to the proof of Theorem 1: Part (a) is established in Section 2 and Part (b) is established in Section 3.

1.4 Notation

The set of all s-sparse vectors in $\mathbb{R}^N$ is denoted by Σ_s, while the set of all effectively s-sparse vectors in $\mathbb{R}^N$ is denoted by Σ_s^{eff}, i.e.,

$$\Sigma_s^{\text{eff}} := \left\{ \mathbf{x} \in \mathbb{R}^N : \|\mathbf{x}\|_1 \leq \sqrt{s} \|\mathbf{x}\|_2 \right\}.$$

Given a measurement vector $\mathbf{y} \in [-\mu, +\mu]^m$, we define the index sets $\mathcal{I}_{\text{nonsat}}$, $\mathcal{I}_{\text{possat}}$, and $\mathcal{I}_{\text{negsat}}$ corresponding to nonsaturated, positively saturated, and negatively saturated measurements by

$$\begin{aligned} \mathcal{I}_{\text{nonsat}} &:= \{ i \in [\![1:m]\!] : |y_i| < \mu \} && \text{of cardinality } m_{\text{nonsat}}, \\ \mathcal{I}_{\text{possat}} &:= \{ i \in [\![1:m]\!] : y_i = +\mu \} && \text{of cardinality } m_{\text{possat}}, \\ \mathcal{I}_{\text{negsat}} &:= \{ i \in [\![1:m]\!] : y_i = -\mu \} && \text{of cardinality } m_{\text{negsat}}. \end{aligned}$$

Often, the dependence on $\mathbf{y}$ is not explicitly stated because the context is clear, but sometimes writing e.g. $\mathcal{I}_{\text{nonsat}}(\mathbf{y})$ will prove necessary.

Given a matrix $\mathbf{A}$ and and index set $\mathcal{I}$, the matrix $\mathbf{A}_\mathcal{I}$ represents the row-submatrix of $\mathbf{A}$ where only the rows indexed by $\mathcal{I}$ are selected — note the deviation from the widespread usage in compressive sensing where $\mathbf{A}_\mathcal{I}$ would represent a column-submatrix.

2 The small-magnitude regime

This section is devoted to proving Part (a) of Theorem 1. We first collect some supporting results before turning to the proof itself.

2.1 Auxiliary lemmas

We start by stating a necessary and sufficient condition for a vector $\mathbf{x} \in \mathbb{R}^N$ acquired via $\mathbf{y} = \mathcal{S}(\mathbf{A}\mathbf{x})$ to be exactly recovered as a solution of (1).

Lemma 2. Let $\mathbf{x} \in \mathbb{R}^N$ be acquired via $\mathbf{y} = \mathcal{S}(\mathbf{A}\mathbf{x})$. Then $\mathbf{x}$ is the unique minimizer of $\|\mathbf{z}\|_1$ subject to $\mathcal{S}(\mathbf{A}\mathbf{z}) = \mathbf{y}$ if and only if

$$(2) \quad \sum_{j \in S} \operatorname{sgn}(x_j) u_j < \sum_{\ell \in \overline{S}} |u_\ell| \quad \text{for all } \mathbf{u} \in \mathbb{R}^N \setminus \{\mathbf{0}\} \text{ obeying } \begin{cases} \langle \mathbf{a}_i, \mathbf{u} \rangle \geq \langle \mathbf{a}_i, \mathbf{x} \rangle + \mu, & i \in \mathcal{I}_{\text{negsat}}, \\ \langle \mathbf{a}_i, \mathbf{u} \rangle = 0, & i \in \mathcal{I}_{\text{nonsat}}, \\ \langle \mathbf{a}_i, \mathbf{u} \rangle \leq \langle \mathbf{a}_i, \mathbf{x} \rangle - \mu, & i \in \mathcal{I}_{\text{possat}}, \end{cases}$$

where S denotes the support of $\mathbf{x}$.

Proof. Let us assume that (2) holds, and let $\mathbf{z} \neq \mathbf{x}$ satisfy $\mathcal{S}(\mathbf{A}\mathbf{z}) = \mathbf{y}$. Writing $\mathbf{z} = \mathbf{x} - \mathbf{u}$ for some $\mathbf{u} \in \mathbb{R}^N \setminus \{\mathbf{0}\}$,

$$\begin{array}{llcll} \langle \mathbf{a}_i, \mathbf{z} \rangle \leq -\mu, & i \in \mathcal{I}_{\text{negsat}} & \text{becomes} & \langle \mathbf{a}_i, \mathbf{u} \rangle \geq \langle \mathbf{a}_i, \mathbf{x} \rangle + \mu, & i \in \mathcal{I}_{\text{negsat}}, \\ \langle \mathbf{a}_i, \mathbf{z} \rangle = \langle \mathbf{a}_i, \mathbf{x} \rangle, & i \in \mathcal{I}_{\text{nonsat}} & \text{becomes} & \langle \mathbf{a}_i, \mathbf{u} \rangle = 0, & i \in \mathcal{I}_{\text{nonsat}}, \\ \langle \mathbf{a}_i, \mathbf{z} \rangle \geq +\mu, & i \in \mathcal{I}_{\text{possat}} & \text{becomes} & \langle \mathbf{a}_i, \mathbf{u} \rangle \leq \langle \mathbf{a}_i, \mathbf{x} \rangle - \mu, & i \in \mathcal{I}_{\text{possat}}. \end{array}$$

These are the linear constraints in (2), so $\sum_{j \in S} \operatorname{sgn}(x_j) u_j < \sum_{\ell \in \overline{S}} |u_\ell|$ holds. It follows that

$$\begin{aligned} \|\mathbf{z}\|_1 &= \sum_{j \in S} |z_j| + \sum_{\ell \in \overline{S}} |z_\ell| \geq \sum_{j \in S} \operatorname{sgn}(x_j) z_j + \sum_{\ell \in \overline{S}} |z_\ell| = \sum_{j \in S} \operatorname{sgn}(x_j) x_j - \sum_{j \in S} \operatorname{sgn}(x_j) u_j + \sum_{\ell \in \overline{S}} |u_\ell| \\ &> \sum_{j \in S} \operatorname{sgn}(x_j) x_j = \|\mathbf{x}\|_1. \end{aligned}$$

This proves that $\mathbf{x}$ is the unique minimizer of $\|\mathbf{z}\|_1$ subject to $\mathcal{S}(\mathbf{A}\mathbf{z}) = \mathbf{y}$.

