PERFORMANCE ANALYSIS OF ONE-BIT GROUP-SPARSE SIGNAL RECONSTRUCTION

Niklas Koep, Arash Behboodi, Rudolf Mathar

Institute for Theoretical Information Technology
RWTH Aachen University, Germany

ABSTRACT

We consider the reconstruction of group-sparse vectors from sign measurements of random projections. In particular, we establish conditions on the number of measurements under which such signal ensembles can be recovered up to a prescribed accuracy. The results rely on a mixed restricted isometry property as first employed by Foucart in the context of 1-bit compressed sensing, as well as certain results on random hyperplane tessellations. The paper fills a gap in the literature by establishing that group-sparse signals can be estimated from 1-bit observations with the same number of measurements as required for the reconstruction of block-sparse signals from unquantized measurements. We confirm the correct behavior of the recovery schemes in a series of numerical experiments.

***Index Terms*—** Compressed sensing, group-sparsity, group restricted isometry property, hard thresholding

1. INTRODUCTION

In various domains of science and engineering, one may be faced with the problem of estimating signals in a high-dimensional space from low-dimensional observations. Naturally, this task is generally ill-posed unless prior knowledge is imposed on the signals of interest. The theory of *compressed sensing* (CS) makes precise the conditions under which such an undertaking is feasible [1, 2, 3]. The most recent developments in this direction extend the ideas of sparse recovery into the setting of nonlinear observations [4, 5] which includes modeling of saturation effects [6] and *analog-to-digital converters* (ADCs) [7]. A particularly attractive paradigm which emerged in this context is that of 1-bit compressed sensing [8] in which an acquisition system only retains the sign information of each projected coordinate. While most works in this direction such as [9, 10] focus their attention on the reconstruction of sparse vectors, only a limited number of works address other low-complexity structures [11]. For instance, in [12], the first and last named author already investigated the reconstruction of block-sparse signals from binary observations with a flat coefficient structure. In the present work, this empirical work is complemented by a theoretical analysis of related recovery schemes in the context of group-sparse signal recovery.

Notation

Throughout the paper, we denote vectors and matrices by lower- and uppercase boldface letters, respectively. Given two vectors $\mathbf{x}, \mathbf{z} \in \mathbb{R}^d$, we denote by $\mathbf{x} \circ \mathbf{z}$ the Hadamard product with $(\mathbf{x} \circ \mathbf{z})_i = x_i z_i$. For two nonnegative numbers $a, b \in \mathbb{R}$, we write $a \lesssim b$ if there is an absolute constant $C > 0$ such that $a \leq Cb$.

This work was partially supported by the Deutsche Forschungsgemeinschaft (DFG) project QuaCoSS (grant MA 1184/27-1).

2. SIGNAL AND ACQUISITION MODEL

We consider the task of reconstructing group-sparse vectors from 1-bit quantized random projections. In particular, we assume that the coefficients of d-dimensional vectors are grouped into G nonoverlapping groups according to the following definition.

Definition 2.1 (Group partition)**.** *A collection $\mathcal{I} = \{\mathcal{I}_1, \ldots, \mathcal{I}_G\}$ of subsets $\mathcal{I}_i \subseteq [d] := \{1, \ldots, d\}$ is called a group partition of $[d]$ if $\mathcal{I}_i \cap \mathcal{I}_j = \emptyset \;\forall i \neq j$, and $\bigcup_{i=1}^{G} \mathcal{I}_i = [d]$.*

Note that this definition does not assume that the elements in $\mathcal{I}_i$ are consecutive indices, nor that the cardinality of the individual sets are identical. For convenience, we denote the cardinality of the largest group by $g = \max_{i \in [G]} |\mathcal{I}_i|$. Denote by $\mathbf{x}_{\mathcal{I}_i} \in \mathbb{R}^d$ the restriction of $\mathbf{x}$ to the indices in $\mathcal{I}_i$, *i.e.*, $(\mathbf{x}_{\mathcal{I}_i})_j = x_j \cdot \mathbb{1}_{\{j \in \mathcal{I}_i\}}$ for $j \in [d]$ where $\mathbb{1}_{\{\cdot\}}$ denotes the binary indicator function. Then a signal $\mathbf{x}$ is called s-group-sparse (w. r. t. the group partition $\mathcal{I}$) if it is supported on at most s groups. Given a group partition $\mathcal{I}$, we naturally associate with it the following family of mixed norms.

Definition 2.2 (Group ℓ_p-norms)**.** *Let $\mathbf{x} \in \mathbb{R}^d$. Then, for $p \geq 1$, the group ℓ_p-norm on $\mathbb{R}^d$ is defined as $\|\mathbf{x}\|_{\mathcal{I},p} := \left(\sum_{i=1}^{G} \|\mathbf{x}_{\mathcal{I}_i}\|_2^p\right)^{1/p}$.*

