

Robust 1-bit Compressed Sensing and Sparse Logistic Regression: A Convex Programming Approach

Yaniv Plan and Roman Vershynin

Abstract—This paper develops theoretical results regarding noisy 1-bit compressed sensing and sparse binomial regression. We demonstrate that a single convex program gives an accurate estimate of the signal, or coefficient vector, for both of these models. We show that an s -sparse signal in \mathbb{R}^n can be accurately estimated from $m = O(s \log(n/s))$ single-bit measurements using a simple convex program. This remains true even if each measurement bit is flipped with probability nearly 1/2. Worst-case (adversarial) noise can also be accounted for, and uniform results that hold for all sparse inputs are derived as well. In the terminology of sparse logistic regression, we show that $O(s \log(2n/s))$ Bernoulli trials are sufficient to estimate a coefficient vector in \mathbb{R}^n which is approximately s -sparse. Moreover, the same convex program works for virtually all generalized linear models, in which the link function may be unknown. To our knowledge, these are the first results that tie together the theory of sparse logistic regression to 1-bit compressed sensing. Our results apply to general signal structures aside from sparsity; one only needs to know the size of the set K where signals reside. The size is given by the mean width of K , a computable quantity whose square serves as a robust extension of the dimension.

Index Terms—Compressed sensing, data compression, parameter estimation, quantization, regression analysis, signal reconstruction.

I. INTRODUCTION

A. One-bit Compressed Sensing

IN modern data analysis, a pervasive challenge is to recover extremely high-dimensional signals from seemingly inadequate amounts of data. Research in this direction is being conducted in several areas including compressed sensing, sparse approximation, and low-rank matrix recovery. The key is to take into account the signal structure, which in essence reduces the dimension of the signal space. In compressed sensing and sparse approximation, this structure is *sparsity*—we say that a vector in \mathbb{R}^n is s -sparse if it has s nonzero entries. In low-rank matrix recovery, one restricts to matrices with low rank.

The standard assumption in these fields is that one has access to *linear measurements* of the form

$$y_i = \langle \mathbf{a}_i, \mathbf{x} \rangle, \quad i = 1, 2, \dots, m \quad (\text{I.1})$$

where $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m \in \mathbb{R}^n$ are known measurement vectors and $\mathbf{x} \in \mathbb{R}^n$ is the signal to be recovered. Typical compressed

sensing results state that when \mathbf{a}_i are iid random vectors drawn from a certain distribution (e.g., Gaussian), $m \sim s \log(2n/s)$ measurements suffice for robust recovery of s -sparse signals \mathbf{x} (see [8]).

In the recently introduced problem of *1-bit compressed sensing* [5], the measurements are no longer linear but rather consist of single bits. If there is no noise, the measurements are modeled as

$$y_i = \text{sign}(\langle \mathbf{a}_i, \mathbf{x} \rangle), \quad i = 1, 2, \dots, m \quad (\text{I.2})$$

where $\text{sign}(x) = 1$ if $x \geq 0$ and $\text{sign}(x) = -1$ if $x < 0$.¹ On top of this, noise may be introduced as random or adversarial bit flips.

The 1-bit measurements are meant to model quantization in the extreme case. It is interesting to note that when the signal-to-noise ratio is low, numerical experiments demonstrate that such extreme quantization can be optimal [17] when constrained to a fixed bit budget. The webpage <http://dsp.rice.edu/1bitCS/> is dedicated to the rapidly growing literature on 1-bit compressed sensing. Further discussion of this recent literature will be given in Section III-A; we note for now that this paper presents the first theoretical accuracy guarantees in the noisy problem using a polynomial-time solver (given by a convex program).

B. Noisy 1-bit Measurements

We propose the following general model for noisy 1-bit compressed sensing. We assume that the measurements, or response variables, $y_i \in \{-1, 1\}$, are drawn independently at random satisfying

$$\mathbb{E} y_i = \theta(\langle \mathbf{a}_i, \mathbf{x} \rangle), \quad i = 1, 2, \dots, m \quad (\text{I.3})$$

where θ is some function, which automatically must satisfy $-1 \leq \theta(z) \leq 1$. A key point in our results is that θ may be *unknown* or unspecified; one only needs to know the measurements y_i and the measurement vectors \mathbf{a}_i in order to recover \mathbf{x} . Thus, there is an unknown nonlinearity in the measurements. See [4] and [16] for earlier connections between the 1-bit problem and nonlinear measurements.

