ONE-BIT COMPRESSED SENSING WITH NON-GAUSSIAN MEASUREMENTS

ALBERT AI, ALEX LAPANOWSKI, YANIV PLAN, AND ROMAN VERSHYNIN

Abstract. In one-bit compressed sensing, previous results state that sparse signals may be robustly recovered when the measurements are taken using Gaussian random vectors. In contrast to standard compressed sensing, these results are not extendable to natural non-Gaussian distributions without further assumptions, as can be demonstrated by simple counter-examples involving extremely sparse signals. We show that approximately sparse signals that are not extremely sparse can be accurately reconstructed from single-bit measurements sampled according to a sub-gaussian distribution, and the reconstruction comes as the solution to a convex program.

Keywords: 1-bit compressed sensing; quantization; signal reconstruction; convex programming

1. Introduction

In the standard noiseless compressed sensing model, one has access to linear measurements of the form

$$y_i = \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle, \qquad i = 1, 2, \ldots, m$$

where $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathbb{R}^n$ are known measurement vectors and $\boldsymbol{x} \in \mathbb{R}^n$ is a sparse signal which one wishes to reconstruct (see, e.g., [2]). Let $\|\boldsymbol{x}\|_0$ denote the number of nonzero entries in $\boldsymbol{x}$. Typical results state that when the measurement vectors are chosen randomly from a sub-gaussian distribution, and $\|\boldsymbol{x}\|_0 \le s$, then $m = O(s \log(n/s))$ measurements are sufficient for robust recovery of the signal $\boldsymbol{x}$ (see, [2]).

In noiseless one-bit compressed sensing, the measurements are compressed to single bits, and thus they take the form

$$y_i = \operatorname{sign}\left(\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle\right), \qquad i = 1, 2, \ldots, m. \tag{1.1}$$

Here, the sign function is defined by $\operatorname{sign}(t) = 1$ when $t \ge 0$ and -1 otherwise. Clearly, the magnitude of $\boldsymbol{x}$ is lost in these measurements and so the goal is to approximate the direction of $\boldsymbol{x}$. Thus we may assume without loss of generality that $\boldsymbol{x} \in S^{n-1}$.

One-bit compressed sensing was introduced in [1] to model extreme quantization in compressed sensing. The webpage `http://dsp.rice.edu/1bitCS/` details the recent literature that concerns

Date: Submitted August 2012.

2000 *Mathematics Subject Classification.* 94A12; 60D05; 90C25.

Y.P. was supported by an NSF Postdoctoral Research Fellowship under award No. 1103909. R.V. was supported by NSF grant 1001829. A.A. and A.L. were REU students supported by the NSF grant 0918623.

theoretical and algorithmic results on one-bit compressed sensing, as well as applications and extensions to quantization with more than two bits. Let us review the existing theoretical results on one-bit quantization.

Suppose that the signal $\boldsymbol{x} \in \mathbb{R}^n$ satisfies $\|\boldsymbol{x}\|_0 \leq s$. Gupta et al. [3] assume that the measurement vectors $\boldsymbol{a}_i$ are Gaussian and demonstrate that the support of $\boldsymbol{x}$ can tractably be recovered from either 1) $O(s \log n)$ nonadaptive measurements assuming a constant dynamic range of $\boldsymbol{x}$ (i.e. the magnitude of all nonzero entries of $\boldsymbol{x}$ is assumed to lie between two constants), or 2) $O(s \log n)$ adaptive measurements. Jacques et al. [4] introduce a certain *binary ϵ-stable embedding property* which is a one-bit analogue to the *restricted isometry property* of standard compressed sensing. They demonstrate that Gaussian measurement ensembles satisfy this property with high probability (given enough measurements). Assuming the binary ϵ-stable embedding property holds, they show that any estimate of $\boldsymbol{x}$ which is both s-sparse and approximately matches the data, will be accurate. In particular, $O(s \log n)$ Gaussian measurements are sufficient to have a relative error bounded by any fixed constant. These results are robust to noise.

Plan and Vershynin [7, 8] show that one may reconstruct a sparse signal $\boldsymbol{x}$ from single-bit measurements by *convex programming*, for which tractable solvers exist. [7] considers the noiseless case and [8] considers the noisy case (and also sparse logistic regression). In [8] and the present paper, the model for the signal $\boldsymbol{x}$ is allowed to be quite general, with sparsity as a special case. Indeed, suppose $\boldsymbol{x}$ belongs to some known set K, which is meant to encode the *model of the signal structure*. For example, in order to encode sparsity, one could let K be the set

$$S_{n,s} := \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_0 \leq s, \|\boldsymbol{x}\|_2 \leq 1\}.$$

The recovery is achieved in [8] by solving the optimization problem

$$\max \sum_{i=1}^{m} y_i \langle \boldsymbol{a}_i, \boldsymbol{x}' \rangle \qquad \text{subject to} \qquad \boldsymbol{x}' \in K. \tag{1.2}$$

If K is a convex set then (1.2) is a convex optimization problem, so it can be solved by a variety of convex optimization solvers.

However, the reader may note that the set of sparse vectors $S_{n,s}$ is extremely non-convex. To overcome this, it was proposed in [8] to take K to be an approximate convex relaxation of $S_{n,s}$ (see [7, Lemma 3.1]), namely

$$K = K_{n,s} := \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_2 \leq 1, \|\boldsymbol{x}\|_1 \leq \sqrt{s}\}. \tag{1.3}$$

It was shown in [8] that $m = O(s \log(n/s))$ Gaussian measurements are sufficient to accurately recover $\boldsymbol{x}$ by solving the convex optimization problem (1.2).

A natural question is whether reconstruction of $\boldsymbol{x}$ from one-bit measurements is still feasible when measurements are taken using random vectors with *non-Gaussian* coordinates. A simple counterexample shows that this is not generally possible even when the coordinates are sub-gaussian. Suppose that all coordinates of $\boldsymbol{a}_i$ are in $\{-1,1\}$. For example, one may let the coordinates be independent symmetric Bernoulli random variables. Then the vectors

$$\boldsymbol{x} = (1, \frac{1}{2}, 0, \ldots, 0) \quad \text{and} \quad \boldsymbol{x}' = (1, -\frac{1}{2}, 0, \ldots, 0)$$

clearly satisfy $\operatorname{sign}(\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle) = \operatorname{sign}(\langle \boldsymbol{a}_i, \boldsymbol{x}'\rangle)$. This shows that one can not distinguish the two very different signals $\boldsymbol{x}$ and $\boldsymbol{x}'$ by such measurements,[1] even if infinitely many measurements are taken.

One may ask whether this counterexample has typical or worst-case behavior. In this paper, we demonstrate that the latter is the case—*a difficulty can only arise for extremely sparse signals.* Namely, we show that under the assumption

$$\|\boldsymbol{x}\|_\infty \ll \|\boldsymbol{x}\|_2 = 1, \tag{1.4}$$

an approximate recovery of $\boldsymbol{x}$ is still possible with general sub-gaussian measurements, and it is achieved by the convex program (1.2). Furthermore, we prove that for the distributions that are near Gaussian (in total variation), an approximate recovery of $\boldsymbol{x}$ is possible even without the assumption (1.4).

1.1. **Main Results.** We shall assume that the signal set K lies in the unit Euclidean ball in $\mathbb{R}^n$, which we shall denote B_2^n. The quality of recovery of a signal $\boldsymbol{x} \in K$ will depend on K through a single geometric parameter – the *Gaussian mean width* of K. It is defined as

$$w(K) = \mathbb{E} \sup_{\boldsymbol{x} \in K-K} \langle \boldsymbol{g}, \boldsymbol{x} \rangle,$$

where $\boldsymbol{g}$ denotes a standard Gaussian random vector in $\mathbb{R}^n$, i.e. a vector with independent $N(0,1)$ random coordinates. The reader may refer to [8, Section 2] for a brief overview of the properties of mean width.

The main purpose of this paper is to allow the measurement vectors $\boldsymbol{a}_i$ to have general *sub-gaussian* (rather than Gaussian) independent coordinates. Recall that a random variable a is sub-gaussian if its distribution is dominated by a centered normal distribution. This property can be expressed in several equivalent ways, see [11, Section 5.2.3]. One convenient way to define a sub-gaussian random variable is to require that its moments be bounded by the corresponding moments of $N(0,1)$, so that $(\mathbb{E}\,|a|^p)^{1/p} = O(\sqrt{p})$ as $p \to \infty$. Formally, a is called sub-gaussian if

$$\kappa := \sup_{p \geq 1} p^{-1/2} (\mathbb{E}\,|a|^p)^{1/p} < \infty. \tag{1.5}$$

The quantity κ is called the *sub-gaussian norm* of a. The class of sub-gaussian random variables includes in particular normal, Bernoulli and all bounded random variables.

Our main result is a generalization of [8, Theorem 1.1], which states that when the measurement vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ are Gaussian, then

$$\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_2^2 \lesssim \frac{w(K)}{\sqrt{m}}$$

[1] One can normalize the signals $\boldsymbol{x}$ and $\boldsymbol{x}'$ to lie on S^{n-1}, and the same phenomenon clearly persists.

with high probability. Our generalization allows $\boldsymbol{a}_i$ to have coordinates with sub-gaussian distributions. The only important difference is that the error now has an additive dependence on $\|\boldsymbol{x}\|_\infty$. This serves to exclude extremely sparse signals, which can destroy recovery, according to the example we discussed above.