We extend the notation $\|\cdot\|_{\mathcal{I},p}$ to $p = 0$ in which case we define the pseudonorm $\|\mathbf{x}\|_{\mathcal{I},0} := |\{i \in [G] : \mathbf{x}_{\mathcal{I}_i} \neq \mathbf{0}\}|$ which counts the number of groups a vector is supported on. With this definition in place, we define the set

$$\Sigma_{\mathcal{I},s} = \{\mathbf{x} \in \mathbb{S}^{d-1} : \|\mathbf{x}\|_{\mathcal{I},0} \leq s\}$$

of s-group-sparse vectors on the sphere w. r. t. the group partition $\mathcal{I}$. We will also make frequent use of a slightly more general signal set, namely the so-called effectively s-group-sparse vectors defined as

$$\mathcal{E}_{\mathcal{I},s} = \{\mathbf{x} \in \mathbb{S}^{d-1} : \|\mathbf{x}\|_{\mathcal{I},1} \leq \sqrt{s}\}.$$

Note that we have by the Cauchy-Schwarz inequality that every s-group-sparse vector is naturally effectively s-group-sparse, and hence $\Sigma_{\mathcal{I},s} \subset \mathcal{E}_{\mathcal{I},s}$ (see also Lem. 3.4 below).

Given an (effectively) group-sparse vector $\mathring{\mathbf{x}}$, we now consider measurements of the form $\mathbf{y} = \operatorname{sgn}(\mathbf{A}\mathring{\mathbf{x}}) \in \{\pm 1\}^m$ where the sign operator sgn is applied component-wise, and we use the convention $\operatorname{sgn}(a) = -1$ if $a < 0$, and $\operatorname{sgn}(a) = 1$ if $a \geq 0$. Clearly, any norm information about $\mathring{\mathbf{x}}$ is lost in this acquisition model as sgn is invariant under positive scaling of its argument. The task we are concerned with in this paper is therefore the estimation of $\mathring{\mathbf{x}}$ up to normalization from knowledge of the measurement matrix $\mathbf{A}$, the quantized projections $\mathbf{y} \in \{\pm 1\}^m$, as well as possibly prior knowledge about the (effective) group-sparsity level s of $\mathring{\mathbf{x}}$.

978-1-5386-4658-8/18/$31.00 ©2019 IEEE

3. GROUP-SPARSE SIGNAL RECOVERY

In this section, we analyze the performance of three recovery schemes adopted for the task of group-sparse signal reconstruction. We begin with the analysis of two approaches based on convex programming.

3.1. Convex Programming

Consider the following optimization problem adopted from [10]:

$$\underset{\mathbf{x}}{\text{minimize}} \ \ \|\mathbf{x}\|_{\mathcal{I},1} \ \text{ s.t. } \ \mathbf{y} = \operatorname{sgn}(\mathbf{A}\mathbf{x}), \|\mathbf{A}\mathbf{x}\|_1 = 1. \qquad (\mathrm{P}_1)$$

Note that we can eliminate the sgn operator, and replace the first constraint by $\mathbf{y} \circ \mathbf{A}\mathbf{x} \geq \mathbf{0}$ without changing the solution of the problem since the second constraint removes the zero vector from the feasible set. Similarly, the second constraint can be replaced by the linear constraint $\langle \mathbf{y}, \mathbf{A}\mathbf{x} \rangle = 1$ as long as $\mathbf{x}$ satisfies the first constraint. In other words, due to the nature of the objective function, Prob. (P_1) admits a representation as a *second-order cone program* (SOCP). For convenience of notation, we will also denote the output of Prob. (P_1) by $\Delta_{\mathrm{CP},\mathcal{I}}(\mathbf{y})$. The analysis of this recovery strategy relies on two properties of the measurement matrix $\mathbf{A}$. The first property we require is a result which states that $\mathbf{A}$ uniformly tessellates subsets of the unit Euclidean sphere according to the following definition.

Definition 3.1 (ε-Tessellation). *A matrix $\mathbf{A} \in \mathbb{R}^{m \times d}$ is said to induce an ε-tessellation on a set $\mathcal{K} \subset \mathbb{S}^{d-1}$ if*

$$\forall \mathbf{x}, \mathbf{z} \in \mathcal{K} \text{ with } \operatorname{sgn}(\mathbf{A}\mathbf{x}) = \operatorname{sgn}(\mathbf{A}\mathbf{z}) : \|\mathbf{x} - \mathbf{z}\|_2 \leq \varepsilon.$$

Since measurement consistency is enforced by the first constraint of Prob. (P_1), the definition of ε-tessellations immediately yields a recovery guarantee for Prob. (P_1). More precisely, if $\mathbf{A}$ induces an ε-tessellation on $\mathcal{E}_{\mathcal{I},s}$, then we have that $\mathring{\mathbf{x}} \in \mathcal{E}_{\mathcal{I},s}$ and ℓ_2-normalized minimizers $\hat{\mathbf{z}}$ of Prob. (P_1) are at most ε apart (in the Euclidean sense), provided that $\hat{\mathbf{z}}$ is effectively group-sparse. This in turn can be established by imposing that $\mathbf{A}$ satisfies the following generalized *restricted isometry property* (RIP).