Conversely, let us assume that $\mathbf{x}$ is the unique minimizer of $\|\mathbf{z}\|_1$ subject to $\mathcal{S}(\mathbf{A}\mathbf{z}) = \mathbf{y}$. Let $\mathbf{u} \in \mathbb{R}^N \setminus \{\mathbf{0}\}$ obey the linear constraints in (2). For $t \in (0,1)$, the vector $t\mathbf{u}$ also satisfies these constraints (as a convex combination of $\mathbf{u}$ and $\mathbf{0}$, which both satisfy the constraints). Furthermore, if t is small enough, then

$$\operatorname{sgn}(x_j - t u_j) = \operatorname{sgn}(x_j) \qquad \text{for all } j \in S.$$

Defining $\mathbf{z} := \mathbf{x} - t\mathbf{u}$, we easily check that $\mathcal{S}(\mathbf{A}\mathbf{z}) = \mathbf{y}$. The minimality of $\mathbf{x}$ therefore implies that $\|\mathbf{x}\|_1 < \|\mathbf{z}\|_1$. It follows that

$$\begin{aligned}\|\mathbf{x}\|_1 &< \sum_{j \in S} |z_j| + \sum_{\ell \in \overline{S}} |z_\ell| = \sum_{j \in S} \operatorname{sgn}(x_j - tu_j)(x_j - tu_j) + \sum_{\ell \in \overline{S}} t|u_\ell| \\ &= \sum_{j \in S} \operatorname{sgn}(x_j) x_j - t \sum_{j \in S} \operatorname{sgn}(x_j) u_j + t \sum_{\ell \in \overline{S}} |u_\ell| = \|\mathbf{x}\|_1 - t \left(\sum_{j \in S} \operatorname{sgn}(x_j) u_j - \sum_{\ell \in \overline{S}} |u_\ell| \right).\end{aligned}$$

Rearranging this inequality yields $\sum_{j \in S} \operatorname{sgn}(x_j) u_j < \sum_{\ell \in \overline{S}} |u_\ell|$, as expected. □

As a consequence of Lemma 2, we emphasize that the recovery of a sparse vector as solution of (1) is guaranteed under the null space property, not for the matrix $\mathbf{A}$ itself but for a row-submatrix. Recall that a matrix $\mathbf{B}$ is said to satisfy the null space property (NSP for short) of order s if (see e.g. [7, Definition 4.1])

$$\|\mathbf{v}_S\|_1 < \|\mathbf{v}_{\overline{S}}\|_1 \qquad \text{for all } \mathbf{v} \in \ker(\mathbf{B}) \setminus \{\mathbf{0}\} \text{ and all } S \subseteq [\![1:N]\!] \text{ with } \operatorname{card}(S) \le s.$$

Corollary 3. Let $\mathbf{x} \in \mathbb{R}^N$ be a fixed s-sparse vector. If $\mathbf{A}_{\mathcal{I}_{\text{nonsat}}(\mathbf{A}\mathbf{x})}$ satisfies the null space property of order s, then $\mathbf{x}$ is the unique minimizer of $\|\mathbf{z}\|_1$ subject to $\mathcal{S}(\mathbf{A}\mathbf{z}) = \mathcal{S}(\mathbf{A}\mathbf{x})$.

Proof. Let $\mathbf{u} \in \mathbb{R}^N \setminus \{\mathbf{0}\}$ obey the linear constraints in (2). In particular, we see that $\mathbf{u}$ belongs to the null space of $\mathbf{A}_{\mathcal{I}_{\text{nonsat}}(\mathbf{A}\mathbf{x})}$. Thus, for any index S of size at most s, we have $\sum_{j \in S} |u_j| < \sum_{\ell \in \overline{S}} |u_\ell|$. We immediately derive, with $S := \operatorname{supp}(\mathbf{x})$, that $\sum_{j \in S} \operatorname{sgn}(x_j) u_j < \sum_{\ell \in \overline{S}} |u_\ell|$. This establishes (2). The announced result now follows from Lemma 2. □

The difficulty in establishing that the row-submatix $\mathbf{A}_{\mathcal{I}_{\text{nonsat}}(\mathbf{A}\mathbf{x})}$ satisfies the null space property for a Gaussian matrix $\mathbf{A}$ comes from the fact that the index set $\mathcal{I}_{\text{nonsat}}(\mathbf{A}\mathbf{x})$ is random, too. So we will show that the size $m_{\text{nonsat}}(\mathbf{A}\mathbf{x})$ of $\mathcal{I}_{\text{nonsat}}(\mathbf{A}\mathbf{x})$ is suitably large and that all index sets $\mathcal{I}$ with this suitably large size yield row-submatrices $\mathbf{A}_{\mathcal{I}}$ satisfying the null space property. The number $m_{\text{nonsat}}(\mathbf{A}\mathbf{x})$ of nonsaturated measurement is intuitively large when $\|\mathbf{x}\|_2$ is small. In fact, a relation between $\|\mathbf{x}\|_2$ and $m_{\text{nonsat}}(\mathbf{A}\mathbf{x}) = \sum_i X_i$, with $X_i = 1$ if $|\langle \mathbf{a}_i, \mathbf{x} \rangle| < \mu$ and $X_i = 0$ otherwise, can be obtained from Hoeffding's inequality, but it will only be valid when $\mathbf{x}$ is fixed. To make our results uniform over all sparse vectors $\mathbf{x}$ simultaneously, we will rely on the following lemma[1] and its corollary.

Lemma 4. Let $\delta \in (0,1)$. Suppose that the random matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ has independent $\mathcal{N}(0, \sigma^2)$ entries. There are absolute constants $C_1, c_1, \gamma_1 > 0$ such that, if $m \ge C_1 \delta^{-12} s \ln(eN/s)$, then with

[1] the powers of δ are not optimal in Lemma 4 — our current rendition of the original result of [13] is not optimal and this original result is not optimal either (see [4] for an improvement).

probability at least $1 - \gamma_1 \exp(-c_1 \delta^4 m)$, one has

$$\left| \langle \mathrm{sgn}(\mathbf{A}\mathbf{u}), \mathbf{A}\mathbf{v} \rangle - \sqrt{\frac{2}{\pi}} \sigma m \left\langle \frac{\mathbf{u}}{\|\mathbf{u}\|_2}, \mathbf{v} \right\rangle \right| \leq \delta \sqrt{\frac{2}{\pi}} \sigma m \|\mathbf{v}\|_2 \tag{3}$$

for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$ that are sums of two effectively s-sparse vectors.

The property (3) was dubbed sign product embedding property in [11], but it was established earlier in [13]. In fact, [13, Proposition 4.3] showed that, with failure probability at most $\gamma \exp(-c\delta^4 m)$, (3) is valid for all $\mathbf{u}, \mathbf{v}$ in an arbitrary set K provided $m \geq C\delta^{-12} w(K)^2$. Here, $w(K)$ denotes the mean width of K defined by

$$w(K) := \mathbb{E}\left[\sup_{\mathbf{x} \in K - K} \langle \mathbf{g}, \mathbf{x} \rangle \right],$$

where $\mathbf{g} \in \mathbb{R}^N$ represents a random vector with independent $\mathcal{N}(0,1)$ entries. To derive Lemma 4, we have taken $K = (\Sigma_s^{\mathrm{eff}} \cap B_2^N) - (\Sigma_s^{\mathrm{eff}} \cap B_2^N)$ and we have used [13, (3.3)] to bound its mean width as $w(K) \leq 2w(\Sigma_s^{\mathrm{eff}} \cap B_2^N) \leq C' \sqrt{s \ln(eN/s)}$. As a matter of fact, the result in this section only uses the instantiation of (3) to the case $\mathbf{u} = \mathbf{v}$ — a result that has a 'restricted isometry flavor' and that can also be obtained via the more classical strategy of concentration inequality followed by covering arguments.[2]

Corollary 5. Let $\delta \in (0,1)$. Suppose that the random matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ has independent $\mathcal{N}(0, \sigma^2)$ entries. There are absolute constants $C_1, c_1, \gamma_1 > 0$ such that, if $m \geq C_1 \delta^{-12} s \ln(eN/s)$, then with probability at least $1 - \gamma_1 \exp(-c_1 \delta^4 m)$, one has

$$(1-\delta) \sqrt{\frac{2}{\pi}} \sigma m \|\mathbf{u}\|_2 \leq \|\mathbf{A}\mathbf{u}\|_1 \leq (1+\delta) \sqrt{\frac{2}{\pi}} \sigma m \|\mathbf{u}\|_2 \tag{4}$$

for all $\mathbf{u} \in \mathbb{R}^N$ that are sums of two effectively s-sparse vectors.