We will consider the noisy measurement model in which each 1-bit measurement is flipped with small probability

$$y_i = \varepsilon_i \operatorname{sign}(\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle). \tag{1.6}$$

Above, $\{\varepsilon_i\}$ is a collection of i.i.d. Bernoulli random variables satisfying $P(\varepsilon_i = 1) = 1 - p$ and $P(\varepsilon_i = -1) = p$.

Theorem 1.1 (Estimating a signal with random bit flips)**.** *Let $a \in \mathbb{R}$ be a symmetric, sub-gaussian, and unit variance random variable with κ as in* (1.5)*. Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ be independent random vectors in $\mathbb{R}^n$ whose coordinates are i.i.d. copies of a. Consider a signal set $K \subseteq B_2^n$, and fix $\boldsymbol{x} \in K$ satisfying $\|\boldsymbol{x}\|_2 = 1$. Let $\boldsymbol{y}$ follow the 1-bit measurement model of Equation* (1.6)*. Then for each $\beta > 0$, with probability at least $1 - 4e^{-\beta^2}$, the solution $\widehat{\boldsymbol{x}}$ to the optimization problem (1.2) satisfies*

$$\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_2^2 \le C \left(\kappa^3 \|\boldsymbol{x}\|_\infty^{1/2} + \frac{\kappa}{\sqrt{m}(2-p)} (w(K) + \beta) \right). \tag{1.7}$$

In this theorem and later, C and c denote positive absolute constants, which can be different from line to line.

A proof of Theorem 1.1 is given in Section 4.

This theorem can be easily specialized to sparse (and approximately sparse) signals. To this end, we consider $K = K_{n,s}$ as in (1.3). A standard computation (see [8, Equation 3.3]) shows that

$$w(K_{n,s}) \le C \sqrt{s \log(2n/s)}.$$

Then the following corollary follows directly from Theorem 1.1.

Corollary 1.2 (Estimating a sparse signal with random bit flips)**.** *Let $K = K_{n,s}$, $s \ge 1$, and let everything else be as in Theorem 1.1. Then with probability at least $1 - 4\exp\{-2s\log(2n/s)\} \ge 1 - \frac{1}{n^2}$, the solution $\widehat{\boldsymbol{x}}$ to the optimization problem (1.2) satisfies*

$$\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_2^2 \le C \left(\kappa^3 \|\boldsymbol{x}\|_\infty^{1/2} + \kappa \sqrt{\frac{s \log(n/s)}{m(2-p)^2}} \right).$$

In words, this result yields that if the signal is approximately s-sparse, but not extremely sparse so that $\|\boldsymbol{x}\|_\infty \ll \|\boldsymbol{x}\|_2 = 1$, then with high probability $\boldsymbol{x}$ can be accurately recovered from

$$m = O(s \log(n/s))$$

general sub-gaussian measurements—provided that at most a constant fraction of bits are randomly flipped. Interestingly, accurate reconstruction is still possible even if nearly half of the bits are flipped.

We also establish a version of Theorem 1.1 under a statistical model, which also corresponds to additive noise before quantization. We take the *generalized linear model* in which each measurement is modeled by a random variable y_i taking values in $\{-1, 1\}$ such that

$$\mathbb{E}(y_i|\boldsymbol{a}_i) = \theta(\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle), \qquad i = 1, 2, \ldots, m. \tag{1.8}$$

Conditionally on $\{\boldsymbol{a}_i\}$, the measurements y_i are assumed independent. $\theta : \mathbb{R}^d \to [-1, 1]$ is a measurable function, which may even be unknown or unspecified. We only assume that $\theta(t) \in C^3(\mathbb{R})$, the first three derivatives being bounded by τ_1, τ_2, τ_3 respectively, and that

$$\mathbb{E}\,\theta(g)g =: \lambda > 0 \tag{1.9}$$

where $g \sim N(0, 1)$. To see why this is a natural assumption, notice that $\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle \sim \mathcal{N}(0, 1)$ if $\boldsymbol{a}_i$ are standard Gaussian random vectors and $\|x\|_2 = 1$; thus

$$\mathbb{E}\, y_i \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle = \mathbb{E}\,\theta(g)g = \lambda.$$

For example, in sparse logistic regression one would take

$$\theta(t) = \tanh(t/2),$$

with bounds $\tau_1 = 0.5$, $\tau_2 \approx 0.19$, $\tau_3 \approx 0.083$ and $\lambda \approx 0.41$.

To note another important example, observe that the setting of Theorem 1.1 is described by choosing $\theta(t) = \mathrm{sign}(t)$ and disregarding the differentiability requirements. In this case, $\lambda = \mathbb{E}\,\theta(g)g = \mathbb{E}\,|g| = \sqrt{2/\pi}$.

The following is a version of Theorem 1.1 under this noisy or statistical model.

Theorem 1.3 (Estimating a spread signal in the generalized linear model)**.** *We remain in the setting of Theorem 1.1, but with random measurements y_i modeled as in Equation (1.8). Then for each $\beta > 0$, with probability at least $1 - 4e^{-\beta^2}$, the solution $\widehat{\boldsymbol{x}}$ to the optimization problem (1.2) satisfies*

$$\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_2^2 \leq C\left(\frac{\kappa^4}{\lambda}(\tau_2 + \tau_3)\|\boldsymbol{x}\|_\infty + \frac{\kappa}{\lambda\sqrt{m}}(w(K) + \beta)\right). \tag{1.10}$$

For Gaussian measurement vectors $\boldsymbol{a}_i$, a version of this theorem was proved in [8].

The proof of Theorem 1.3 is provided in Section 3.

An interested reader may specialize this result to sparse signals $\boldsymbol{x}$ as we did before, i.e. by taking $K = K_{n,s}$ and noting as in Corollary 1.2 that $w(K_{n,s}) \leq C\sqrt{\log(2n/s)}$.

Our last result is about sub-Gaussian distributions, which nevertheless are close to Gaussian in total variation. For such measurements, it is reasonable to expect that the same conclusions as for Gaussian measurements hold, i.e., that the theorems above hold for all signals $\boldsymbol{x}$ without any dependence on $\|\boldsymbol{x}\|_\infty$. We confirm that this is the case. Suppose that the coordinates of $\boldsymbol{a}_i$ are i.i.d. copies of a random variable a that satisfies the total variation bound

$$\|a - g\|_{TV} := \sup_A |P(a \in A) - P(g \in A)| \leq \varepsilon$$

where $g \sim N(0,1)$. In the case when $\theta(t) = \mathrm{sign}(t)$, one has

$$\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_2^2 \lesssim \varepsilon^{1/8} + \frac{w(K)}{\sqrt{m}},$$

and in the case when $\theta(t) \in C^2$ one has

$$\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_2^2 \lesssim \varepsilon^{1/2} + \frac{w(K)}{\sqrt{m}}.$$

Above, the $\lesssim$ notation hides dependence on $\kappa, \lambda, \tau_1, \tau_2$ and a numeric constant. The precise results and their proofs are provided in the appendix as Theorems 5.1 and 5.4, respectively.

Acknowledgements

The authors would like to thank the anonymous reviewers for thoughtful comments which helped to simplify the presentation and tighten results.

2. General proof structure

In this section we give the general structure behind the proofs of our theorems. We also give a general lemma which may be useful to other researchers who wish to develope theory for 1-bit compressed sensing under different measurement models.

It will be convenient to define the (rescaled) objective function for our convex program (1.2):

$$f_{\boldsymbol{x}}(\boldsymbol{x}') := \frac{1}{m} \sum_{i=1}^{m} y_i \langle \boldsymbol{a}_i, \boldsymbol{x}' \rangle.$$

Note that this is a random function whose distribution depends on the distribution of $\{\boldsymbol{a}_i\}$ and choice of θ. In order to demonstrate that $\hat{\boldsymbol{x}}$ is a good approximation of $\boldsymbol{x}$, we will need to control the expectation of f and the variation of f around its expectation. It turns out that f uniformly concentrates around its expectation value regardless of θ, but with dependence on the sub-Gaussian norm of $\boldsymbol{a}_i$.

Proposition 2.1 (Concentration)**.** *For each $\beta > 0$,*

$$P\left(\sup_{\boldsymbol{z} \in K-K} |f_{\boldsymbol{x}}(\boldsymbol{z}) - \mathbb{E}\, f_{\boldsymbol{x}}(\boldsymbol{z})| \geq C\kappa \frac{w(K) + \beta}{\sqrt{m}} \right) \leq 4e^{-\beta^2}.$$

The proof of Proposition 2.1 is provided in Section 4.1.

We turn to the expectation of $f_{\boldsymbol{x}}(\boldsymbol{x}')$. In the special case of standard normal measurement vectors $\boldsymbol{a}_i$, it is not hard to show that

$$\mathbb{E}\, f_{\boldsymbol{x}}(\boldsymbol{x}') = \lambda \langle \boldsymbol{x}, \boldsymbol{x}' \rangle.$$

(See [8, Lemma 4.1].) If we allow sub-gaussian measurement vectors, the equality no longer holds, but under some conditions on $\boldsymbol{x}$ and θ, we still have

$$\mathbb{E}\, f_{\boldsymbol{x}}(\boldsymbol{x}') \approx \lambda \langle \boldsymbol{x}, \boldsymbol{x}' \rangle.$$

We will prove the above approximate equality in subsequent sections, but for now we see how it implies accurate reconstruction of $\boldsymbol{x}$.