In compressed sensing, it is typical to choose the measurement vectors \mathbf{a}_i at random (see [8]). In this paper, we choose \mathbf{a}_i to be independent standard Gaussian random vectors in \mathbb{R}^n . Although this assumption can be relaxed to allow for correlated coordinates (see Section III-D), discrete distributions are not permitted. Indeed, unlike traditional compressed sensing, accurate noiseless 1-bit compressed sensing is provably impossible

¹For concreteness, we set $\text{sign}(0) = 1$; this choice is arbitrary and could be replaced with $\text{sign}(0) = -1$.

Manuscript received February 05, 2012; accepted June 22, 2012. Date of publication September 04, 2012; date of current version December 19, 2012. Y. Plan was supported by a National Science Foundation (NSF) Postdoctoral Research Fellowship under Award 1103909. R. Vershynin was supported by the NSF under Grants DMS 0918623 and 1001829.

The authors are with the Department of Mathematics, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: yplan@umich.edu; romanv@umich.edu).

Communicated by M. Elad, Associate Editor for Signal Processing.

Digital Object Identifier 10.1109/TIT.2012.2207945

for some discrete distributions of \mathbf{a}_i (e.g., for Bernoulli distribution, see [23]). Summarizing, the model (I.3) has two sources of randomness:

- 1) the measurement vectors \mathbf{a}_i are independent standard Gaussian random vectors;
- 2) given $\{\mathbf{a}_i\}$, the measurements y_i are independent $\{-1, 1\}$ valued random variables.

Note that (I.3) is the *generalized linear model* in statistics, and θ is known as the inverse of the *link function*; the particular choice $\theta(z) = \tanh(z/2)$ corresponds to logistic regression. The statisticians may prefer to switch \mathbf{x} with $\boldsymbol{\beta}$, \mathbf{a}_i with \mathbf{x}_i , n with p and m with n , but we prefer to keep our notation which is standard in compressed sensing.

Notice that in the noiseless 1-bit compressed sensing model (I.2), all information about the magnitude of \mathbf{x} is lost in the measurements. Similarly, in the noisy model (I.3), the magnitude of \mathbf{x} may be absorbed into the definition of θ . Thus, our goal will be to estimate the projection of \mathbf{x} onto the Euclidean sphere, $\mathbf{x}/\|\mathbf{x}\|_2$. Without loss of generality, we thus assume that $\|\mathbf{x}\|_2 = 1$ in most of our discussion that follows.

We shall make a single assumption on the function θ defining the model (I.3), namely that

$$\mathbb{E}\theta(g)g =: \lambda > 0 \quad (\text{I.4})$$

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

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

Thus our assumption is simply that the 1-bit measurements y_i are positively correlated with the corresponding linear measurements $\langle \mathbf{a}_i, \mathbf{x} \rangle$.² Standard 1-bit compressed sensing (I.2) is a partial case of model (I.3) with $\theta(z) = \text{sign}(z)$. In this case, λ achieves its maximal value: $\lambda = \mathbb{E}|g| = \sqrt{2/\pi}$. In general, λ plays a role similar to a signal-to-noise ratio.

C. Signal Set

To describe the structure of possible signals, we assume that \mathbf{x} lies in some set $K \subset B_2^n$ where B_2^n denotes the Euclidean unit ball in \mathbb{R}^n . A key characteristic of the size of the signal set is its *mean width* $w(K)$, defined as

$$w(K) := \mathbb{E} \sup_{\mathbf{x} \in K - K} \langle \mathbf{g}, \mathbf{x} \rangle \quad (\text{I.5})$$

where \mathbf{g} is a standard normal Gaussian vector in \mathbb{R}^n and $K - K$ denotes the Minkowski difference.³ The notion of mean width is closely related to that of the Gaussian complexity, which is widely used in statistical learning theory to measure the size of classes of functions (see [3] and [19]). An intuitive explanation of the mean width, its basic properties, and simple examples are

²If $\mathbb{E}\theta(g)g < 0$, we could replace y_i with $-y_i$ to change the sign; thus our assumption is really that the correlation is nonzero: $\mathbb{E}\theta(g)g \neq 0$.

³Specifically, $K - K = \{\mathbf{x} - \mathbf{y} : \mathbf{x}, \mathbf{y} \in K\}$.

given in Section II. The important point is that $w(K)^2$ can serve as the *effective dimension* of K .