2.2 Main result

It is time to prove Part (a) of Theorem 1. In the formal statement below, $\Delta(\mathbf{y})$ denotes the output of the recovery scheme (1).

Proposition 6. Suppose that the random matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ has independent $\mathcal{N}(0, \sigma^2)$ entries. There are absolute constants $C, c, \gamma, \alpha > 0$ such that, if $m \geq Cs \ln(eN/s)$, then with probability at least $1 - \gamma \exp(-cm)$, one has

$$\Delta(\mathcal{S}(\mathbf{A}\mathbf{x})) = \mathbf{x}$$

for all s-sparse $\mathbf{x} \in \mathbb{R}^N$ satisfying $\sigma \|\mathbf{x}\|_2 \leq \alpha \mu$.

[2] such an approach would improve the powers of δ, too.

Proof. We select the constant $\alpha \in (0,1/2)$ small enough so that $\alpha \ln(e/\alpha) \le c'/4$, where $c' > 0$ is an absolute constant appearing later in the proof. By Corollary 3, we have

$$\begin{aligned}
&\mathbb{P}(\Delta(\mathcal{S}(\mathbf{Ax})) \neq \mathbf{x} \text{ for some } s\text{-sparse } \mathbf{x} \in \mathbb{R}^N \text{ with } \sigma\|\mathbf{x}\|_2 \le \alpha\mu)\\
&\qquad \le \mathbb{P}(\mathbf{A}_{\mathcal{I}_{\mathrm{nonsat}}(\mathbf{Ax})} \text{ fails the NSP of order } s \text{ for some } s\text{-sparse } \mathbf{x} \in \mathbb{R}^N \text{ with } \sigma\|\mathbf{x}\|_2 \le \alpha\mu)\\
(5)\qquad &\qquad \le \mathbb{P}(m_{\mathrm{nonsat}}(\mathbf{Ax}) < (1-\alpha)m \text{ for some } s\text{-sparse } \mathbf{x} \in \mathbb{R}^N \text{ with } \sigma\|\mathbf{x}\|_2 \le \alpha\mu)\\
(6)\qquad &\qquad + \mathbb{P}(\mathbf{A}_{\mathcal{I}} \text{ fails the NSP of order } s \text{ for some } \mathcal{I} \subseteq [\![1:m]\!] \text{ of size } \ge (1-\alpha)m).
\end{aligned}$$

With the constant C chosen so that $C \ge 4^{12}C_1$, we can place ourselves in the conditions of applicability of Corollary 5 with $\delta = 1/4$. Thus, for any s-sparse $\mathbf{x} \in \mathbb{R}^N$ with $\sigma\|\mathbf{x}\|_2 \le \alpha\mu$, we have, on the one hand,

$$\|\mathbf{Ax}\|_1 \le \left(1 + \frac{1}{4}\right)\sqrt{\frac{2}{\pi}}\sigma m \|\mathbf{x}\|_2 \le m\sigma\|\mathbf{x}\|_2 \le m\alpha\mu,$$

and, on the other hand,

$$\|\mathbf{Ax}\|_1 = \sum_{i=1}^m |\langle \mathbf{a}_i, \mathbf{x}\rangle| \ge \sum_{i \in \mathcal{I}_{\mathrm{sat}}(\mathbf{Ax})} |\langle \mathbf{a}_i, \mathbf{x}\rangle| \ge \sum_{i \in \mathcal{I}_{\mathrm{sat}}(\mathbf{Ax})} \mu = m_{\mathrm{sat}}(\mathbf{Ax})\mu.$$

Putting these two inequalities together yields

$$m_{\mathrm{sat}}(\mathbf{Ax}) \le \alpha m, \qquad \text{i.e.,} \qquad m_{\mathrm{nonsat}}(\mathbf{Ax}) \ge (1-\alpha)m.$$

This shows that the probability in (5) is bounded above by

$$\gamma_1 \exp(-c_1 m/256).$$

Next, it is known that a random matrix $\mathbf{A}' \in \mathbb{R}^{m' \times N}$ with independent $\mathcal{N}(0,\sigma^2)$ entries satisfies the null space property of order s with probability at least $1 - 2\exp(-c'm' + c''s\ln(eN/s))$, where $c', c'' > 0$ are absolute constants. Consequently, in view of

$$\sum_{m' \ge (1-\alpha)m} \binom{m}{m'} = \sum_{k \le \alpha m} \binom{m}{k} \le \sum_{k \le \alpha m} \frac{m^k}{k!} \le \sum_{k \le \alpha m} \frac{(\alpha m)^k}{k!}\left(\frac{1}{\alpha}\right)^{\alpha m} \le e^{\alpha m}\left(\frac{1}{\alpha}\right)^{\alpha m} = \left(\frac{e}{\alpha}\right)^{\alpha m},$$

a union bound allows us to bound the probability in (6) by

$$\begin{aligned}
\left(\frac{e}{\alpha}\right)^{\alpha m} 2\exp\left(-c'(1-\alpha)m + c''s\ln\left(\frac{eN}{s}\right)\right) &\le 2\exp\left(\alpha\ln\left(\frac{e}{\alpha}\right)m\right)\exp\left(-\frac{c'}{2}m + c''s\ln\left(\frac{eN}{s}\right)\right)\\
&\le 2\exp\left(-\frac{c'}{4}m + c''s\ln\left(\frac{eN}{s}\right)\right) \le 2\exp\left(-\frac{c'}{8}m\right),
\end{aligned}$$

where the last step used $c''s\ln(eN/s) \le (c'/8)m$, which comes from the requirement $m \ge Cs\ln(eN/s)$ with $C \ge 8c''/c'$. All in all, we have obtained

$$\begin{aligned}
&\mathbb{P}(\Delta(\mathcal{S}(\mathbf{Ax})) \neq \mathbf{x} \text{ for some } s\text{-sparse } \mathbf{x} \in \mathbb{R}^N \text{ with } \sigma\|\mathbf{x}\|_2 \le \alpha\mu)\\
&\qquad \le \gamma_1 \exp(-c_1 m/256) + 2\exp\left(-\frac{c'}{8}m\right) \le \gamma\exp(-cm)
\end{aligned}$$

for some appropriate constants $c, \gamma > 0$. The proof is now complete. □

Remark. The results of this section only relied on the null space property and the modified restricted isometry property of Corollary 5 (the former actually being a consequence of the latter). Such properties are more generally valid for subexponential random matrices, see [8] to guide the reasoning. Thus, exact recovery in the small-magnitude regime will also occur for non-Gaussian matrices. In the intermediate-magnitude regime, the arguments proposed in the next section appear tied to the Gaussian case. It is plausible, however, that the results will extend to non-Gaussian matrices as well.

3 The intermediate-magnitude regime

This section is devoted to proving Part (b) of Theorem 1. Again, we first collect some supporting results before turning to the proof itself.