Lemma 2.2. *Fix $\boldsymbol{x} \in K$ with $\|\boldsymbol{x}\|_2 = 1$. Let $\theta : \mathbb{R} \to [-1,1]$ be measurable and suppose that $\boldsymbol{y}$ follows the generalized linear model (1.8). Let $\alpha > 0$ and suppose that for any $\boldsymbol{x}' \in K$,*

$$\left|\mathbb{E} f_{\boldsymbol{x}}(\boldsymbol{x}') - \lambda\langle \boldsymbol{x}, \boldsymbol{x}'\rangle\right| \leq \alpha. \tag{2.1}$$

Then for all $\beta > 0$, with probability at last $1 - 4\varepsilon^{-\beta^2}$ the solution $\hat{\boldsymbol{x}}$ to (1.2) satisfies

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2^2 \leq \frac{4\alpha}{\lambda} + C\kappa \frac{w(K) + \beta}{\lambda\sqrt{m}}.$$

Proof. Fix $\boldsymbol{x}' \in K$. We will show that $f_{\boldsymbol{x}}(\boldsymbol{x}')$ can only be large if $\boldsymbol{x}'$ is near to $\boldsymbol{x}$, and then use this to show that the maximizer $\hat{\boldsymbol{x}}$ must be accurate.

Let $\boldsymbol{z} = \boldsymbol{x}' - \boldsymbol{x} \in K - K$. We have,

$$-\mathbb{E} f_{\boldsymbol{x}}(\boldsymbol{z}) = \mathbb{E} f_{\boldsymbol{x}}(\boldsymbol{x}) - \mathbb{E} f_{\boldsymbol{x}}(\boldsymbol{x}') \geq \langle \lambda\boldsymbol{x}, \boldsymbol{x}\rangle - \langle \lambda\boldsymbol{x}, \boldsymbol{x}'\rangle - 2\alpha \geq \frac{\lambda}{2}\left\|\boldsymbol{x} - \boldsymbol{x}'\right\|_2^2 - 2\alpha.$$

The first inequality follows from Equation (2.1) and the second follows from $\|\boldsymbol{x}\|_2 = 1$.

Further, by Proposition 2.1, we have a lower bound of $1 - 4e^{-\beta^2}$ on the event

$$\sup_{\boldsymbol{z} \in K-K} |f_{\boldsymbol{x}}(\boldsymbol{z}) - \mathbb{E} f_{\boldsymbol{x}}(\boldsymbol{z})| \leq C\kappa \frac{w(K) + \beta}{\sqrt{m}}.$$

In this event, note that

$$f_{\boldsymbol{x}}(\boldsymbol{z}) \leq \mathbb{E} f_{\boldsymbol{x}}(\boldsymbol{z}) + C\kappa \frac{w(K) + \beta}{\sqrt{m}} \leq 2\alpha - \frac{\lambda}{2}\|\boldsymbol{x} - \boldsymbol{x}'\|_2^2 + C\kappa \frac{w(K) + \beta}{\sqrt{m}}.$$

This holds uniformly for all $\boldsymbol{x}' \in K$. Pick $\boldsymbol{x}' = \widehat{\boldsymbol{x}}$. Since $\hat{\boldsymbol{x}}$ maximizes $f_{\boldsymbol{x}}$ we have $f_{\boldsymbol{x}}(\boldsymbol{z}) = f_{\boldsymbol{x}}(\widehat{\boldsymbol{x}}) - f_{\boldsymbol{x}}(\boldsymbol{x}) \geq 0$. Thus the right-hand side of the above inequality is bounded below by 0. Rearranging completes the proof of the lemma. ■

3. Proof of Theorem 1.3

We only need to bound α in Equation (2.1). For convenience, let us denote $y := y_1$ and $\boldsymbol{a} := \boldsymbol{a}_1$. Recalling (1.8), we observe the following equivalences:

$$\mathbb{E} f_{\boldsymbol{x}}(\boldsymbol{x}') = \frac{1}{m}\sum_{i=1}^{m} \mathbb{E}\, y_i \langle \boldsymbol{a}_i, \boldsymbol{x}'\rangle = \mathbb{E}\, y\langle \boldsymbol{a}, \boldsymbol{x}'\rangle = \mathbb{E}(\mathbb{E}\, y\langle \boldsymbol{a}, \boldsymbol{x}'\rangle | \boldsymbol{a}) = \mathbb{E}\,\theta(\langle \boldsymbol{a}, \boldsymbol{x}\rangle)\langle \boldsymbol{a}, \boldsymbol{x}'\rangle. \tag{3.1}$$

We also note that for a standard normal vector $\boldsymbol{g}$, $\mathbb{E}\,\theta(\langle \boldsymbol{g}, \boldsymbol{x}\rangle)\langle \boldsymbol{g}, \boldsymbol{x}'\rangle = \lambda\langle \boldsymbol{x}, \boldsymbol{x}'\rangle$, which satisfies Equation (2.1) with $\alpha = 0$ (see [8, Lemma 4.1]). Thus, we need to show that the expectation in the sub-gaussian case nearly matches the Gaussian case. Such a comparison is a bi-variate version of Berry-Esseen central limit theorem for the function $\theta(\langle \boldsymbol{a}, \boldsymbol{x}\rangle)\langle \boldsymbol{a}, \boldsymbol{x}'\rangle$.

Lemma 3.1 (Berry-Esseen type central limit theorem). *Consider $\boldsymbol{x}, \boldsymbol{z} \in B_2^n$. Let $\boldsymbol{a}$ be a random vector with i.i.d. mean-zero, variance-one, sub-gaussian entries whose sub-gaussian norm is bounded by κ. Let $\boldsymbol{g}$ be a vector with independent standard normal entries. Let $\theta : \mathbb{R} \to \mathbb{R}$ be a measurable function satisfying $\|\theta''\|_\infty \le \tau_2$ and $\|\theta'''\|_\infty \le \tau_3$. Then*

$$|\mathbb{E}\,\theta(\langle \boldsymbol{a}, \boldsymbol{x}\rangle)\langle \boldsymbol{a}, \boldsymbol{z}\rangle - \mathbb{E}\,\theta(\langle \boldsymbol{g}, \boldsymbol{x}\rangle)\langle \boldsymbol{g}, \boldsymbol{z}\rangle| \le C(\tau_2 + \tau_3)\,\mathbb{E}\,a^4 \|\boldsymbol{x}\|_\infty, \tag{3.2}$$

The proof is based on a Lindeberg replacement argument in two variables; it is provided in the appendix. Note that the quality of approximation in this theorem is the same for all $\boldsymbol{z} \in B_2^n$; this will be crucial for our argument.

We apply the lemma to prove Theorem 1.3.

Proof of Theorem 1.3. We set α to be the right-hand side of Equation (3.2). Further, by definition, $\mathbb{E}\,a^4 \le 16\kappa^4$. Thus, we may apply Lemma 2.2 with

$$\alpha \le C\kappa^4(\tau_2 + \tau_3)\|\boldsymbol{x}\|_\infty$$

to complete the proof. ∎

4. Proof of Theorem 1.1

For simplicity, we first assume the noiseless model in (1.1); we will then describe a minor adjustment to the proof to generalize to the random bit flip model in (1.6).

The essential difference from the proof of Theorem 1.3 is that $\theta(t) = \operatorname{sign}(t)$ is not differentiable. One approach would be to approximate θ by a smooth function and apply the Berry-Esseen type central limit theorem (Lemma 3.1). However, we achieve a tighter bound with a different approach. Once again, we only need to bound α in Equation (2.1). This bound is contained in the following proposition. In this section, $\lambda = \mathbb{E}\,|g| = \sqrt{2/\pi}$.

Proposition 4.1 (Expectation). *Consider $\boldsymbol{x} \in S^{n-1}, \boldsymbol{x}' \in B_2^n$. If $\|\boldsymbol{x}\|_\infty \le c/\,\mathbb{E}\,|a|^3$, then*

$$|\mathbb{E}\, f_{\boldsymbol{x}}(\boldsymbol{x}') - \lambda\langle \boldsymbol{x}, \boldsymbol{x}'\rangle| \le C\,\mathbb{E}\,|a|^3 \|\boldsymbol{x}\|_\infty^{1/2}. \tag{4.1}$$

We pause to prove Theorem 1.1 based on the proposition.

Proof of Theorem 1.1. By definition of κ, we have $\mathbb{E}\,|a|^3 \le 3^{3/2}\kappa^3$. Now suppose that $\|\boldsymbol{x}\|_\infty \ge c/\,\mathbb{E}\,|a|^3$. Thus, $\|\boldsymbol{x}\|_\infty \ge c'/\kappa^3$. Then the right-hand side of Equation (1.7) in Theorem 1.1 is lower bounded by

$$C \cdot c'\kappa^{3/2} \ge C \cdot c'.$$

Take $C \ge 4/c'$ in which case the theorem trivially holds since $\hat{\boldsymbol{x}}, \boldsymbol{x} \in B_2^n$.

On the other hand, if $\|\boldsymbol{x}\|_\infty \ge c/\,\mathbb{E}\,|a|^3$ we can apply Proposition 4.1. We apply the proposition to bound α in Lemma 2.2. This gives

$$\alpha = C\kappa^3\|\boldsymbol{x}\|_\infty^{1/2}$$

and completes the proof of the theorem in the noiseless case.