Definition 3.2 (Group-RIP). *Let $\mathbf{A} \in \mathbb{R}^{m \times d}$. Then $\mathbf{A}$ satisfies the* (ℓ_2, ℓ_1) group restricted isometry property *(group-RIP) of order s with constant $\delta_s \in (0,1)$ if*

$$(1 - \delta_s)\|\mathbf{x}\|_2 \leq \|\mathbf{A}\mathbf{x}\|_1 \leq (1 + \delta_s)\|\mathbf{x}\|_2 \ \forall \mathbf{x} \in \Sigma_{\mathcal{I},s}. \qquad (1)$$

If Eq. (1) *also holds for $\mathcal{E}_{\mathcal{I},s}$, we say that $\mathbf{A}$ satisfies the effective (ℓ_2, ℓ_1) group-RIP.*

Note that verifying either of these properties for a given matrix $\mathbf{A}$ is generally intractable [13]. As is customary in the compressed sensing literature, we will instead appeal to random measurement ensembles, and rely on high-probability bounds to establish the necessary properties. In particular, we assume from here on that $\mathbf{A}$ is populated by independent copies of a standard Gaussian random variable. Before turning our attention to the question when Def. 3.1 and 3.2 hold with high probability, we first record the following statement about the effective group-sparsity of minimizers of Prob. (P_1).

Lemma 3.1. *Let $\mathring{\mathbf{x}} \in \mathcal{E}_{\mathcal{I},s}$, and assume $\mathbf{A} \in \mathbb{R}^{m \times d}$ satisfies the effective (ℓ_2, ℓ_1) group-RIP with constant δ_t of order $t = 4s((1+\delta_t)/(1-\delta_t))^2$. Then with $\mathbf{y} = \operatorname{sgn}(\mathbf{A}\mathring{\mathbf{x}})$, every minimizer $\hat{\mathbf{x}} = \Delta_{\mathrm{CP},\mathcal{I}}(\mathbf{y})$ of Prob.* (P_1) *is effectively t-group-sparse.*

Due to space limitations, we skip the proof details here, and only remark that the proof proceeds in the same way as that of Lem. 4 in [14] with a few minor modifications to account for the group-sparsity structure. Moreover, instead of fixing a value for the group-RIP constant δ_t, we opt to treat it as a free parameter instead, and merely impose that minimizers $\hat{\mathbf{x}}$ are effectively t-group-sparse, yielding the condition $t = 4s((1+\delta_t)/(1-\delta_t))^2$ in Lem. 3.1. The next result now follows immediately from Lem. 3.1 and the definition of ε-tessellations.

Lemma 3.2. *Fix $\delta_t \in (0,1)$, and assume $\mathbf{A} \in \mathbb{R}^{m \times d}$ satisfies the group-RIP on $\mathcal{E}_{t,s}$ with $t = 4s((1+\delta_t)/(1-\delta_t))^2$. Moreover, assume that $\mathbf{A}$ induces an ε-tessellation on $\mathcal{E}_{t,s}$. Then for every $\mathring{\mathbf{x}} \in \mathcal{E}_{\mathcal{I},s}$, it holds with $\mathbf{y} = \operatorname{sgn}(\mathbf{A}\mathring{\mathbf{x}})$ that every normalized minimizer $\hat{\mathbf{z}} = \hat{\mathbf{x}}/\|\hat{\mathbf{x}}\|_2$ of Prob.* (P_1) *satisfies $\|\mathring{\mathbf{x}} - \hat{\mathbf{z}}\|_2 \leq \varepsilon$.*

We now turn to establishing that the (effective) group-RIP holds with high probability. To that end, we first require the definition of the so-called (Gaussian) mean width

$$w(\mathcal{K}) = \mathbb{E}_{\mathbf{g}} \sup_{\mathbf{x} \in \mathcal{K}} \langle \mathbf{g}, \mathbf{x} \rangle \qquad (2)$$

where $\mathbf{g} \in \mathbb{R}^d$ denotes a standard Gaussian random vector. Note that by standard results in convex analysis, this definition immediately implies that w is invariant under application of the convex hull, *i.e.*,

$$w(\mathcal{K}) = w(\operatorname{conv}(\mathcal{K})). \qquad (3)$$

We will use the following result due to Plan and Vershynin which is a simplified version of Lem. 2.1 in [15] to assert the (effective) group-RIP of $\mathbf{A}$.

Theorem 3.1. *Let $\mathbf{A} \in \mathbb{R}^{m \times d}$ be a standard Gaussian random matrix, and let $\mathcal{K} \subset \mathbb{S}^{d-1}$. Set $\tilde{\mathbf{A}} = m^{-1}\sqrt{\pi/2}\mathbf{A}$, and fix $\delta \in (0,1)$. Then it holds with probability at least $1 - \eta$ that $\sup_{\mathbf{x} \in \mathcal{K}} |\|\tilde{\mathbf{A}}\mathbf{x}\|_1 - 1| \leq \delta$ provided $m \gtrsim \delta^{-2}(w(\mathcal{K})^2 + \log(\eta^{-1}))$.*

In order to apply Thm. 3.1, it is necessary to estimate the mean width of $\Sigma_{\mathcal{I},s}$. This is the content of the following statement.