The main example of interest is where K encodes sparsity. If $K = K_{n,s}$ is the convex hull of the unit s -sparse vectors in \mathbb{R}^n , the mean width of this set computed in (II.2) and (III.3) is

$$w(K_{n,s}) \sim (s \log(2n/s))^{1/2}. \quad (\text{I.6})$$

D. Main Results

We propose the following solver to estimate the signal \mathbf{x} from the 1-bit measurements y_i . It is given by the optimization problem

$$\max \sum_{i=1}^m y_i \langle \mathbf{a}_i, \mathbf{x}' \rangle \quad \text{subject to} \quad \mathbf{x}' \in K. \quad (\text{I.7})$$

This can be described even more compactly as

$$\max \langle \mathbf{y}, \mathbf{A}\mathbf{x}' \rangle \quad \text{subject to} \quad \mathbf{x}' \in K \quad (\text{I.8})$$

where \mathbf{A} is the $m \times n$ measurement matrix with rows \mathbf{a}_i and $\mathbf{y} = (y_1, \dots, y_m)$ is the vector of 1-bit measurements.

If the set K is convex, (I.7) is a *convex program*, and therefore, it can be solved in an algorithmically efficient manner. This is the situation we will mostly care about, although our results below apply for general, nonconvex signal sets K as well.

Theorem 1.1 (Fixed Signal Estimation, Random Noise): Let $\mathbf{a}_1, \dots, \mathbf{a}_m$ be independent standard Gaussian random vectors in \mathbb{R}^n , and let K be a subset of the unit Euclidean ball in \mathbb{R}^n . Fix $\mathbf{x} \in K$ satisfying $\|\mathbf{x}\|_2 = 1$. Assume that the measurements y_1, \dots, y_m follow the model above.⁴ Then for each $\beta > 0$, with probability at least $1 - 4 \exp(-2\beta^2)$ the solution $\hat{\mathbf{x}}$ to the optimization problem (I.7) satisfies

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq \frac{8}{\lambda \sqrt{m}} (w(K) + \beta).$$

As an immediate consequence, we see that *the signal $\mathbf{x} \in K$ can be effectively estimated from $m = O(w(K)^2)$ 1-bit noisy measurements*. The following result makes this statement precise.

Corollary 1.2 (Number of Measurements): Let $\delta > 0$ and suppose that

$$m \geq C\delta^{-2}w(K)^2.$$

Then, under the assumptions of Theorem 1.1, with probability at least $1 - 8 \exp(-c\delta^2 m)$, the solution $\hat{\mathbf{x}}$ to the optimization problem (I.7) satisfies

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq \delta/\lambda.$$

Here and in the rest of this paper, C and c denote positive absolute constants whose values may change from instance to instance.

⁴Specifically, our assumptions are that y_i are $\{-1, 1\}$ valued random variables that are independent given $\{\mathbf{a}_i\}$, and that (I.3) holds with some function θ satisfying (I.4).

Theorem 1.1 is concerned with an arbitrary but fixed signal $\mathbf{x} \in K$, and with a stochastic model on the noise in the measurements. We will show how to strengthen these results to cover all signals $\mathbf{x} \in K$ uniformly, and to allow for worst-case (adversarial) noise. Such noise can be modeled as flipping some fixed percentage of arbitrarily chosen bits, and it can be measured using the Hamming distance $d_H(\tilde{\mathbf{y}}, \mathbf{y}) = \sum_{i=1}^m \mathbf{1}_{\{\tilde{y}_i \neq y_i\}}$ between $\tilde{\mathbf{y}}, \mathbf{y} \in \{-1, 1\}^m$.

We present the following theorem in the same style as Corollary 1.2.

Theorem 1.3 (Uniform Estimation, Adversarial Noise): Let $\mathbf{a}_1, \dots, \mathbf{a}_m$ be independent standard Gaussian random vectors in \mathbb{R}^n , and let K be a subset of the unit Euclidean ball in \mathbb{R}^n . Let $\delta > 0$ and suppose that

$$m \geq C\delta^{-6}w(K)^2. \quad (\text{I.9})$$

Then with probability at least $1 - 8\exp(-c\delta^2m)$, the following event occurs. Consider a signal $\mathbf{x} \in K$ satisfying $\|\mathbf{x}\|_2 = 1$ and its (unknown) uncorrupted 1-bit measurements $\tilde{\mathbf{y}} = (\tilde{y}_1, \dots, \tilde{y}_m)$ given as

$$\tilde{y}_i = \text{sign}(\langle \mathbf{a}_i, \mathbf{x} \rangle), \quad i = 1, 2, \dots, m.$$

Let $\mathbf{y} = (y_1, \dots, y_m) \in \{-1, 1\}^m$ be any (corrupted) measurements satisfying $d_H(\tilde{\mathbf{y}}, \mathbf{y}) \leq \tau m$. Then, the solution $\hat{\mathbf{x}}$ to the optimization problem (I.7) with input \mathbf{y} satisfies

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq \delta \sqrt{\log(e/\delta)} + 11\tau \sqrt{\log(e/\tau)}. \quad (\text{I.10})$$

This uniform result will follow from a deeper analysis than the fixed-signal result, Theorem 1.1. Its proof will be based on the recent results from [22] on random hyperplane tessellations of K .