In the random bit flip model of Equation (1.6), $\mathbb{E} f_{\boldsymbol{x}}(\boldsymbol{x}')$ is scaled by a factor of $\mathbb{E}\,\varepsilon_i = 2 - p$. This has the effective of scaling λ and the right-hand side of Equation (4.1) by a factor of $2 - p$. Thus, we complete the proof by applying Lemma 2.2 with

$$\alpha = C(2-p)\kappa^3 \|\boldsymbol{x}\|_\infty^{1/2} \quad \text{and} \quad \lambda = \sqrt{2/\pi}(2-p).$$

■

In order to prove Proposition 4.1, we will need to use two different known one-dimensional Berry-Esseen results (see [9, Theorems 2.1.24 and 2.1.30]), stated below in simplified form for the convenience of the reader:

Theorem 4.2 (One-dimensional Berry-Esseen central limit theorem)**.** *Let $\boldsymbol{z}$ be a random vector with n independent, mean-zero, entries satisfying $\mathbb{E}\,\|\boldsymbol{z}\|_2^2 = 1$. Set*

$$S_n = \sum_{i=1}^n z_i \quad \textit{and} \quad \beta^3 := \mathbb{E}\,\|\boldsymbol{z}\|_3^3 := \sum_{i=1}^n \mathbb{E}\,|z_i|^3$$

and let g be a standard normal random variable. We have

$$\int_{-\infty}^{\infty} |P(S_n \le t) - P(g \le t)|\, dt \le 9\beta^3 \tag{4.2}$$

and

$$\sup_t |P(S_n \le t) - P(g \le t)| \le 10\beta^3. \tag{4.3}$$

We now prove Proposition 4.1 using a simple geometric argument. Define the vector

$$\boldsymbol{v}_x := \mathbb{E}\operatorname{sign}(\langle \boldsymbol{a}, \boldsymbol{x}\rangle)\boldsymbol{a}$$

and note that the conclusion of Proposition 4.1 states that

$$\boldsymbol{v}_x \approx \lambda \boldsymbol{x}.$$

In order to prove that the above approximate equality holds, we will derive it from two *scalar* inequalities:

$$\langle \boldsymbol{v}_x, \boldsymbol{x}\rangle \approx \lambda \quad \text{and} \quad \|\boldsymbol{v}_x\|_2 \lesssim \lambda. \tag{4.4}$$

(Indeed, the first of these approximate identities states that $\boldsymbol{v}_x$ is near a hyperplane with normal $\boldsymbol{x}$, and the second one states that $\boldsymbol{v}_x$ is nearly in the ball which intersects that hyperplane at the point $\lambda\boldsymbol{x}$.) Thus we reduce the problem to proving (4.4).

Lemma 4.3. $|\langle \boldsymbol{v}_x, \boldsymbol{x}\rangle - \lambda| \le C\,\mathbb{E}\,|a|^3 \|\boldsymbol{x}\|_3^3 \le C\,\mathbb{E}\,|a|^3\,\|\boldsymbol{x}\|_\infty$.

Proof. Recall that by definition of $\boldsymbol{v}_x$,

$$\langle \boldsymbol{v}_x, \boldsymbol{x}\rangle = \mathbb{E}\operatorname{sign}(\langle \boldsymbol{a}, \boldsymbol{x}\rangle)\langle \boldsymbol{a}, \boldsymbol{x}\rangle = \mathbb{E}\,|\langle \boldsymbol{a}, \boldsymbol{x}\rangle|\,.$$

Note that $\lambda = \sqrt{2/\pi} = \mathbb{E}\,|g|$ and thus, to prove the lemma, we wish to bound the difference $\big|\,\mathbb{E}\,|\langle \boldsymbol{a}, \boldsymbol{x}\rangle| - \mathbb{E}\,|g|\,\big|$. We have

$$\big|\,\mathbb{E}\,|\langle \boldsymbol{a}, \boldsymbol{x}\rangle| - \mathbb{E}\,|g|\,\big| = \left|\int_0^\infty P(|\langle \boldsymbol{a}, \boldsymbol{x}\rangle| \geq t) - P(|g| \geq t)dt\right| = 2\left|\int_0^\infty P(\langle \boldsymbol{a}, \boldsymbol{x}\rangle \geq t) - P(g \geq t)dt\right|.$$

To bound the right-hand side, we apply the Berry-Esseen result in Equation (4.2) which bounds the above quantity by

$$C\sum_{i=1}^n \mathbb{E}\,|x_i a_i|^3 = C\,\mathbb{E}\,|a|^3 \|\boldsymbol{x}\|_3^3.$$

■

To bound $\|\boldsymbol{v}_x\|_2$ we will apply the Berry-Esseen Theorem with $\boldsymbol{z} = \boldsymbol{v}_x / \|\boldsymbol{v}_x\|_2$. This will require first a rough two-sided bound on $\|\boldsymbol{v}_x\|_2$ and also an upper bound on $\|\boldsymbol{v}_x\|_\infty$. We establish these in the following two lemmas.

Lemma 4.4. *Suppose that* $\|\boldsymbol{x}\|_\infty \leq c/\,\mathbb{E}\,|a|^3$. *Then* $\frac{1}{2} \leq \|\boldsymbol{v}_x\|_2 \leq 1$.

Proof. For the lower bound, using Lemma 4.3, we have

$$\|\boldsymbol{v}_x\|_2 = \|\boldsymbol{v}_x\|_2 \|\boldsymbol{x}\|_2 \geq |\langle \boldsymbol{v}_x, \boldsymbol{x}\rangle| \geq \lambda - C\,\mathbb{E}\,|a|^3 \|\boldsymbol{x}\|_\infty .$$

Since $\lambda = \sqrt{2/\pi}$, and $\|\boldsymbol{x}\|_\infty \leq c/\,\mathbb{E}\,|a|^3$, the right-hand side is greater than $1/2$, as long as we take $c \leq (\sqrt{2/\pi} - 1/2)/C$.

In the other direction, we have

$$\|\boldsymbol{v}_x\|_2^2 = \langle \boldsymbol{v}_x, \boldsymbol{v}_x\rangle = \mathbb{E}\,\mathrm{sign}(\langle \boldsymbol{a}, \boldsymbol{x}\rangle)\langle \boldsymbol{a}, \boldsymbol{v}_x\rangle \leq \mathbb{E}\,|\langle \boldsymbol{a}, \boldsymbol{v}_x\rangle| \leq (\mathbb{E}\langle \boldsymbol{a}, \boldsymbol{v}_x\rangle^2)^{1/2} = \|\boldsymbol{v}_x\|_2 . \tag{4.5}$$

It follows that $\|\boldsymbol{v}_x\|_2 \leq 1$. ■

Lemma 4.5. *Suppose that* $\|\boldsymbol{x}\|_\infty \leq c/\,\mathbb{E}\,|a|^3$. *Then,* $\|\boldsymbol{v}_x\|_\infty \leq C\,\mathbb{E}\,|a|^3 \|\boldsymbol{x}\|_\infty$.

Proof. Establishing the notation $\langle \boldsymbol{a}, \boldsymbol{x}\rangle = \sum_{k=1}^n a_k x_k$ where without loss of generality, $x_i \geq 0$, define for convenience $S = \sum_{k\neq i}^n a_k x_k$ and let F_S be the cumulative distribution function of S. Consider an arbitrary constant r.

$$\begin{aligned}
|\mathbb{E}\,\theta(S + rx_i)\cdot r| &= \left| r\int_{\mathbb{R}} \mathrm{sign}(t + rx_i)dF_S(t)\right| \\
&= |r|\,|P(S \geq -rx_i) - P(S < -rx_i)| = |r|P(|S| \leq |r|x_i) \\
&\leq |r|\,P(|g| \leq |r|\,x_i) + |r|\cdot|P(|g| \leq |r|\,x_i) - P(|S| \leq |r|x_i)| .
\end{aligned}$$

The second term in the last inequality may be bounded using the Berry-Esseen result in Equation (4.3). This gives

$$|\mathbb{E}\,\theta(S + rx_i)\cdot r| \leq |r|\left\{\sqrt{\frac{2}{\pi}}|r|x_i + 2\left(\sum_{k\neq i} x_k^2\right)^{-3/2} \mathbb{E}\,|a|^3 \sum_{k\neq i} |x_k|^3\right\}.$$

Note $\|\boldsymbol{x}\|_3^3 \le \|\boldsymbol{x}\|_\infty \|\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|_\infty \le c/\mathbb{E}\,|a|^3 \le 1/8$, where the last inequality follows since $\mathbb{E}\,|a|^3 \ge (\mathbb{E}\,a^2)^{3/2} = 1$ and we take $c \le 1/8$. Then $x_i^3 \le 1/8$, $x_i^2 \le 1/4$, so that $\sum_{k\neq i} x_i^2 \ge 3/4$. Observing furthermore that $\|\boldsymbol{x}\|_\infty \ge \sum_{k\neq i} x_k^2 \|\boldsymbol{x}\|_\infty \ge \sum_{k\neq i} |x_k|^3$, we have the bound

$$|\mathbb{E}\,\theta(S + rx_i)\cdot r| \le C\,|r|^2\,x_i + C\,|r|\,\mathbb{E}\,|a|^3 \|\boldsymbol{x}\|_\infty.$$

We may express a single coordinate of $v_x = \mathbb{E}\,\theta(\langle \boldsymbol{a}, \boldsymbol{x}\rangle)\cdot \boldsymbol{a}$ as $\mathbb{E}\,\theta(\langle \boldsymbol{a}, \boldsymbol{x}\rangle)\cdot a_i$. Then,

$$\begin{aligned}
|\,\mathbb{E}\,\theta(\langle \boldsymbol{a}, \boldsymbol{x}\rangle)\cdot a_i| &\le \int_{\mathbb{R}} |\mathbb{E}\,\theta(S + tx_i)\cdot t|\,dF_{a_i}(t)\\
&\le \int_{\mathbb{R}} \left(Ct^2 x_i + C|t|\,\mathbb{E}\,|a|^3 \|\boldsymbol{x}\|_\infty\right) dF_{a_i}(t)\\
&= Cx_i\,\mathbb{E}\,a_i^2 + C\,\mathbb{E}\,|a|^3 \|\boldsymbol{x}\|_\infty\,\mathbb{E}\,|a_i|\\
&\le Cx_i + C\,\mathbb{E}\,|a|^3 \|\boldsymbol{x}\|_\infty.
\end{aligned}$$

Observing that $\mathbb{E}\,|a|^3 \ge \mathbb{E}\,a^2 = 1$ completes the proof of the lemma. ∎

We now prove Proposition 4.1.