Lemma 3.3. *It holds that $w(\Sigma_{\mathcal{I},s}) \leq \sqrt{2s\log(2eG/s)} + \sqrt{sg}$.*

Proof. First, note that

$$\begin{aligned} w(\Sigma_{\mathcal{I},s}) &= \mathbb{E}_{\mathbf{g}} \sup_{\mathbf{x} \in \Sigma_{\mathcal{I},s}} \langle \mathbf{x}, \mathbf{g} \rangle = \mathbb{E} \max_{\mathcal{T}} \sup_{\mathbf{x} \in \mathbb{S}^{d-1}_{\mathcal{T}}} \langle \mathbf{x}, \mathbf{g} \rangle \\ &= \mathbb{E} \max_{\mathcal{T}} \|\mathbf{g}_{\mathcal{T}}\|_2 \\ &\leq \max_{\mathcal{T}} \mathbb{E}\|\mathbf{g}_{\mathcal{T}}\|_2 + \mathbb{E} \max_{\mathcal{T}} \left| \|\mathbf{g}_{\mathcal{T}}\|_2 - \mathbb{E}\|\mathbf{g}_{\mathcal{T}}\|_2 \right| \end{aligned}$$

where the subsequent maxima are taken over all possible group index sets $\mathcal{T} \subset \mathcal{I}$ with $|\mathcal{T}| = s$, and we write $\mathbf{g}_{\mathcal{T}}$ for the restriction of $\mathbf{g}$ to $\bigcup_{S \in \mathcal{T}} S$. For the first term on the r.h.s., we have by Jensen's inequality that $\max_{\mathcal{T}} \mathbb{E}\|\mathbf{g}_{\mathcal{T}}\|_2 \leq \sqrt{sg}$. For the second term, note that $\|\cdot\|_2$ is 1-Lipschitz continuous by definition. The Gaussian concentration inequality therefore shows that the centered random variable $X_{\mathcal{T}} := \|\mathbf{g}_{\mathcal{T}}\|_2 - \mathbb{E}\|\mathbf{g}_{\mathcal{T}}\|_2$ is subgaussian since by [16, Thm. 5.5], we have for $\lambda \in \mathbb{R}$ that $\mathbb{E}\exp(\lambda X_{\mathcal{T}}) \leq \exp(\lambda^2/2)$. By a common bound on the expected maximum of a sequence of independent subgaussian random variables (see, *e.g.*, [17, Prop. 7.29]), this implies $\mathbb{E}\max_{\mathcal{T}} |X_{\mathcal{T}}| \leq \sqrt{2\log\left(2\binom{G}{s}\right)} \leq \sqrt{2s\log(2eG/s)}$ which yields the announced result. ☐

The next result, whose proof follows that of Lem. 3.1 in [10], and is skipped here in the interest of space, now allows us to bound the mean width of $\mathcal{E}_{\mathcal{I},s}$.

Lemma 3.4. *It holds that* $\operatorname{conv}(\Sigma_{\mathcal{I},s}) \subset \mathcal{E}_{\mathcal{I},s} \subset 2\operatorname{conv}(\Sigma_{\mathcal{I},s})$.

By the invariance property (3) of w, as well as homogeneity of the mean width according to Eq. (2), Lem. 3.4 then immediately yields

$$w(\mathcal{E}_{\mathcal{I},s}) \leq 2w(\Sigma_{\mathcal{I},s}). \tag{4}$$

Next, we turn to the ε-tessellation property of $\mathbf{A}$. The probability that $\mathbf{A}$ induces an ε-tessellation on $\mathcal{E}_{\mathcal{I},s}$ will be fully determined by the covering number $\mathfrak{N}$ of $\mathcal{E}_{\mathcal{I},s}$ w. r. t. the Euclidean metric, *i.e.*, the cardinality of the smallest subset $\mathcal{N}$ of $\mathcal{E}_{\mathcal{I},s}$ such that every point in $\mathcal{E}_{\mathcal{I},s}$ falls within an ε-distance of at least one point in $\mathcal{N}$. A bound on the covering number of $\mathcal{E}_{\mathcal{I},s}$ will in turn follow from the previous estimate on $w(\mathcal{E}_{\mathcal{I},s})$ via Sudakov's inequality. The following result due to Bilyk and Lacey establishes a high-probability bound for Gaussian random matrices to induce ε-tessellations on subsets of the unit Euclidean sphere.

Theorem 3.2 ([18, Thm. 1.5])**.** *Let* $\mathbf{A} \in \mathbb{R}^{m\times d}$ *be a standard Gaussian random matrix. Then there exist constants* $0 < c < 1 < C$ *such that with probability at least* $1 - (2\mathfrak{N}(\mathcal{K}, \|\cdot\|_2, c\varepsilon))^{-2}$*, the matrix* $\mathbf{A}$ *induces an* ε*-tessellation on the set* $\mathcal{K} \subset \mathbb{S}^{d-1}$ *provided that* $m \geq C\varepsilon^{-1}\log\mathfrak{N}(\mathcal{K}, \|\cdot\|_2, c\varepsilon)$.

We are now ready to state our first main result which establishes a recovery guarantee for Prob. (P_1).

Theorem 3.3. *Let* $\mathbf{A} \in \mathbb{R}^{m\times d}$ *be a standard Gaussian random matrix, and assume* $\varepsilon \leq 1/3$*. Then the conclusion of Lem. 3.2 holds with probability at least*

$$1 - c\exp\left(-\tilde{c}\big(s\log\left(\tfrac{G}{16s}\right) + sg\big)\right)$$

provided that $m \gtrsim \varepsilon^{-3}(s\log(G/s) + sg)$.