3.1 Auxiliary lemmas

The subsequent arguments rely in part on a property established in [14, Theorem 3.1] concerning random tessellations of the 'effectively sparse sphere'. Note that, if we overlook the powers of δ, then the result can easily be derived from Lemma 4.

Lemma 7. *Let $\delta \in (0,1)$. Suppose that the random matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ has independent $\mathcal{N}(0,\sigma^2)$ entries. There are absolute constants $C_2, c_2, \gamma_2 > 0$ such that, if $m \geq C_2 \delta^{-4} s \ln(eN/s)$, then with probability at least $1 - \gamma_2 \exp(-c_2 \delta^4 m)$, one has, for all $\mathbf{u}, \mathbf{v} \in \Sigma_s^{\rm eff}$ with $\|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1$,*

$$[\mathrm{sgn}(\mathbf{A}\mathbf{u}) = \mathrm{sgn}(\mathbf{A}\mathbf{v})] \Longrightarrow [\|\mathbf{u} - \mathbf{v}\|_2 \leq \delta].$$

The subsequent arguments also rely on a uniform concentration property for the ℓ_1-norm of vectors of saturated measurements. We establish this property in the rest of this subsection. Its statement involves a function $\widetilde{\mathcal{S}} = \widetilde{\mathcal{S}}_\mu$ associated to the saturation function $\mathcal{S} = \mathcal{S}_\mu$ and defined, for $t > 0$, by

$$\widetilde{\mathcal{S}}(t) = \sqrt{\frac{2}{\pi}} t \left(1 - \exp\left(-\frac{\mu^2}{2t^2}\right)\right) + 2\mu Q\left(\frac{\mu}{t}\right).$$

Here we used the customary notation of Q-function for the tail probability of a standard normal distribution, i.e., Q is the decreasing function from $(-\infty,+\infty)$ into $(0,1)$ given by $Q(t) = \mathbb{P}(g \geq t)$, where $g \sim \mathcal{N}(0,1)$. It is worth pointing out that, if $t \leq \beta\mu$ for some $\beta > 0$ and if $\tau := \mu/t$, then

$$\widetilde{\mathcal{S}}(t) \geq \sqrt{\frac{2}{\pi}} t \left(1 - \exp\left(-\frac{\tau^2}{2}\right)\right) + \frac{t}{\beta} 2Q(\tau) \geq \eta_\beta t \tag{7}$$

for some $\eta_\beta > 0$ depending only on β. The latter inequality used the fact that either $1-\exp(-\tau^2/2)$ or $2Q(\tau)$ is larger than some absolute constant (since $1-\exp(-\tau^2/2)$ increases from 0 to 1 for $\tau \in [0,\infty)$ and $2Q(\tau)$ decreases from 1 to 0 for $\tau \in [0,\infty)$). It is also worth pointing out that, if $t \le \beta\mu$, then

$$\widetilde{\mathcal{S}}'(t) = \sqrt{\frac{2}{\pi}}\left(1-\exp\left(-\frac{\mu^2}{2t^2}\right)\right) \ge \nu_\beta, \qquad \nu_\beta := \sqrt{\frac{2}{\pi}}\left(1-\exp\left(-\frac{1}{2\beta^2}\right)\right). \tag{8}$$

These observations being made, we can state and prove the aforementioned uniform concentration property.

Lemma 8. Let $\delta \in (0,1)$ and let $\beta_0 \ge \alpha_0 > 0$ be fixed. Suppose that the random matrix $\mathbf{A} \in \mathbb{R}^{m\times N}$ has independent $\mathcal{N}(0,\sigma^2)$ entries. There are constants $C_3 = C_3(\alpha_0,\beta_0), c_3 = c_3(\beta_0), \gamma_3 > 0$ such that, if $m \ge C_3\delta^{-4}s\ln(eN/s)$, then with probability at least $1-\gamma_3\exp(-c_3\delta^2 m)$, one has

$$(1-\delta)m\,\widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2) \le \|\mathcal{S}(\mathbf{A}\mathbf{u})\|_1 \le (1+\delta)m\,\widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2)$$

for all effectively s-sparse $\mathbf{u} \in \mathbb{R}^N$ satisfying $\alpha_0\mu \le \sigma\|\mathbf{u}\|_2 \le \beta_0\mu$.

Proof. The proof consists of the several steps detailed below.
Expectation calculation: for any fixed $\mathbf{u} \in \mathbb{R}^N$,

$$\mathbb{E}\left(\|\mathcal{S}(\mathbf{A}\mathbf{u})\|_1\right) = m\widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2).$$

Indeed, since $\|\mathcal{S}(\mathbf{A}\mathbf{u})\|_1 = \sum_{i=1}^m \mathcal{S}(|\langle \mathbf{a}_i,\mathbf{u}\rangle|)$ and $\langle \mathbf{a}_i,\mathbf{u}\rangle$ is a Gaussian random variable with mean zero and standard deviation $\sigma\|\mathbf{u}\|_2$, it is enough to show that $\mathbb{E}(\mathcal{S}(\theta g)) = \widetilde{\mathcal{S}}(\theta)$ when $g \sim \mathcal{N}(0,1)$. This simply follows from

$$\begin{aligned}\mathbb{E}(\mathcal{S}(\theta g)) &= \int_{-\infty}^{\infty} \mathcal{S}(\theta t)\frac{\exp(-t^2/2)}{\sqrt{2\pi}}dt = 2\left[\int_0^{\mu/\theta} \theta t\frac{\exp(-t^2/2)}{\sqrt{2\pi}}dt + \int_{\mu/\theta}^{\infty} \mu\frac{\exp(-t^2/2)}{\sqrt{2\pi}}dt\right]\\ &= \sqrt{\frac{2}{\pi}}\theta\left(1-\exp\left(-\frac{\mu^2}{2\theta^2}\right)\right) + 2\mu Q\left(\frac{\mu}{\theta}\right) = \widetilde{\mathcal{S}}(\theta).\end{aligned}$$

Concentration inequality: For any fixed $\mathbf{u} \in \mathbb{R}^N$ with $\sigma\|\mathbf{u}\|_2 \le \beta_0\mu$,

$$\mathbb{P}\left(\left|\|\mathcal{S}(\mathbf{A}\mathbf{u})\|_1 - m\widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2)\right| > \varepsilon m\widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2)\right) \le 2\exp\left(-\frac{\eta_{\beta_0}^2\varepsilon^2 m}{2}\right).$$