Proof of Proposition 4.1. Define $\boldsymbol{z} = \boldsymbol{v}_x/\|\boldsymbol{v}_x\|_2$ and note that $\|\boldsymbol{z}\|_\infty = \|\boldsymbol{v}_x\|_\infty / \|\boldsymbol{v}_x\|_2$. Applying Lemma 4.4 and Lemma 4.5 yields $\|\boldsymbol{z}\|_\infty \le CE|a|^3 \|\boldsymbol{x}\|_\infty$. Hence,

$$\begin{aligned}
\|\boldsymbol{v}_x\|_2 = \langle \boldsymbol{v}_x, \boldsymbol{v}_x/\|\boldsymbol{v}_x\|_2\rangle &= \mathbb{E}\,\mathrm{sign}(\langle \boldsymbol{a}, \boldsymbol{x}\rangle)\langle \boldsymbol{a}, \boldsymbol{z}\rangle\\
&\le \mathbb{E}\,\mathrm{sign}(\langle \boldsymbol{a}, \boldsymbol{z}\rangle)\langle \boldsymbol{a}, \boldsymbol{z}\rangle = \langle \boldsymbol{v}_z, \boldsymbol{z}\rangle\\
&\le \lambda + C\,\mathbb{E}\,|a|^3 \|\boldsymbol{z}\|_\infty \le \lambda + C(\mathbb{E}\,|a|^3)^2 \|\boldsymbol{x}\|_\infty.
\end{aligned}$$

In the last line, we used Lemma 4.3. Together with Lemma 4.3 we have now verified both geometric constraints in (4.4).

Combining results, we have

$$\begin{aligned}
|\langle \boldsymbol{v}_x, \boldsymbol{x}'\rangle - \langle \lambda\boldsymbol{x}, \boldsymbol{x}'\rangle|^2 &\le \|\boldsymbol{v}_x\|_2^2 - \lambda^2 + 2\lambda(\lambda - \langle \boldsymbol{v}_x, \boldsymbol{x}\rangle)\\
&= (\|\boldsymbol{v}_x\|_2 + \lambda)(\|\boldsymbol{v}_x\|_2 - \lambda) + 2\lambda(\lambda - \langle \boldsymbol{v}_x, \boldsymbol{x}\rangle)\\
&\le C\left((\mathbb{E}\,|a|^3)^2 \|\boldsymbol{x}\|_\infty + \mathbb{E}\,|a|^3 \|\boldsymbol{x}\|_3^3\right).
\end{aligned}$$

Recall that $\|\boldsymbol{x}\|_\infty \ge \|\boldsymbol{x}\|_3^3$ and thus the first term is dominant. We may collect terms to conclude

$$|\,\mathbb{E}\,f_{\boldsymbol{x}}(\boldsymbol{x}') - \langle \lambda\boldsymbol{x}, \boldsymbol{x}'\rangle| \le C\,\mathbb{E}\,|a|^3 \|\boldsymbol{x}\|_\infty^{1/2}.$$

This completes the proof of Proposition 4.1. ∎

4.1. **Concentration: Proof of Proposition 2.1.** We need to control the random variable

$$Z := \sup_{\boldsymbol{z}\in K-K} |f_{\boldsymbol{x}}(\boldsymbol{z}) - \mathbb{E}\,f_{\boldsymbol{x}}(\boldsymbol{z})|.$$

This will be done using techniques from probability in Banach spaces, following the argument in [8, Proposition 4.2]. The symmetrization lemma below allows us to essentially replace Z by the random variable

$$Z' := \sup_{\boldsymbol{z} \in K-K} \frac{1}{m} \left| \sum_{i=1}^{m} \varepsilon_i y_i \langle \boldsymbol{a}_i, \boldsymbol{z} \rangle \right|.$$

where ε_i denote independent symmetric Bernoulli random variables.

Lemma 4.6 (Symmetrization)**.** *We have*

$$\mathbb{E}\, Z \le 2\, \mathbb{E}\, Z'. \tag{4.6}$$

Furthermore, for each $t > 0$ we have the deviation inequality

$$P(Z \ge 2\, \mathbb{E}\, Z + t) \le 4P(Z' > t/2). \tag{4.7}$$

The proof of this result is identical to the proof of [8, Lemma 5.1].

The following is a standard Gaussian concentration inequality, which is a simple application of [5, Theorem 7.1].

Lemma 4.7 (Gaussian concentration)**.** *Given a set $K \subseteq B_2^n$, we have*

$$P\left(\sup_{\boldsymbol{z} \in K-K} \langle \boldsymbol{g}, \boldsymbol{z} \rangle - w(K) > r \right) \le e^{-r^2/8}, \;\; r > 0.$$

The following inequality is a specialization of [6, Lemma 4.6]. (In contrast to [6, Lemma 4.6] we allow the consideration of a semi-norm, but the proof of the inequality remains unchanged.)

Lemma 4.8 (Contraction Principle)**.** *Consider sequences of independent symmetric random variables η_i and ξ_i such that for some scalar $M \ge 1$, and every i and $t > 0$,*

$$P(|\eta_i| > t) \le M P(|\xi_i| > t).$$

Let $\|\cdot\|$ denote a semi-norm. Then for any finite sequence x_i and a scalar $p \ge 1$, we have

$$\mathbb{E} \left(\left\| \sum_{i=1}^{n} \eta_i x_i \right\| \right)^p \le \mathbb{E} \left(M \left\| \sum_{i=1}^{n} \xi_i x_i \right\| \right)^p.$$

We will first apply Lemma 4.8 to derive a moment bound on Z'. We then convert the moment bound back into a tail bound which we plug into the right-hand side of Equation (4.7).

Because $\varepsilon_i y_i \boldsymbol{a}_i$ has the same distribution as $\boldsymbol{a}_i$, and by the symmetry of $K - K$,

$$\mathbb{E}(Z')^p = \mathbb{E} \left(\sup_{\boldsymbol{z} \in K-K} \frac{1}{m} \sum_{i=1}^{m} \langle \boldsymbol{a}_i, \boldsymbol{z} \rangle \right)^p = \mathbb{E} \left(\sup_{\boldsymbol{z} \in K-K} \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{n} (a_i)_j z_j \right)^p.$$

We apply Lemma 4.8 with $(a_i)_j$ in place of η_i, $e_i e_j^*$ in place of x_i (where e_i is the i-th standard basis vector), ξ_i as independent $N(0,1)$ random variables, and the matrix semi-norm defined by

$\|A\| := \sup_{\boldsymbol{z}\in K-K} \sum_{i,j} A_{i,j} z_j$. To this end, recall that $(a_i)_j$ are distributed identically with a. Since a is a sub-gaussian random variable, it follows from definition (1.5) that

$$P(|a| > t) \leq CP(|g| \cdot \kappa > t), \quad t > 0.$$

Therefore an application of Lemma 4.8 allows us to replace $(a_i)_j$ by $(C\kappa)(g_i)_j$ and thus conclude that

$$\mathbb{E}(Z')^p \leq \mathbb{E}\left(\sup_{\boldsymbol{z}\in K-K} \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n} C\kappa(g_i)_j z_j\right)^p = \mathbb{E}\left(\frac{C\kappa}{\sqrt{m}} \sup_{\boldsymbol{z}\in K-K}\langle \boldsymbol{g}, \boldsymbol{z}\rangle\right)^p. \tag{4.8}$$

To further develop this inequality, we express the Gaussian concentration tail bound (Lemma 4.7) in terms of moment bounds. For convenience, define

$$\xi = \sup_{\boldsymbol{z}\in K-K}\langle \boldsymbol{g}, \boldsymbol{z}\rangle.$$

Using Lemma 4.7 and the equivalence of sub-gaussian properties, for instance in [11, Lemma 5.5], we have

$$(\mathbb{E}(\xi - w(K))_+^p)^{1/p} \leq C\sqrt{p}.$$

Above $(\xi - w(K))_+ := \max(\xi - w(K), 0)$. Applying Minkowski's inequality gives

$$(\mathbb{E}\,\xi^p)^{1/p} \leq (\mathbb{E}(\xi - w(K))_+^p)^{1/p} + (\mathbb{E}\,w(K)^p)^{1/p} \leq C\sqrt{p} + w(K).$$

Combine this with Equation (4.8) to give the moment bound

$$(\mathbb{E}(Z')^p)^{1/p} \leq C \cdot \frac{\kappa(\sqrt{p} + w(K))}{\sqrt{m}}. \tag{4.9}$$

For convenience, set $\beta = \sqrt{p}$. We now use Markov's inquality to convert the moment bound to a tail bound. Set $t = e \cdot (\mathbb{E}(Z')^p)^{1/p}$ and note that by Equation (4.9) above,

$$t \leq C \cdot \frac{\kappa(\beta + w(K))}{\sqrt{m}}.$$

Further, by Markov's inequality we have

$$P(Z' \geq t) \leq \frac{\mathbb{E}(Z')^p}{t^p} \leq e^{-\beta^2} \tag{4.10}$$

To complete the proof of the proposition, apply Lemma 4.6: The moment bound (4.9) with $p = 1$ controls $\mathbb{E}(Z')$ and the tail bound (4.10) controls the right-hand side of Equation (4.7).