Proof. According to Thm. 3.1, one requires $m \gtrsim \delta_t^{-2} w(\mathcal{E}_{t,s})^2$ measurements for $\mathbf{A}$ to satisfy the group-RIP on $\mathcal{E}_{t,s}$ with failure probability at most $\exp(-w(\mathcal{E}_{t,s})^2)$. On the other hand, Sudakov's inequality (see, *e.g.*, [19, Thm. 8.1.13]) yields

$$\log\mathfrak{N}(\mathcal{E}_{t,s}, \|\cdot\|_2, \varepsilon) \lesssim \varepsilon^{-2} w(\mathcal{E}_{t,s})^2.$$

Thm. 3.2 therefore establishes that $m \gtrsim \varepsilon^{-3} w(\mathcal{E}_{t,s})^2$ measurements suffice for $\mathbf{A}$ to induce an ε-tessellation on $\mathcal{E}_{t,s}$ with failure probability at most $c_0\exp(-c_1\varepsilon^{-2}w(\mathcal{E}_{t,s})^2)$. With the choice $\delta_t = \varepsilon$, we have by the union bound that $\mathbf{A}$ satisfies both properties with failure probability at most $c_0'\exp(-c_1' w(\mathcal{E}_{t,s})^2)$. Define now $\gamma_\varepsilon = ((1+\varepsilon)/(1-\varepsilon))^2$. Invoking Eq. (4) in combination with Lem. 3.3 therefore yields

$$m \gtrsim \varepsilon^{-3}\gamma_\varepsilon(s\log(G/s) + sg).$$

With our restriction $\varepsilon \leq 1/3$, we have $\gamma_\varepsilon \leq 4$ which in turn yields the desired scaling for m after absorbing the constant into the notation. Moreover, substituting the bound for $w(\mathcal{E}_{t,s})$ into the last estimate of the failure probability yields $c\exp(-\tilde{c}\gamma_\varepsilon(s\log(\frac{G}{4s\gamma_\varepsilon}) + sg))$. Using that $1 \leq \gamma_\varepsilon \leq 4$ concludes the proof. □

Remark 3.1. In order for the failure probability in the proof of Thm. 3.3 to yield sensible values, it is necessary for $\log(\frac{G}{4s\gamma_\varepsilon}) + g$ to remain positive. This yields the very mild condition $\varepsilon \leq 1 - 2\big(\exp\left(\frac{1}{2}\left[\log\left(\frac{G}{4s}\right) + g\right]\right) + 1\big)^{-1}$.

Correlation Maximization

We may also appeal to an earlier result due to Plan and Vershynin [20] which establishes a nonuniform recovery guarantee for a fixed signal $\mathring{\mathbf{x}} \in \mathbb{R}^d$. Adopted to the setting of group-sparse recovery, we suggest to solve the program

$$\underset{\mathbf{x}}{\text{maximize}} \ \langle \mathbf{y}, \mathbf{A}\mathbf{x}\rangle \ \text{ s.t. } \|\mathbf{x}\|_{\mathcal{I},1} \leq \sqrt{s}, \|\mathbf{x}\|_2 \leq 1 \tag{P_2}$$

which maximizes the correlation between quantized and unquantized observations over the set of effectively s-group-sparse vectors. For notational convenience, we also write $\Delta_{\mathrm{corr},\mathcal{I}}(\mathbf{y})$ to denote minimizers of Prob. (P_2). The performance of this recovery scheme is again determined by the mean width of the signal set $\mathcal{E}_{\mathcal{I},s}$. The following result is a simple consequence of Thm. 1.1 in [20], as well as Lem. 3.4 in combination with Lem. 3.3.

Theorem 3.4. *Let* $\mathring{\mathbf{x}} \in \mathcal{E}_{\mathcal{I},s}$*, and* $\mathbf{y} = \operatorname{sgn}(\mathbf{A}\mathring{\mathbf{x}})$ *with* $\mathbf{A} \in \mathbb{R}^{m\times d}$ *denoting a standard Gaussian random matrix. Then with probability at least* $1-\eta$*, every normalized minimizer* $\hat{\mathbf{z}} = \hat{\mathbf{x}}/\|\hat{\mathbf{x}}\|_2$ *of Prob.* (P_2) *with* $\hat{\mathbf{x}} = \Delta_{\mathrm{corr},\mathcal{I}}(\mathbf{y})$ *satisfies* $\|\mathring{\mathbf{x}} - \hat{\mathbf{z}}\|_2 \leq \varepsilon$ *provided*

$$m \gtrsim \varepsilon^{-4}\big(s\log(G/s) + sg + \log(\eta^{-1})\big).$$

Remark 3.2. One can translate Thm. 3.4 into a form which holds uniformly over the entire signal set at the expense of an additional log-factor, as well as worse powers of ε^{-1} required for m (cf. [20, Thm. 1.3]).