We interpret $\|\mathcal{S}(\mathbf{A}\mathbf{u})\|_1$ as a function of the vectorization of $\sigma^{-1}\mathbf{A}$, which is a standard Gaussian random vector of size mN. This is a Lipschitz function with constant $\sqrt{m}\sigma\|\mathbf{u}\|_2$, as seen from

$$\begin{aligned}
\big|\|\mathcal{S}(\mathbf{A}\mathbf{u})\|_1 - \|\mathcal{S}(\mathbf{B}\mathbf{u})\|_1\big| &= \left|\sum_{i=1}^m \left(|\mathcal{S}(\langle \mathbf{a}_i, \mathbf{u}\rangle)| - |\mathcal{S}(\langle \mathbf{b}_i, \mathbf{u}\rangle)|\right)\right| \leq \sum_{i=1}^m |\mathcal{S}(\langle \mathbf{a}_i, \mathbf{u}\rangle) - \mathcal{S}(\langle \mathbf{b}_i, \mathbf{u}\rangle)| \\
&\leq \sum_{i=1}^m |\langle \mathbf{a}_i, \mathbf{u}\rangle - \langle \mathbf{b}_i, \mathbf{u}\rangle| = \sum_{i=1}^m |\langle \mathbf{a}_i - \mathbf{b}_i, \mathbf{u}\rangle| \leq \sum_{i=1}^m \|\mathbf{a}_i - \mathbf{b}_i\|_2 \|\mathbf{u}\|_2 \\
&\leq \sqrt{m}\left[\sum_{i=1}^m \|\mathbf{a}_i - \mathbf{b}_i\|_2^2\right]^{1/2} \|\mathbf{u}\|_2 = \sqrt{m}\|\mathbf{u}\|_2 \|\mathbf{A} - \mathbf{B}\|_2 = \sqrt{m}\sigma\|\mathbf{u}\|_2 \|\sigma^{-1}\mathbf{A} - \sigma^{-1}\mathbf{B}\|_2.
\end{aligned}$$

In view of the concentration of measures for Lipschitz functions (see e.g. [7, Theorem 8.40]) and of $\mathbb{E}(\|\mathcal{S}(\mathbf{A}\mathbf{u})\|_1) = m\widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2)$, we obtain

$$\mathbb{P}\left(\Big|\|\mathcal{S}(\mathbf{A}\mathbf{u})\|_1 - m\widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2)\Big| > \varepsilon m\widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2)\right) \leq 2\exp\left(-\frac{\varepsilon^2 m^2 \widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2)^2}{2m\sigma^2\|\mathbf{u}\|_2^2}\right) \leq 2\exp\left(-\frac{\eta_{\beta_0}^2 \varepsilon^2 m}{2}\right),$$

where the last step made use of (7).

Uniform concentration on effectively sparse hollowed balls: if $m \geq C_{\alpha_0,\beta_0}\delta^{-4}s\ln(eN/s)$, then

$$\begin{aligned}
\mathbb{P}\left(\Big|\|\mathcal{S}(\mathbf{A}\mathbf{u})\|_1 - m\widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2)\Big| > \delta m\widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2) \text{ for some } \mathbf{u} \in \Sigma_s^{\mathrm{eff}} \text{ with } \alpha_0\mu \leq \sigma\|\mathbf{u}\|_2 \leq \beta_0\mu\right) \\
\leq \gamma\exp\left(-c_{\beta_0}\delta^2 m\right).
\end{aligned}$$

The proof of this statement — a reformulation of the statement of the lemma — is based on a covering argument, starting with the following estimation of the covering number of the set $\Sigma_s^{\mathrm{eff}} \cap B_2^N$ of effectively s-sparse vectors with ℓ_2-norm at most equal to one:

$$\mathcal{N}(\Sigma_s^{\mathrm{eff}} \cap B_2^N, \rho) \leq \mathcal{N}(\Sigma_t \cap B_2^N, \rho/4) \leq \left(\frac{eN}{t}\right)^t \left(1 + \frac{8}{\rho}\right)^t, \qquad t := \left\lceil \frac{4s}{\rho^2} \right\rceil.$$

We focus on justifying the first inequality, as the second inequality is classical. Let then $\widetilde{\mathbf{u}}_1, \ldots, \widetilde{\mathbf{u}}_n$ be the elements of a $(\rho/4)$-net of $\Sigma_t \cap B_2^N$ with $n = \mathcal{N}(\Sigma_t \cap B_2^N, \rho/4)$, and let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be their best approximations from $\Sigma_s^{\mathrm{eff}} \cap B_2^N$. Given $\mathbf{u} \in \Sigma_s^{\mathrm{eff}} \cap B_2^N$, we consider an index set T of t largest absolute entries of $\mathbf{u}$, and we choose $\widetilde{\mathbf{u}}_k$ such that $\|\mathbf{u}_T - \widetilde{\mathbf{u}}_k\|_2 \leq \rho/4$. Then

$$\begin{aligned}
\|\mathbf{u} - \mathbf{u}_k\|_2 &\leq \|\mathbf{u} - \widetilde{\mathbf{u}}_k\|_2 + \|\widetilde{\mathbf{u}}_k - \mathbf{u}_k\|_2 \leq 2\|\mathbf{u} - \widetilde{\mathbf{u}}_k\|_2 \leq 2\|\mathbf{u}_{\overline{T}}\|_2 + 2\|\mathbf{u}_T - \widetilde{\mathbf{u}}_k\|_2 \\
&\leq \frac{1}{\sqrt{t}}\|\mathbf{u}\|_1 + \frac{\rho}{2} \leq \sqrt{\frac{s}{4s/\rho^2}} + \frac{\rho}{2} = \rho, \tag{9}
\end{aligned}$$

where the first inequality in (9) followed from [7, Theorem 2.5]. This proves that $\mathbf{u}_1, \ldots, \mathbf{u}_n$ is a ρ-net for $\Sigma_s^{\mathrm{eff}} \cap B_2^N$, hence establishes the required inequality.

With $\omega := \beta_0\mu/\sigma$, we now place ourselves in the situation where the concentration inequality holds for all $\omega\mathbf{u}_k$, $k = 1, \ldots, n$, with the choices $\varepsilon = \delta/2$ and $\rho = [(5\alpha_0\eta_{\beta_0})/(22\beta_0)]\delta$. We also place ourselves in the situation where the property of Corollary 5 holds with $\delta = 1/4$. All of this occurs with failure probability bounded by

$$\begin{aligned} n \times 2 \exp\left(- \frac{\eta_{\beta_0}^2 \varepsilon^2 m}{2} \right) + \gamma_1 \exp\left(-c_1 m/256\right) &\leq \gamma \exp\left(-c'_{\beta_0} \delta^2 m + c''_{\alpha_0,\beta_0} \delta^{-2} s \ln\left(\frac{eN}{s}\right)\right) \\ &\leq \gamma \exp\left(-c_{\beta_0} \delta^2 m\right), \qquad (10) \end{aligned}$$

where the last inequality is a consequence of $m \geq C_{\alpha_0,\beta_0} \delta^{-4} s \ln(eN/s)$ with the constant C_{α_0,β_0} chosen large enough in comparison with $c''_{\alpha_0,\beta_0}/c'_{\beta_0}$ (and with C_1). Under these conditions, we shall prove that $|\|\mathcal{S}(\mathbf{A}\mathbf{u})\|_1 - m\widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2)| \leq \delta m \widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2)$ for all $\mathbf{u} \in \Sigma_s^{\rm eff}$ with $\alpha_0\mu \leq \sigma\|\mathbf{u}\|_2 \leq \beta_0\mu$. To do so, consider such a vector $\mathbf{u} \in \Sigma_s^{\rm eff}$ with $\alpha_0\mu/\sigma \leq \|\mathbf{u}\|_2 \leq \beta_0\mu/\sigma = \omega$ and chose $\mathbf{u}_k \in \Sigma_s^{\rm eff} \cap B_2^N$ such that $\|\omega^{-1}\mathbf{u} - \mathbf{u}_k\|_2 \leq \rho$, i.e., $\|\mathbf{u} - \omega\mathbf{u}_k\|_2 \leq \rho\omega$. Let us observe that

$$\begin{aligned} \left| \|\mathcal{S}(\mathbf{A}\mathbf{u})\|_1 - m\widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2) \right| &\leq \left| \|\mathcal{S}(\mathbf{A}(\omega\mathbf{u}_k))\|_1 - m\widetilde{\mathcal{S}}(\sigma\|\omega\mathbf{u}_k\|_2) \right| && (11) \\ &\quad + m \left| \widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2) - \widetilde{\mathcal{S}}(\sigma\|\omega\mathbf{u}_k\|_2) \right| && (12) \\ &\quad + \left| \|\mathcal{S}(\mathbf{A}\mathbf{u})\|_1 - \|\mathcal{S}(\mathbf{A}(\omega\mathbf{u}_k))\|_1 \right|. && (13) \end{aligned}$$