5. Conclusion

In contrast to standard compressed sensing, one-bit compressed sensing is infeasible when the measurement vectors are Bernoulli and the signal is extremely sparse. Nevertheless, we show that when the signal is sparse, but not overly sparse, it may be recovered from Bernoulli (or more generally, sub-gaussian) one-bit measurements. To our knowledge, these are the first theoretical results in one-bit compressed sensing that specifically allow non-Gaussian measurements.

5.1. **Proof of Lemma 3.1.** We apply a Lindeberg replacement argument in a way similar to [10, Proposition D.2]. Define $v_j = (x_j, z_j)$, and let $\boldsymbol{g} \in \mathbb{R}^n$ be a vector of independent standard Gaussian variables which is also independent of $\boldsymbol{a}$. Define $S_i = \sum_{j=1}^{i-1} a_j v_j + \sum_{j=i+1}^{n} g_j v_j$ and $\phi(v) = \theta(x)z$ (where $v = (x, z)$). Define $(S_i)_1$ to be the x component and $(S_i)_2$ to be the z component. Then note by telescoping,

$$|\mathbb{E}\,\theta(\langle \boldsymbol{a}, \boldsymbol{x}\rangle)\langle \boldsymbol{a}, \boldsymbol{z}\rangle - \mathbb{E}\,\theta(\langle \boldsymbol{g}, \boldsymbol{x}\rangle)\langle \boldsymbol{g}, \boldsymbol{z}\rangle| = \left|\mathbb{E}\,\phi\left(\sum_{j=1}^{n} a_j v_j\right) - \mathbb{E}\,\phi\left(\sum_{j=1}^{n} g_j v_j\right)\right|$$

$$\leq \sum_{i=1}^{n} |\mathbb{E}\,\phi(S_i + a_i v_i) - \mathbb{E}\,\phi(S_i + g_i v_i)|.$$

By Taylor's theorem with remainder, we have

$$\phi(S_i + a_i v_i) = \phi(S_i) + \sum_{|\alpha|=1} (a_i v_i)^\alpha \partial^\alpha \phi(S_i) + \frac{1}{2} \sum_{|\alpha|=2} (a_i v_i)^\alpha \partial^\alpha \phi(S_i) + \frac{1}{6} \sum_{|\alpha|=3} (a_i v_i)^\alpha \partial^\alpha \phi(S_i')$$

for some S_i' on the line segment joining S_i and $S_i + a_i v_i$. A similar result holds for $\phi(S_i + g_i v_i)$ with respective S_i''. Observe that since $\mathbb{E}\,a = \mathbb{E}\,g = 0$ and $\mathbb{E}\,a^2 = \mathbb{E}\,g^2 = 1$, the zeroth to second order terms cancel upon taking expectations in the difference.

$$|\mathbb{E}\,\phi(S_i + a_i v_i) - \mathbb{E}\,\phi(S_i + g_i v_i)| = \frac{1}{6}\left|\mathbb{E} \sum_{|\alpha|=3} (a_i v_i)^\alpha \partial^\alpha \phi(S_i') - \mathbb{E} \sum_{|\alpha|=3} (g_i v_i)^\alpha \partial^\alpha \phi(S_i'')\right|.$$

Consider the first expectation on the right hand side. Observe that the partials in the error vanish except when at most one partial is taken on the second argument of ϕ, yielding either $\theta''(x)$ or $\theta'''(x)z$. Furthermore, note that since S_i' is on the line segment joining S_i and $S_i + a_i v_i$, we may apply the bound $|(S_i')_2| \leq |(S_i)_2| + |a_i z_i|$ to conclude

$$\mathbb{E}\left|\sum_{|\alpha|=3} (a_i v_i)^\alpha \partial^\alpha \phi(S_i')\right| \leq C\,\mathbb{E}\,|a_i|^3 (x_i^2 |z_i| + |x_i|^3)(\|\theta''\|_\infty + \|\theta'''\|_\infty (|(S_i)_2| + |a_i z_i|))$$

$$= C(x_i^2|z_i| + |x_i|^3)(\tau_2\,\mathbb{E}\,|a_i|^3 + \tau_3(\mathbb{E}\,|(S_i)_2 a_i^3| + |z_i|\,\mathbb{E}\,a_i^4))$$

Observe that $(S_i)_2$ and a_i are independent, and $(\mathbb{E}\,|(S_i)_2|)^2 \leq \mathbb{E}(S_i)_2^2 \leq 1$ by Cauchy-Schwarz and the fact that the variance of an independent sum is a sum of variances. Further observing that $|z_i| \leq 1$ and $\mathbb{E}\,|a|^3 \leq \mathbb{E}\,a^4$, we may collect terms to conclude

$$\mathbb{E}\left|\sum_{|\alpha|=3} (a_i v_i)^\alpha \partial^\alpha \phi(S_i')\right| \leq C\|\boldsymbol{x}\|_\infty (|x_i z_i| + x_i^2)(\tau_2 + 2\tau_3)\,\mathbb{E}\,a_i^4.$$

A similar bound follows for the remainder from the Gaussian expansion, and observe that the Gaussian remainder can be absorbed since $\mathbb{E}\, a^4 \geq \mathbb{E}\, a^2 = 1$. Note that

$$\|\boldsymbol{x}\|_\infty \left(\sum_{i=1}^n |x_i z_i| + x_i^2 \right) \leq \|\boldsymbol{x}\|_\infty (\|\boldsymbol{x}\|_2 \|\boldsymbol{z}\|_2 + \|\boldsymbol{x}\|_2^2) \leq 2\|x\|_\infty$$

so that summing over i from 1 to n,

$$\left| \mathbb{E}\, \theta(\langle \boldsymbol{a}, \boldsymbol{x} \rangle) \langle \boldsymbol{a}, \boldsymbol{z} \rangle - \mathbb{E}\, \theta(\langle \boldsymbol{g}, \boldsymbol{x} \rangle) \langle \boldsymbol{g}, \boldsymbol{z} \rangle \right| \leq C(\tau_2 + \tau_3)\, \mathbb{E}\, a^4 \|\boldsymbol{x}\|_\infty,$$

which completes the proof of the lemma.

5.2. **Total variation: sign function.** We consider the setting of Theorem 1.1, where $\theta(t) = (2-p)\operatorname{sign}(t)$, with the additional assumption that $\|a - g\|_{TV} \leq \varepsilon$.

Theorem 5.1 (Estimating a signal with no noise)**.** *We remain in the setting of Theorem 1.1 with the additional condition $\|a - g\|_{TV} \leq \varepsilon$. Then for each $\beta > 0$, with probability at least $1 - 4e^{-\beta^2}$, the solution $\widehat{\boldsymbol{x}}$ to the optimization problem (1.2) satisfies*

$$\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_2^2 \leq C\sqrt{\kappa}\varepsilon^{1/8} + \frac{C\kappa}{(2-p)\sqrt{m}}(w(K) + \beta).$$

To prove the theorem, we only need bound α in Lemma 2.2. This is contained in the following proposition. For simplicity, we will prove the theorem in the noiseless case when $p = 0$, but note that the noisy case follows from a simple rescaling argument as in the proof of Theorem 1.1. Below, $\lambda = \sqrt{2/\pi}$.

Proposition 5.2 (Expectation)**.** *For $\boldsymbol{x}, \boldsymbol{x}' \in B_2^n$,*

$$|\,\mathbb{E}\, f_{\boldsymbol{x}}(\boldsymbol{x}') - \langle \lambda \boldsymbol{x}, \boldsymbol{x}' \rangle| \leq C(\mathbb{E}\, a^4)^{1/8} \varepsilon^{1/8}.$$

We now prove the theorem.

Proof of Theorem 5.1. By definition $\mathbb{E}\, a^4 \leq 16\kappa^4$. Thus, we may apply Lemma 2.2 with

$$\alpha = C\sqrt{\kappa}\varepsilon^{1/8}$$

to complete the proof. ■

To prove the proposition, we proceed with similar steps to the proof of Theorem 1.1, starting with the following lemma.