3.2. Group Hard Thresholding

In addition to the convex programs considered in the previous sections, we also consider a simple recovery procedure adopted from [14] to the case of group-sparse signal reconstruction which does not rely on convex programming. Given a vector $\mathring{\mathbf{x}} \in \mathbb{R}^d$, and its quantized projections $\mathbf{y} = \operatorname{sgn}(\mathbf{A}\mathring{\mathbf{x}})$, we define the recovery map

$$\Delta_{\mathrm{HT},\mathcal{I}}(\mathbf{y}) = \underset{\mathbf{x}\in\Sigma_{\mathcal{I},s}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{A}^\top\mathbf{y}\|_{\mathcal{I},1} = \mathcal{H}_{\mathcal{I},s}(\mathbf{A}^\top\mathbf{y})$$

where $\mathcal{H}_{\mathcal{I},s} \colon \mathbb{R}^d \to \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_{\mathcal{I},0} \leq s\}$ denotes the so-called *group hard thresholding operator* which only retains the s groups of a vector with largest ℓ_2-norm. The recovery performance of this procedure is summarized in the following result which was first established by Foucart for the case of sparse vectors. The proof proceeds in the same way as the proof of Thm. 8 in [14] as soon as one replaces the sets S and T by the respective group index sets which $\mathring{\mathbf{x}}$ and $\hat{\mathbf{x}}$ are supported on.

Lemma 3.5. *Assume* $\mathbf{A} \in \mathbb{R}^{m\times d}$ *satisfies the group-RIP of order* $2s$ *with constant* δ_{2s}*. Then for any* $\mathring{\mathbf{x}} \in \Sigma_{\mathcal{I},s}$*, we have with* $\hat{\mathbf{x}} = \Delta_{\mathrm{HT},\mathcal{I}}(\operatorname{sgn}(\mathbf{A}\mathring{\mathbf{x}}))$*, and* $\hat{\mathbf{z}} = \hat{\mathbf{x}}/\|\hat{\mathbf{x}}\|_2$ *that* $\|\mathring{\mathbf{x}} - \hat{\mathbf{z}}\|_2 \leq 4\sqrt{5}\sqrt{\delta_{2s}}$.

Translating the error bound above into the form $\|\mathring{\mathbf{x}} - \hat{\mathbf{z}}\|_2 \leq \varepsilon$, and invoking Thm. 3.1 with Lem. 3.3, this immediately yields the following result.

Theorem 3.5. *Let* $\mathbf{A} \in \mathbb{R}^{m\times d}$ *be a standard Gaussian random matrix with* $m \gtrsim \varepsilon^{-4}(s\log(G/s) + sg + \log(\eta^{-1}))$*. Then with probability at least* $1-\eta$*, any* $\mathring{\mathbf{x}} \in \Sigma_{\mathcal{I},s}$ *is approximated by* $\hat{\mathbf{z}} = \hat{\mathbf{x}}/\|\hat{\mathbf{x}}\|$ *with* $\hat{\mathbf{x}} = \Delta_{\mathrm{HT},\mathcal{I}}(\operatorname{sgn}(\mathbf{A}\mathring{\mathbf{x}}))$ *such that* $\|\mathring{\mathbf{x}} - \hat{\mathbf{z}}\|_2 \leq \varepsilon$.

Remark 3.3. The dependence of m on ε in Thm. 3.5 is the same as in Thm. 3.4. Note, however, that the statement of Thm. 3.5 holds uniformly over the entire signal ensemble $\Sigma_{\mathcal{I},s}$ while Thm. 3.4 only holds for fixed vectors.

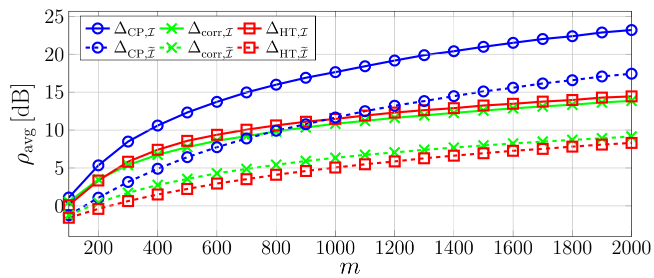

Fig. 1. SNR vs. number of measurements. The dashed lines represent the performance when the group-sparsity structure is ignored.

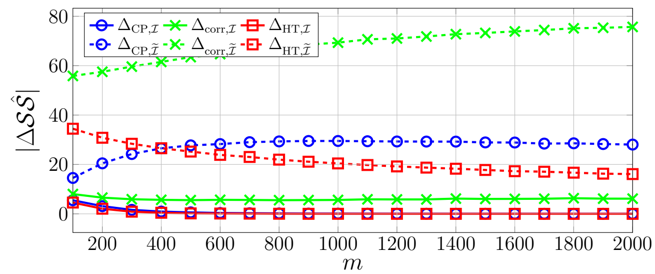

Fig. 2. Support estimation error. Dashed lines correspond to the performance w. r. t. $\widetilde{\mathcal{I}} = \{\{1\},\ldots,\{d\}\}$.

Unlike in Lem. 3.2, the recovery quality in Lem. 3.5 above is only dependent on the group-RIP constant. While such a result can also be established for Lem. 3.2 by only appealing to the group-RIP of $\mathbf{A}$, establishing a guarantee of the form $\|\mathring{\mathbf{x}} - \hat{\mathbf{z}}\|_2 \leq C\sqrt{\delta_t}$ requires one to restrict δ_t to the interval $(0, 1/5]$ (cf. [14, Thm. 9]). Moreover, instead of relying on the group-RIP, it is also possible to appeal to a slightly more general notion of restricted isometries known as *sign product embeddings* (SPEs) as coined in [21] to establish Thm. 3.5. While the SPE on $\Sigma_{\mathcal{I},s}$ immediately implies the group-RIP, the best known bound requires m to scale with δ_t^{-6} rather than δ_t^{-2} as required by Thm. 3.1 (see [22, Sec. 2] for details).