By the concentration inequalities, the right-hand side of (11) is bounded as

$$\left| \|\mathcal{S}(\mathbf{A}(\omega\mathbf{u}_k))\|_1 - m\widetilde{\mathcal{S}}(\sigma\|\omega\mathbf{u}_k\|_2) \right| \leq \varepsilon m \widetilde{\mathcal{S}}(\sigma\|\omega\mathbf{u}_k\|_2) \leq \varepsilon m \widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2) + \varepsilon m \left| \widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2) - \widetilde{\mathcal{S}}(\sigma\|\omega\mathbf{u}_k\|_2) \right|.$$

The latter terms combines with the term in (12), which, in view of $|\widetilde{\mathcal{S}}'(t)| \leq \sqrt{2/\pi}$, is bounded as

$$m \left| \widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2) - \widetilde{\mathcal{S}}(\sigma\|\omega\mathbf{u}_k\|_2) \right| \leq m \sqrt{\frac{2}{\pi}} \sigma \left| \|\mathbf{u}\|_2 - \|\omega\mathbf{u}_k\|_2 \right| \leq \frac{4}{5} m\sigma \|\mathbf{u} - \omega\mathbf{u}_k\|_2 \leq \frac{4}{5} m\sigma\rho\omega.$$

As for the term in (13), it reduces to

$$\begin{aligned} \left| \|\mathcal{S}(\mathbf{A}\mathbf{u})\|_1 - \|\mathcal{S}(\mathbf{A}(\omega\mathbf{u}_k))\|_1 \right| &= \left| \sum_{i=1}^m |\mathcal{S}(\langle \mathbf{a}_i, \mathbf{u}\rangle)| - |\mathcal{S}(\langle \mathbf{a}_i, \omega\mathbf{u}_k\rangle)| \right| \leq \sum_{i=1}^m |\mathcal{S}(\langle \mathbf{a}_i, \mathbf{u}\rangle) - \mathcal{S}(\langle \mathbf{a}_i, \omega\mathbf{u}_k\rangle)| \\ &\leq \sum_{i=1}^m |\langle \mathbf{a}_i, \mathbf{u} - \omega\mathbf{u}_k\rangle| = \|\mathbf{A}(\mathbf{u} - \omega\mathbf{u}_k)\|_1 \leq \left(1 + \frac{1}{4}\right)\sqrt{\frac{2}{\pi}} \sigma m \|\mathbf{u} - \omega\mathbf{u}_k\|_2 \\ &\leq m\sigma\rho\omega. \end{aligned}$$

Altogether, we obtain

$$\left| \|\mathcal{S}(\mathbf{A}\mathbf{u})\|_1 - m\widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2) \right| \leq \varepsilon m \widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2) + \left(\frac{9}{5} + \frac{4\varepsilon}{5}\right) m\sigma\rho\omega \leq \frac{\delta}{2} m \widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2) + \frac{11}{5}\rho m \beta_0 \mu.$$

We conclude by noticing that

$$\beta_0 \mu = \frac{\beta_0}{\alpha_0}\alpha_0\mu \leq \frac{\beta_0}{\alpha_0}\sigma\|\mathbf{u}\|_2 \leq \frac{\beta_0}{\alpha_0}\frac{1}{\eta_{\beta_0}}\widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2),$$

so that the choice $\rho := [(5\alpha_0\eta_{\beta_0})/(22\beta_0)]\delta$ yields the announced result that

$$|\|\mathcal{S}(\mathbf{A}\mathbf{u})\|_1 - m\widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2)| \leq \delta m \widetilde{\mathcal{S}}(\sigma\|\mathbf{u}\|_2).$$

This estimate is valid for all $\mathbf{u} \in \Sigma_s^{\rm eff}$ satisfying $\alpha_0\mu \leq \sigma\|\mathbf{u}\|_2 \leq \beta_0\mu$ with failure probability bounded as in (10). □

3.2 Main result

It is finally time to prove Part (b) of Theorem 1. In the formal statement below, $\Delta(\mathbf{y})$ again denotes the output of the recovery scheme (1).

Proposition 9. Let $\delta \in (0,1)$. Suppose that the random matrix $\mathbf{A} \in \mathbb{R}^{m\times N}$ has independent $\mathcal{N}(0,\sigma^2)$ entries. For any $\beta \geq \alpha$, where $\alpha > 0$ be the absolute constant from Proposition 6, there are constants $C' = C'(\beta), c' = c'(\beta) > 0$ depending only on β and an absolute constant $\gamma' > 0$ such that, if $m \geq C'\delta^{-4} s \ln(eN/s)$, then with probability at least $1 - \gamma' \exp(-c'\delta^4 m)$, one has

$$\|\mathbf{x} - \Delta(\mathcal{S}(\mathbf{A}\mathbf{x}))\|_2 \leq \delta \|\mathbf{x}\|_2$$

for all s-sparse $\mathbf{x} \in \mathbb{R}^N$ satisfying $\alpha\mu \leq \sigma\|\mathbf{x}\|_2 \leq \beta\mu$.

Proof. By possibly reducing η_β, we may assume that $\kappa_\beta := 4/\eta_\beta$ is an integer. Setting $\alpha_0 := \alpha/\kappa_\beta$, and $\beta_0 = 2\beta$, we place ourselves in the situation where

(i) the property of Corollary 5 holds for $\delta = \dfrac{1}{4}$ and s replaced by $\kappa_\beta^2 s$;

(ii) the property of Lemma 7 holds for $\delta = \delta_2 := \dfrac{\delta}{4}$ and s replaced by $\kappa_\beta^2 s$;

(iii) the property of Lemma 8 holds for $\delta = \delta_3 := \min\left\{\dfrac{\alpha\nu_{2\beta}\delta}{8+\alpha\nu_{2\beta}\delta}, \dfrac{1}{2}\right\}$ and s replaced by $\kappa_\beta^2 s$.