Lemma 5.3. $|\langle \boldsymbol{v}_{\boldsymbol{x}}, \boldsymbol{x} \rangle - \lambda| = \left| \mathbb{E}\, |\langle \boldsymbol{a}, \boldsymbol{x} \rangle| - \sqrt{\frac{2}{\pi}} \right| \leq 4(\mathbb{E}\, a^4 + \mathbb{E}\, g^4)^{1/4} \varepsilon^{1/4}.$

Proof. We first prove a variant of the Berry-Esseen result on expectations, applying Lindeberg replacement. Define $S_i = \sum_{j=1}^{i-1} a_j x_j + \sum_{j=i+1}^n g_j x_j$, and $\phi(x)$ to be a bounded twice differentiable function. We will later use an approximation argument to replace ϕ by the absolute value function.

Note by telescoping,

$$|\mathbb{E}\,\phi(\langle \boldsymbol{a}, \boldsymbol{x}\rangle) - \mathbb{E}\,\phi(\langle \boldsymbol{g}, \boldsymbol{x}\rangle)| = \left|\mathbb{E}\,\phi\left(\sum_{i=1}^n a_i x_i\right) - \mathbb{E}\,\phi\left(\sum_{i=1}^n g_i x_i\right)\right|$$

$$\leq \sum_{i=1}^n |\,\mathbb{E}\,\phi(S_i + a_i x_i) - \mathbb{E}\,\phi(S_i + g_i x_i)|.$$

For convenience, dropping subscripts, we now wish to bound $|\,\mathbb{E}\,\phi(S+ax) - \mathbb{E}\,\phi(S+gx)|$.

By Taylor's theorem with remainder, we have

$$\phi(S+ax) = \phi(S) + ax\phi'(S) + R(S, ax)$$

where $|R(S,ax)| \leq (ax)^2 \|\phi''\|_\infty/2$. A similar result holds for $\phi(S+gx)$.

Split $R(S,x)$ into $R_+(S,x) \geq 0$ and $R_-(S,x) \geq 0$. Observe that since $\mathbb{E}\,a = \mathbb{E}\,g = 0$, the zeroth and first order terms cancel upon taking expectations in the difference

$$|\,\mathbb{E}\,\phi(S+ax) - \mathbb{E}\,\phi(S+gx)| = |\,\mathbb{E}\,R(S,ax) - \mathbb{E}\,R(S,gx))|$$

$$\leq |\,\mathbb{E}\,R_+(S,ax) - \mathbb{E}\,R_+(S,gx)| + |\,\mathbb{E}\,R_-(S,ax) - \mathbb{E}\,R_-(S,gx)|.$$

Consider the difference with R_+. We will apply the assumption $\|a-g\|_{TV} \leq \varepsilon$. First, observe that S is independent of both a and g and may be viewed as a constant. Viewing for instance $R_+(S,ax)$ as a function of a,

$$\left|\int_0^M P(R_+(S,ax) > t)dt - \int_0^M P(R_+(S,gx) > t)dt\right| \leq M\varepsilon.$$

Then, consider the tail of the first integral:

$$\int_M^\infty P(R_+(S,ax) > t)dt \leq \int_M^\infty \frac{\mathbb{E}(R_+(S,ax)^2)}{t^2}dt$$

$$= \frac{\mathbb{E}\,R_+(S,ax)^2}{M} \leq \frac{x^4\,\mathbb{E}\,a^4 \|\phi''\|_\infty^2}{4M}.$$

The Gaussian tail yields a similar error. Hence, optimizing over M by choosing

$$M = \frac{x^2(\mathbb{E}\,a^4 + \mathbb{E}\,g^4)^{1/2}\|\phi''\|_\infty}{2\sqrt{\varepsilon}}$$

we have an overall error of

$$|\,\mathbb{E}\,R_+(S,ax) - \mathbb{E}\,R_+(S,gx))| \leq x^2(\mathbb{E}\,a^4 + \mathbb{E}\,g^4)^{1/2}\|\phi''\|_\infty\sqrt{\varepsilon}.$$

The same holds for the difference with R_-. Finally, summing over the n indices, and using that $\|x\|_2 = 1$,

$$|\mathbb{E}\,\phi(\langle \boldsymbol{a}, \boldsymbol{x}\rangle) - \mathbb{E}\,\phi(\langle \boldsymbol{g}, \boldsymbol{x}\rangle)| \leq 2(\mathbb{E}\,a^4 + \mathbb{E}\,g^4)^{1/2}\|\phi''\|_\infty\sqrt{\varepsilon}.$$

Second, we approximate the absolute value using $\phi(x) := \sqrt{c+x^2} \approx |x|$. Observe for instance that $|\mathbb{E}|\langle \boldsymbol{a}, \boldsymbol{x}\rangle| - \mathbb{E}\,\phi(\langle \boldsymbol{a}, \boldsymbol{x}\rangle)| \leq \sqrt{c}$, and likewise with $\boldsymbol{g}$ in the place of $\boldsymbol{a}$. Evaluating $\phi''(x) = c/(c+x^2)^{3/2}$ with a maximum of $1/\sqrt{c}$ at $x=0$, we may conclude

$$|\langle \boldsymbol{v}_x, \boldsymbol{x}\rangle - \lambda| = \left| \mathbb{E}\,|\langle \boldsymbol{a}, \boldsymbol{x}\rangle| - \mathbb{E}\,|\langle \boldsymbol{g}, \boldsymbol{x}\rangle| \right| \leq 2\sqrt{c} + 2(\mathbb{E}\,a^4 + \mathbb{E}\,g^4)^{1/2}\sqrt{\frac{\varepsilon}{c}}.$$

Choosing $\sqrt{c} = (\mathbb{E}\,a^4 + \mathbb{E}\,g^4)^{1/4}\varepsilon^{1/4}$ completes the proof of the lemma. ■

We now proceed to bound $\|\boldsymbol{v}_x\|_2$, thus obtaining the second geometric constraint required in the proof of the proposition. We apply Lemma 5.3 with $\boldsymbol{z} = \boldsymbol{v}_x/\|\boldsymbol{v}_x\|_2$ in the place of $\boldsymbol{x}$:

$$\begin{aligned} \|\boldsymbol{v}_x\|_2 &= \langle \boldsymbol{v}_{\boldsymbol{x}}, \boldsymbol{v}_{\boldsymbol{x}}/\|\boldsymbol{v}_{\boldsymbol{x}}\|_2\rangle = \mathbb{E}\,\mathrm{sign}(\langle \boldsymbol{a}, \boldsymbol{x}\rangle)\langle \boldsymbol{a}, \boldsymbol{z}\rangle \\ &\leq \mathbb{E}\,\mathrm{sign}(\langle \boldsymbol{a}, \boldsymbol{z}\rangle)\langle \boldsymbol{a}, \boldsymbol{z}\rangle = \langle \boldsymbol{v}_z, \boldsymbol{z}\rangle \leq \lambda + 4(\mathbb{E}\,a^4 + \mathbb{E}\,g^4)^{1/4}\varepsilon^{1/4}. \end{aligned} \tag{5.1}$$

We conclude the proof with the following calculation.

$$\begin{aligned} |\langle \boldsymbol{v}_{\boldsymbol{x}}, \boldsymbol{x}'\rangle - \langle \lambda\boldsymbol{x}, \boldsymbol{x}'\rangle|^2 &\leq \|\boldsymbol{v}_{\boldsymbol{x}}\|_2^2 - \lambda^2 + 2\lambda(\lambda - \langle \boldsymbol{v}_{\boldsymbol{x}}, \boldsymbol{x}\rangle) \\ &= (\|\boldsymbol{v}_{\boldsymbol{x}}\|_2 + \lambda)(\|\boldsymbol{v}_{\boldsymbol{x}}\|_2 - \lambda) + 2\lambda(\lambda - \langle \boldsymbol{v}_{\boldsymbol{x}}, \boldsymbol{x}\rangle) \\ &\leq 16(\mathbb{E}\,a^4 + \mathbb{E}\,g^4)^{1/4}\varepsilon^{1/4}. \end{aligned}$$

The last inequality follows from Equation (5.1) and Lemmas 5.3 and 4.4. Proposition 5.2 is a consequence of absorbing constants.

5.3. **Total variation: smooth noise model.** We consider the setting of Theorem 1.3, with the additional assumption that $\|a - g\|_{TV} \leq \varepsilon$. We also relax the assumption on $\theta(t)$, defined as in (1.8), to $\theta(t) \in C^2$.

Theorem 5.4 (Estimating a signal with noise)**.** *We remain in the setting of Theorem 1.3 with the additional condition $\|a - g\|_{TV} \leq \varepsilon$, and also relax the condition on $\theta(t)$ to $\theta(t) \in C^2$. Then for each $\beta > 0$, with probability at least $1 - 4e^{-\beta^2}$, the solution $\widehat{\boldsymbol{x}}$ to the optimization problem (1.2) satisfies*

$$\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_2^2 \leq C\left((\kappa^3 + 1)(\tau_1 + \tau_2)\sqrt{\varepsilon} + \frac{\kappa}{\lambda\sqrt{m}}(w(K) + \beta)\right).$$

We bound α in Lemma 2.2.

Proposition 5.5 (Expectation)**.** *For $\boldsymbol{x} \in S^{n-1}, \boldsymbol{x}' \in B_2^n$,*

$$|\mathbb{E}\,f_{\boldsymbol{x}}(\boldsymbol{x}') - \langle \lambda\boldsymbol{x}, \boldsymbol{x}'\rangle| \leq 8(\mathbb{E}\,a^6 + \mathbb{E}\,g^6)^{1/2}(\tau_1 + \tau_2)\sqrt{\varepsilon}.$$

We now prove the theorem.