4. NUMERICAL EVALUATION

In this section, we conduct an empirical performance analysis of the proposed recovery strategies. Throughout, we set $d = 1000$, and split the support set $[d]$ into $G = 100$ nonoverlapping groups. Moreover, we consider $s = 5$ active groups chosen uniformly at random such that each realization contains $g \cdot s = 10 \cdot 5 = 50$ nonzero coefficients with each individual entry drawn i.i.d. from the standard Gaussian distribution. Finally, we project each vector on $\mathbb{S}^{d-1}$ by normalizing it. For each parameter combination, we run 1000 Monte Carlo trials. We compare the three presented group-sparse recovery methods with their sparse counterparts which we obtain by replacing $\mathcal{I}$ with $\widetilde{\mathcal{I}} := \{\{1\},\ldots,\{d\}\}$. Note that for $\Delta_{\mathrm{corr},\widetilde{\mathcal{I}}}$ and $\Delta_{\mathrm{HT},\widetilde{\mathcal{I}}}$, we provide each recovery method with the total sparsity level $s \cdot g$.

4.1. Reconstruction Performance

In the first experiment, we consider the recovery performance in terms of the average *signal-to-noise ratio* (SNR) according to $\rho_{\mathrm{avg}}\,[\mathrm{dB}] = -20\log_{10}\|\mathring{\mathbf{x}} - \hat{\mathbf{x}}/\|\hat{\mathbf{x}}\|_2\|_2$ where $\hat{\mathbf{x}}$ denotes the output of some recovery method. The results are shown in Fig. 1. Unsurprisingly, each method's performance improves with increasing m with $\Delta_{\mathrm{CP},\mathcal{I}}$ clearly outperforming its competitors. More surprisingly, however, is the fact $\Delta_{\mathrm{CP},\widetilde{\mathcal{I}}}$ starts to outperform both the hard thresholding approach and $\Delta_{\mathrm{corr},\mathcal{I}}$ for $m \geq 1000$. In general, the experiments confirm the relation between the different scaling behaviors required for each method as established in Thm. 3.3, 3.4 and 3.5. More precisely, the dependence on ε is most favorable for $\Delta_{\mathrm{CP},\mathcal{I}}$, while $\Delta_{\mathrm{corr},\mathcal{I}}$ and $\Delta_{\mathrm{HT},\mathcal{I}}$ fall slightly behind due to their dependence on ε^{-4} rather than ε^{-3}.

4.2. Group Support Identification

In many practical applications, rather than aiming at recovering the individual components of a vector, one might instead only be interested in identifying the active groups. We conduct a simple experiment to gauge how well the individual recovery schemes fare in this context. For simplicity, we restrict our attention to genuinely group-sparse vectors again. To investigate how well the individual reconstruction methods manage to identify the group support set $\mathcal{S} = \mathrm{supp}_{\mathcal{I}}(\mathbf{x}) = \{i \in [G] : \mathbf{x}_{\mathcal{I}_i} \neq \mathbf{0}\}$, we consider the symmetric set difference $\mathcal{S}\Delta\hat{\mathcal{S}} = (\mathcal{S} \setminus \hat{\mathcal{S}}) \cup (\hat{\mathcal{S}} \setminus \mathcal{S})$ between the ground truth $\mathcal{S}$ and estimated index set $\hat{\mathcal{S}} \subset [G]$, respectively. Given an estimator $\hat{\mathbf{x}}$, denote by $\mathcal{I}^* = \{\mathcal{I}_1^*,\ldots,\mathcal{I}_G^*\}$ the nonincreasing group rearrangement of $\hat{\mathbf{x}}$ such that $\|\hat{\mathbf{x}}_{\mathcal{I}_1^*}\|_2 \geq \ldots \geq \|\hat{\mathbf{x}}_{\mathcal{I}_G^*}\|_2$ with $\mathcal{I}_i^* = \mathcal{I}_{\pi(i)}$, and $\pi\colon [G] \to [G]$ a permutation. Next, define the group index set $\mathcal{J}^{(n)} = \{\mathcal{I}_1^*,\ldots,\mathcal{I}_n^*\}$. With slight abuse of notation, we write $\hat{\mathbf{x}}_{\mathcal{J}^{(n)}}$ for the restriction of $\hat{\mathbf{x}}$ to the index set $\bigcup_{J\in\mathcal{J}^{(n)}} J$. We then iterate over $n = 1,\ldots,G-1$ until the stopping criterion

$$\left\|\hat{\mathbf{x}}_{\mathcal{J}^{(n+1)}} - \hat{\mathbf{x}}_{\mathcal{J}^{(n)}}\right\|_2 / \left\|\hat{\mathbf{x}}_{\mathcal{J}^{(n+1)}}\right\|_2 \leq \mu$$

is met at some iteration $n^\star$ for the prescribed tolerance $\mu = 10^{-3}$, and set $\hat{\mathcal{S}} = \{\pi(1),\ldots,\pi(n^\star)\}$.