All of this occurs with failure probability bounded above by

$$\gamma_1 \exp(-c_1 m/256) + \gamma_2 \exp(-c_2\delta_2^4 m) + \gamma_3 \exp(-c_3(\beta_0)\delta_3^2 m) \leq \gamma' \exp(-c'(\beta)\delta^4 m),$$

provided that $m \geq \max\{4^{12}C_1, C_2\delta_2^{-4}, C_3(\alpha_0,\beta_0)\delta_3^{-4}\}\kappa_\beta^2 s \ln(eN/\kappa_\beta^2 s)$, which is guaranteed by the requirement $m \geq C'(\beta)\delta^{-4} s \ln(eN/s)$ for a sufficiently large constant $C'(\beta)$. Under these conditions, we shall prove that, for any s-sparse $\mathbf{x} \in \mathbb{R}^N$, we have

$$\|\mathbf{x} - \mathbf{x}^\sharp\|_2 \leq \delta\|\mathbf{x}\|_2, \qquad \text{where} \qquad \mathbf{x}^\sharp := \Delta(\mathcal{S}(\mathbf{A}\mathbf{x})). \tag{14}$$

For this purpose, we introduce the convex combination of $\mathbf{x}$ and $\mathbf{x}^\sharp$ defined by $\mathbf{x}^\flat = (1-\lambda)\mathbf{x} + \lambda\mathbf{x}^\sharp$, $\lambda := \min\{1, \|\mathbf{x}\|_2/\|\mathbf{x}^\sharp\|_2\}$, i.e.,

$$\mathbf{x}^\flat = \begin{cases} \mathbf{x}^\sharp, & \text{if } \|\mathbf{x}^\sharp\|_2 \leq \|\mathbf{x}\|_2, \\ \left(1 - \dfrac{\|\mathbf{x}\|_2}{\|\mathbf{x}^\sharp\|_2}\right)\mathbf{x} + \dfrac{\|\mathbf{x}\|_2}{\|\mathbf{x}^\sharp\|_2}\mathbf{x}^\sharp, & \text{if } \|\mathbf{x}^\sharp\|_2 > \|\mathbf{x}\|_2. \end{cases}$$

This auxiliary vector is introduced because $\|\mathbf{x}^\flat\|_2 \leq 2\|\mathbf{x}\|_2 \leq 2\beta\mu/\sigma$ is readily seen, while a similar bound for $\|\mathbf{x}^\sharp\|_2$ is not immediately obvious (although it is true as a consequence of the error bound (14) to be established). We now claim that it is enough to prove that

$$\|\mathbf{x} - \mathbf{x}^\flat\|_2 \leq \frac{\delta}{2}\|\mathbf{x}\|_2. \tag{15}$$

Indeed, in case $\|\mathbf{x}^\sharp\|_2 \le \|\mathbf{x}\|_2$, the inequality $\|\mathbf{x} - \mathbf{x}^\sharp\|_2 \le \delta \|\mathbf{x}\|_2$ is clear from the fact that $\mathbf{x}^\sharp = \mathbf{x}^\flat$, and in case $\|\mathbf{x}^\sharp\|_2 > \|\mathbf{x}\|_2$, rewriting (15) as $\|(\|\mathbf{x}\|_2/\|\mathbf{x}^\sharp\|_2)(\mathbf{x} - \mathbf{x}^\sharp)\|_2 \le (\delta/2)\|\mathbf{x}\|_2$ yields

$$\|\mathbf{x} - \mathbf{x}^\sharp\|_2 \le \frac{\delta}{2}\|\mathbf{x}^\sharp\|_2 \le \frac{\delta}{2}\|\mathbf{x}\|_2 + \frac{\delta}{2}\|\mathbf{x} - \mathbf{x}^\sharp\|_2, \qquad \text{hence} \qquad \|\mathbf{x} - \mathbf{x}^\sharp\|_2 \le \frac{\delta/2}{1-\delta/2}\|\mathbf{x}\|_2 \le \delta\|\mathbf{x}\|_2.$$

So it now remains to validate (15). First, we observe that $\mathbf{x}^\flat$ is effectively $(\kappa_\beta^2 s)$-sparse. Indeed, noticing that $\mathcal{S}(\mathbf{A}\mathbf{x}^\flat) = \mathcal{S}(\mathbf{A}\mathbf{x})$, we have, on the one hand,

$$\|\mathcal{S}(\mathbf{A}\mathbf{x}^\flat)\|_1 = \|\mathcal{S}(\mathbf{A}\mathbf{x})\|_1 \ge (1-\delta_3) m \widetilde{\mathcal{S}}(\sigma\|\mathbf{x}\|_2).$$

Since $\delta_3 \le 1/2$ and $\widetilde{\mathcal{S}}(\sigma\|\mathbf{x}\|_2) \ge \eta_\beta \sigma \|\mathbf{x}\|_2$, we derive

$$\|\mathcal{S}(\mathbf{A}\mathbf{x}^\flat)\|_1 \ge \frac{\eta_\beta}{2} m \sigma \|\mathbf{x}\|_2. \tag{16}$$

On the other hand, decomposing $[\![1:N]\!]$ into groups $T_0, T_1, T_2, \ldots$ of $t = \kappa_\beta^2 s$ indices according to decreasing magnitudes of the entries of $\mathbf{x}^\flat$, we have

$$\|\mathcal{S}(\mathbf{A}\mathbf{x}^\flat)\|_1 \le \|\mathbf{A}\mathbf{x}^\flat\|_1 = \left\| \mathbf{A}\left(\sum_{k\ge0} \mathbf{x}^\flat_{T_k}\right)\right\|_1 \le \sum_{k \ge 0} \left\| \mathbf{A}\mathbf{x}^\flat_{T_k}\right\|_1 \le \left(1 + \frac{1}{4}\right)\sqrt{\frac{2}{\pi}}\sigma m \sum_{k\ge0} \|\mathbf{x}^\flat_{T_k}\|_2.$$

Using $\|\mathbf{x}^\flat_{T_0}\|_2 \le \|\mathbf{x}^\flat\|_2$ and the classical inequality $\sum_{k\ge1}\|\mathbf{x}^\flat_{T_k}\|_2 \le \|\mathbf{x}^\flat\|_1/\sqrt{t}$, we obtain

$$\|\mathcal{S}(\mathbf{A}\mathbf{x}^\flat)\|_1 \le \sigma m \left[\|\mathbf{x}^\flat\|_2 + \sqrt{\frac{1}{t}}\|\mathbf{x}^\flat\|_1\right].$$

Taking into account that $\|\mathbf{x}^\sharp\|_1 \le \|\mathbf{x}\|_1$ to obtain

$$\|\mathbf{x}^\flat\|_1 \le (1-\lambda)\|\mathbf{x}\|_1 + \lambda\|\mathbf{x}^\sharp\|_1 \le \|\mathbf{x}\|_1,$$

we deduce from $\|\mathbf{x}\|_1 \le \sqrt{s}\|\mathbf{x}\|_2$ that

$$\|\mathcal{S}(\mathbf{A}\mathbf{x}^\flat)\|_1 \le \sigma m \left[\|\mathbf{x}^\flat\|_2 + \sqrt{\frac{s}{t}}\|\mathbf{x}\|_2\right] = \sigma m \left[\|\mathbf{x}^\flat\|_2 + \frac{1}{\kappa_\beta}\|\mathbf{x}\|_2\right]. \tag{17}$$

Combining (16) and (17) gives

$$\|\mathbf{x}^\flat\|_2 \ge \left[\frac{\eta_\beta}{2} - \frac{1}{\kappa_\beta}\right]\|\mathbf{x}\|_2 = \frac{1}{\kappa_\beta}\|\mathbf{x}\|_2,$$

where the last step followed from the choice of κ_β. In particular, we notice that

$$\frac{\|\mathbf{x}^\flat\|_1}{\|\mathbf{x}^\flat\|_2} \le \frac{\|\mathbf{x}\|_1}{(1/\kappa_\beta)\|\mathbf{x}\|_2} \le \kappa_\beta\sqrt{s},$$

which justifies our claim that $\mathbf{x}^\flat$ is effectively $(\kappa_\beta^2 s)$-sparse.