Proof of Theorem 5.4. By definition of κ we have $\mathbb{E}\,a^6 \leq 216\kappa^6$ which implies $\sqrt{\mathbb{E}\,a^6 + \mathbb{E}\,g^6} \leq C(\kappa^3 + 1)$. Thus, we apply Lemma 2.2 with

$$\alpha = C(\kappa^3 + 1)(\tau_1 + \tau_2)\sqrt{\varepsilon}$$

to complete the proof of the theorem. ■

We now prove the proposition.

Proof of Proposition 5.5. Recalling the steps in Section 3, observe that the left hand side of the inequality is expressible as

$$|\,\mathbb{E}\, f_{\boldsymbol{x}}(\boldsymbol{x}') - \langle \lambda \boldsymbol{x}, \boldsymbol{x}'\rangle| = |\,\mathbb{E}\,\theta(\langle \boldsymbol{a}, \boldsymbol{x}\rangle)\langle \boldsymbol{a}, \boldsymbol{x}'\rangle - \mathbb{E}\,\theta(\langle \boldsymbol{g}, \boldsymbol{x}\rangle)\langle \boldsymbol{g}, \boldsymbol{x}'\rangle|.$$

The statement of the proposition becomes similar to that of Lemma 3.1. Using the same notation and proceeding as in its proof (including the use of $\boldsymbol{z}$ in place of $\boldsymbol{x}'$), we apply Lindeberg replacement:

$$|\,\mathbb{E}\,\theta(\langle \boldsymbol{a}, \boldsymbol{x}\rangle)\langle \boldsymbol{a}, \boldsymbol{z}\rangle - \mathbb{E}\,\theta(\langle \boldsymbol{g}, \boldsymbol{x}\rangle)\langle \boldsymbol{g}, \boldsymbol{z}\rangle| \leq \sum_{i=1}^{n} |\,\mathbb{E}\,\phi(S_i + a_i v_i) - \mathbb{E}\,\phi(S_i + a_i v_i)|.$$

As before, we Taylor expand, except only to second order error:

$$\phi(S_i + a_i v_i) = \phi(S_i) + \sum_{|\alpha|=1} (a_i v_i)^{\alpha} \partial^{\alpha} \phi(S_i) + R(S_i, a_i v_i)$$

where $R(S_i, a_i v_i) = \frac{1}{2}\sum_{|\alpha|=2} (a_i v_i)^{\alpha} \partial^{\alpha}\phi(S_i')$ for some S_i' on the line segment joining S_i and $S_i + a_i v_i$. A similar result holds with $\phi(S_i + g_i v_i)$, with respective S_i''.

Split $R(S, v)$ into $R_+(S, v) \geq 0$ and $R_-(S, v) \geq 0$. Observe that since $\mathbb{E}\, a = \mathbb{E}\, g = 0$, the zeroth and first order terms cancel upon taking expectations in the difference

$$|\,\mathbb{E}\,\phi(S_i + a_i v_i) - \mathbb{E}\,\phi(S_i + g_i v_i)| = |\,\mathbb{E}\, R(S_i, a_i v_i) - \mathbb{E}\, R(S_i, g_i v_i))|$$

$$\leq |\,\mathbb{E}\, R_+(S_i, a_i v_i) - \mathbb{E}\, R_+(S_i, g_i v_i)| + |\,\mathbb{E}\, R_-(S_i, a_i v_i) - \mathbb{E}\, R_-(S_i, g_i v_i)|.$$

Consider the difference containing R_+. We will apply the assumption $\|a - g\|_{TV} \leq \varepsilon$. First, observe that S_i is independent of both a_i and g_i and may be viewed as a constant (by conditioning on it). Viewing for instance $R_+(S_i, a_i v_i)$ as a function of a_i,

$$\left| \int_0^M P(R_+(S_i, a_i v_i) > t)dt - \int_0^M P(R_+(S_i, g_i v_i) > t)dt \right| \leq M\varepsilon.$$

Then, consider the tail of the first integral:

$$\int_M^{\infty} P(R_+(S_i, a_i v_i) > t)dt \leq \int_M^{\infty} \frac{\mathbb{E}(R_+(S_i, a_i v_i)^2)}{t^2} dt = \frac{\mathbb{E}\, R_+(S_i, a_i v_i)^2}{M}.$$

Recall the explicit form of the remainder and observe that the partials in the error vanish except when at most one partial is taken on the second argument of ϕ, yielding either $\theta'(x)$ or $\theta''(x)z$. Furthermore, note that since S_i' is on the line segment joining S_i and $S_i + a_i v_i$, we may apply the bound $|(S_i')_2| \leq |(S_i)_2| + |a_i z_i|$ to conclude

$$\mathbb{E}\, 4R_+(S_i, a_i v_i)^2 \leq \mathbb{E}\, 4R(S_i, a_i v_i)^2 \leq \mathbb{E}\left(\sum_{|\alpha|=2} a_i^2 |v_i^{\alpha}| (\|\theta'\|_{\infty} + \|\theta''\|_{\infty}(|(S_i)_2| + |a_i z_i|)) \right)^2$$

$$= \mathbb{E}\left(a_i^2(|x_i| + |z_i|)^2 (\tau_1 + \tau_2(|(S_i)_2| + |z_i a_i|))\right)^2$$

Observe that $(S_i)_2$ and a_i are independent, and $(\mathbb{E}\,|(S_i)_2|)^2 \le \mathbb{E}(S_i)_2^2 \le 1$ by Cauchy-Schwarz and that the variance of an independent sum is a sum of variances. Further observing that $|z_i| \le 1$ and for instance $\mathbb{E}\,|a|^5 \le \mathbb{E}\,a^6$, rearranging and collecting terms yields

$$\mathbb{E}\,4R_+(S_i, a_i v_i)^2 \le (|x_i| + |z_i|)^4 \left(4\tau_2^2\,\mathbb{E}\,a_i^6 + \tau_1^2\,\mathbb{E}\,a_i^4 + 4\tau_1\tau_2\,\mathbb{E}\,|a_i|^5\right)$$

$$\le 4(|x_i| + |z_i|)^4(\tau_1 + \tau_2)^2\,\mathbb{E}\,a_i^6.$$

The Gaussian tail yields a similar error. Hence, optimizing over M by choosing

$$M = \frac{1}{\sqrt{\varepsilon}}(|x_i| + |z_i|)^2(\mathbb{E}\,a^6 + \mathbb{E}\,g^6)^{1/2}(\tau_1 + \tau_2)$$

we have overall error

$$|\,\mathbb{E}\,R_+(S_i, a_i v_i) - \mathbb{E}\,R_+(S_i, g_i v_i)| \le 2(|x_i| + |z_i|)^2(\mathbb{E}\,a^6 + \mathbb{E}\,g^6)^{1/2}(\tau_1 + \tau_2)\sqrt{\varepsilon}.$$

The same holds for the difference with R_-. Finally, summing over the n indices, and using that $\|\boldsymbol{x}\|_2 = 1$ and $\|\boldsymbol{z}\|_2 = 1$,

$$|\mathbb{E}\,\phi(\langle \boldsymbol{a}, \boldsymbol{x}\rangle) - \mathbb{E}\,\phi(\langle \boldsymbol{g}, \boldsymbol{x}\rangle)| \le 8(\mathbb{E}\,a^6 + \mathbb{E}\,g^6)^{1/2}(\tau_1 + \tau_2)\sqrt{\varepsilon},$$

which concludes the proof of the proposition. ■

References

[1] Boufounos, P. T., and Baraniuk, R. G. 1-Bit compressive sensing. In *42nd Annual Conference on Information Sciences and Systems (CISS)* (Mar. 2008).

[2] Eldar, C., and Kutyniok, G., Eds. *Compressed Sensing: Theory and applications.* Cambridge University Press, 2012.

[3] Gupta, A., Nowak, R., and Recht, B. Sample complexity for 1-bit compressed sensing and sparse classification. In *International Symposium on Information Theory (ISIT)* (2010), IEEE.

[4] Jacques, L., Laska, J. N., Boufounos, P. T., and Baraniuk, R. G. Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. Preprint. Available at `http://arxiv.org/abs/1104.3160`.

[5] Ledoux, M. *The concentration of measure phenomenon.* American Mathematical Society, Providence, 2001.

[6] Ledoux, M., and Talagrand, M. *Probability in Banach Spaces: isoperimetry and processes.* Springer-Verlag, Berlin, 1991.

[7] Plan, Y., and Vershynin, R. One-bit compressed sensing by linear programming. Preprint. Available at `http://arxiv.org/abs/1109.4299`.

[8] Plan, Y., and Vershynin, R. Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach. Preprint. Available at `http://arxiv.org/abs/1202.1212`.

[9] Strook, D. *Probability Theory: an analytic view*, second ed. Cambridge University Press, 1993.

[10] Tao, T., and Vu, V. Random matrices: The distribution of the smallest singular values. *Geometric And Functional Analysis 20*, 1 (2010), 260–297.

[11] Vershynin, R. Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing: Theory and Applications*, Y. Eldar and G. Kutyniok, Eds. Cambridge University Press, 2012.

Department of Mathematics, University of Michigan, 530 Church St., Ann Arbor, MI 48109, U.S.A.
E-mail address: {aflapan, yplan, romanv}@umich.edu
E-mail address: aai@princeton.edu