The results of this experiment are shown in Fig. 2. While both $\Delta_{\mathrm{CP},\mathcal{I}}$ and $\Delta_{\mathrm{HT},\mathcal{I}}$ manage to almost perfectly recover the exact group support for $m \geq 400$, $\Delta_{\mathrm{corr},\mathcal{I}}$ consistently misidentifies around 6 groups even as the number of measurements increases. On the other hand, despite outperforming both $\Delta_{\mathrm{corr},\mathcal{I}}$ and $\Delta_{\mathrm{HT},\mathcal{I}}$ in the previous experiment, $\Delta_{\mathrm{CP},\widetilde{\mathcal{I}}}$ fails to properly identify the group support if the inherent group structure is not explicitly exploited. This can be explained as follows. Despite being oblivious to the underlying group structure, the estimator $\Delta_{\mathrm{CP},\widetilde{\mathcal{I}}}$ manages to identify the most important coefficients inside each active group. However, on average the program also selects several coefficients which do not belong to one of the active groups which in turn leads to a large group misclassification rate.

5. CONCLUSION

In this paper, we considered the reconstruction of group-sparse vectors from 1-bit observations. We established theoretical reconstruction guarantees for three recovery strategies modeled after existing schemes in the 1-bit compressed sensing literature. In particular, $\Omega\big(\varepsilon^{-\alpha}(s\log(G/s) + sg)\big)$ measurements suffice to estimate group-sparse signals up to an ε-fidelity where the integer power $\alpha \geq 3$ depends on the choice of the recovery procedure. Moreover, we numerically benchmarked the performance of each recovery scheme both in terms of effective signal reconstruction quality and group support identification. For group support recovery in particular, two of the considered methods almost always manage to correctly identify the group support even in highly undersampled regimes.

6. REFERENCES

[1] David L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, 2006.

[2] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[3] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Information Theory*, vol. 52, pp. 489–509, 2006.

[4] Yaniv Plan and Roman Vershynin, "The generalized lasso with non-linear observations," *IEEE Transactions on information theory*, vol. 62, no. 3, pp. 1528–1537, 2016.

[5] Yaniv Plan, Roman Vershynin, and Elena Yudovina, "High-dimensional estimation with geometric constraints," *Information and Inference: A Journal of the IMA*, vol. 6, no. 1, pp. 1–40, 2017.

[6] Simon Foucart and Tom Needham, "Sparse recovery from saturated measurements," *Information and Inference: A Journal of the IMA*, vol. 6, no. 2, pp. 196–212, 2016.

[7] Laurent Jacques, David K Hammond, and Jalal M Fadili, "Dequantizing compressed sensing: When oversampling and non-gaussian constraints combine," *IEEE Transactions on Information Theory*, vol. 57, no. 1, pp. 559–571, 2011.

[8] Petros T Boufounos and Richard G Baraniuk, "1-bit compressive sensing," in *Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on*. IEEE, 2008, pp. 16–21.

[9] Laurent Jacques, Jason N. Laska, Petros T. Boufounos, and Richard G. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *IEEE Transactions on Information Theory*, vol. 59, no. 4, pp. 2082–2102, 2013.

[10] Yaniv Plan and Roman Vershynin, "One-bit compressed sensing by linear programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1275–1297, feb 2013.

[11] Xiangrong Zeng and Mário A. T. Figueiredo, "Robust binary fused compressive sensing using adaptive outlier pursuit," *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7674–7678, 2014.

[12] Niklas Koep and Rudolf Mathar, "Block-sparse signal recovery from binary measurements," *2018 IEEE Statistical Signal Processing Workshop (SSP)*, pp. 293–297, 2018.

[13] Andreas M. Tillmann and Marc E. Pfetsch, "The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 1248–1259, 2014.

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

[15] Yaniv Plan and Roman Vershynin, "Dimension reduction by random hyperplane tessellations," *Discrete & Computational Geometry*, vol. 51, no. 2, pp. 438–461, 2014.

[16] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, OUP Oxford, 2013.

[17] Simon Foucart and Holger Rauhut, *A Mathematical Introduction to Compressive Sensing*, vol. 1, Birkhäuser, Basel, 2013.

[18] Dmitriy Bilyk and Michael T Lacey, "Random tessellations, restricted isometric embeddings, and one bit sensing," *arXiv preprint arXiv:1512.06697*, 2015.

[19] Roman Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.

[20] Yaniv Plan and Roman Vershynin, "Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach," *IEEE Transactions on Information Theory*, vol. 59, no. 1, pp. 482–494, 2013.

[21] Laurent Jacques, Kévin Degraux, and Christophe De Vleeschouwer, "Quantized iterative hard thresholding: Bridging 1-bit and high-resolution quantized compressed sensing," *CoRR*, vol. abs/1305.1786, 2013.

[22] Richard G. Baraniuk, Simon Foucart, Deanna Needell, Yaniv Plan, and Mary Wootters, "One-bit compressive sensing of dictionary-sparse signals," *CoRR*, vol. abs/1606.07531, 2016.