Next, we observe that the directions of $\mathbf{x}$ and $\mathbf{x}^\flat$ are close. To see this, we notice that both $\mathbf{x}$ and $\mathbf{x}^\flat$ are effectively $(\kappa_\beta^2 s)$-sparse and satisfy $\mathrm{sgn}(\mathbf{Ax}) = \mathrm{sgn}(\mathbf{Ax}^\flat)$ as a consequence of $\mathcal{S}(\mathbf{Ax}) = \mathcal{S}(\mathbf{Ax}^\flat)$, hence

$$\left\| \frac{\mathbf{x}}{\|\mathbf{x}\|_2} - \frac{\mathbf{x}^\flat}{\|\mathbf{x}^\flat\|_2} \right\|_2 \le \delta_2. \tag{18}$$

Finally, we prove that the magnitudes of $\mathbf{x}$ and $\mathbf{x}^\flat$ are close. Because both $\sigma\|\mathbf{x}\|_2$ and $\sigma\|\mathbf{x}^\flat\|_2$ are in the interval $[\alpha\mu/\kappa_\beta, 2\beta\mu] = [\alpha_0\mu, \beta_0\mu]$, we can invoke (iii) to obtain

$$\left| \|\mathcal{S}(\mathbf{Ax})\|_1 - m\widetilde{\mathcal{S}}(\sigma\|\mathbf{x}\|_2) \right| \le \delta_3 m \widetilde{\mathcal{S}}(\sigma\|\mathbf{x}\|_2),$$
$$\left| \|\mathcal{S}(\mathbf{Ax}^\flat)\|_1 - m\widetilde{\mathcal{S}}(\sigma\|\mathbf{x}^\flat\|_2) \right| \le \delta_3 m \widetilde{\mathcal{S}}(\sigma\|\mathbf{x}^\flat\|_2).$$

But since $\mathcal{S}(\mathbf{Ax}) = \mathcal{S}(\mathbf{Ax}^\flat)$, we derive that

$$m\left|\widetilde{\mathcal{S}}(\sigma\|\mathbf{x}\|_2) - \widetilde{\mathcal{S}}(\sigma\|\mathbf{x}^\flat\|_2)\right| \le \delta_3 m \left(\widetilde{\mathcal{S}}(\sigma\|\mathbf{x}\|_2) + \widetilde{\mathcal{S}}(\sigma\|\mathbf{x}^\flat\|_2)\right) \le \frac{2\delta_3}{1-\delta_3}\|\mathcal{S}(\mathbf{Ax})\|_1 \le \frac{2\delta_3}{1-\delta_3} m\mu.$$

Moreover, thanks to (8), we have

$$\left|\widetilde{\mathcal{S}}(\sigma\|\mathbf{x}\|_2) - \widetilde{\mathcal{S}}(\sigma\|\mathbf{x}^\flat\|_2)\right| \ge \nu_{\beta_0}|\sigma\|\mathbf{x}\|_2 - \sigma\|\mathbf{x}^\flat\|_2|.$$

We deduce that the magnitudes of $\mathbf{x}$ and $\mathbf{x}^\flat$ satisfy

$$|\|\mathbf{x}\|_2 - \|\mathbf{x}^\flat\|_2| \le \frac{2\delta_3}{1-\delta_3}\frac{\mu}{\nu_{2\beta}\sigma} \le \frac{2\delta_3}{1-\delta_3}\frac{\|\mathbf{x}\|_2}{\alpha\nu_{2\beta}}. \tag{19}$$

We now wrap up the argument by validating(15) from (18) and (19) as follows:

$$\begin{aligned} \|\mathbf{x} - \mathbf{x}^\flat\|_2 &= \left\| \|\mathbf{x}\|_2 \left(\frac{\mathbf{x}}{\|\mathbf{x}\|_2} - \frac{\mathbf{x}^\flat}{\|\mathbf{x}^\flat\|_2} \right) + \left(\|\mathbf{x}\|_2 - \|\mathbf{x}^\flat\|_2 \right) \frac{\mathbf{x}^\flat}{\|\mathbf{x}^\flat\|_2} \right\|_2 \\ &\le \|\mathbf{x}\|_2 \left\| \frac{\mathbf{x}}{\|\mathbf{x}\|_2} - \frac{\mathbf{x}^\flat}{\|\mathbf{x}^\flat\|_2} \right\|_2 + |\|\mathbf{x}\|_2 - \|\mathbf{x}^\flat\|_2| \\ &\le \delta_2\|\mathbf{x}\|_2 + \frac{2\delta_3}{1-\delta_3}\frac{1}{\alpha\nu_{2\beta}}\|\mathbf{x}\|_2 \le \frac{\delta}{2}\|\mathbf{x}\|_2, \end{aligned}$$

where the last equality results from $\delta_2 = \delta/4$ and $\delta_3 \le \alpha\nu_{2\beta}\delta/(8+\alpha\nu_{2\beta}\delta)$. The proof is complete. □

References

[1] CVX Research, Inc. CVX: MATLAB software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, 2014.

[2] A. Ai, A. Lapanowski, Y. Plan, and R. Vershynin. One-bit compressed sensing with non-Gaussian measurements. Linear Algebra and its Applications 441, 222–239, 2014.

[3] P. T. Boufounos and R. G. Baraniuk. 1-bit compressive sensing. In Proceedings of the 42nd Annual Conference on Information Sciences and Systems (CISS), pages 1621. IEEE, 2008.

[4] D. Bilyk and M. Lacey. Random tessellations, restricted isometric embeddings, and one bit sensing. arXiv preprint arXiv:1512.06697, 2015.

[5] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory 52(2), 489–509, 2006.

[6] D. Donoho. Compressed sensing. IEEE Transactions on Information Theory, 52(4), 1289–1306, 2006.

[7] S. Foucart and H. Rauhut. A Mathematical Introduction to Compressive Sensing. Birkhäuser, 2013.

[8] S. Foucart and M.-J. Lai. Sparse recovery with pre-Gaussian random matrices. Studia Mathematica, 200, 91–102, 2010.

[9] C. S. Güntürk, M. Lammers, A. M. Powell, R. Saab, and Ö Yılmaz. Sobolev duals for random frames and $\Sigma\Delta$ quantization of compressed sensing measurements. Foundations of Computational Mathematics, 13(1), 1–36, 2013.

[10] 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 Transactions on Information Theory 59(4), 2082–2102, 2013.

[11] L. Jacques, K. Degraux, and C. De Vleeschouwer. Quantized iterative hard thresholding: Bridging 1-bit and high-resolution quantized compressed sensing. arXiv preprint arXiv:1305.1786, 2013.

[12] Y. Plan and R. Vershynin. One-bit compressed sensing by linear programming. Communications on Pure and Applied Mathematics, 66(8), 1275–1297, 2013.

[13] Y. Plan and R. Vershynin. Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach. IEEE Transactions on Information Theory, 59(1), 482–494, 2013.

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