Remark 1.4 (Sparse Estimation): A remarkable example is for s -sparse signals in \mathbb{R}^n . Recalling the mean width estimate (I.6), we see that our results above imply that *an s -sparse signal in \mathbb{R}^n can be effectively estimated from $m = O(s \log(2n/s))$ 1-bit noisy measurements.* We will make this statement precise in Corollary 3.1 and the remark after it.

Remark 1.5 (Hamming Cube Encoding and Decoding): Let us put Theorem 1.3 in the context of coding in information theory. In the earlier paper [22], we proved that $K \cap S^{n-1}$ can be almost isometrically embedded into the Hamming cube $\{-1, 1\}^m$, with the same m and same probability bound as in Theorem 1.3. Specifically, one has

$$\left| \frac{1}{\pi} d_G(\mathbf{x}, \mathbf{x}') - \frac{1}{m} d_H(\text{sign}(\mathbf{A}\mathbf{x}), \text{sign}(\mathbf{A}\mathbf{x}') \right| \leq \delta \quad (\text{I.11})$$

for all $\mathbf{x}, \mathbf{x}' \in K \cap S^{n-1}$. Above, d_G and d_H denote the geodesic distance in S^{n-1} and the Hamming distance in $\{-1, 1\}^m$, respectively (see Theorem 6.3 below). Thus, the embedding $K \cap S^{n-1} \rightarrow \{-1, 1\}^m$ is given by the map⁵

$$\mathbf{x} \mapsto \text{sign}(\mathbf{A}\mathbf{x}).$$

⁵The sign function is applied to each coordinate of $\mathbf{A}\mathbf{x}$.

This map *encodes* a given signal $\mathbf{x} \in K$ into a binary string $\mathbf{y} = \text{sign}(\mathbf{A}\mathbf{x})$. Conversely, one can accurately and robustly *decode* \mathbf{x} from \mathbf{y} by solving the optimization problem (I.7). This is the content of Theorem 1.3.

Remark 1.6 (Optimality): While the dependence of m on the mean width $w(K)$ in the results above seems to be optimal (see [22] for a discussion), the dependence on the accuracy δ in Theorem 1.3 is most likely not optimal. We are not trying to optimize dependence on δ in this paper, but are leaving this as an open problem. Nevertheless, in some cases, the dependence on δ in Theorem 1.1 is optimal; see Section III-A below.

Theorem 1.3 can be extended to allow for a random noise together with adversarial noise; this is discussed in Remark 4.4 below.

E. Organization

An intuitive discussion of the mean width along with the estimate (I.6) of the mean width when K encodes sparse vectors is given in Section II. In Section III, we specialize our results to a variety of (approximately) sparse signal models—1-bit compressed sensing, sparse logistic regression and low-rank matrix recovery. In Section III-D, we extend our results to allow for correlations in the entries of the measurement vectors.

The proofs of our main results, Theorems 1.1 and 1.3, are given in Sections IV–VI. In Section IV, we quickly reduce these results to the two concentration inequalities that hold uniformly over the set K —Propositions 4.2 and 4.3, respectively. Proposition 4.2 is proved in Section V using standard techniques of probability in Banach spaces. The proof of Proposition 4.3 is deeper; it is based on the recent work of the authors [22] on random hyperplane tessellations. The argument is given in Section VI.

F. Notation

We write $a \sim b$ if $ca \leq b \leq Ca$ for some positive absolute constants c, C (a and b may have dimensional dependence). In order to increase clarity, vectors are written in lower case bold italics (e.g., \mathbf{g}), and matrices are upper case bold italics (e.g., \mathbf{A}). We let \mathbf{g} denote a standard Gaussian random vector whose length will be clear from context; g denotes a standard normal random variable. C, c will denote positive absolute constants whose values may change from instance to instance. Given a vector \mathbf{v} in \mathbb{R}^n and a subset $T \subset \{1, \dots, n\}$, we denote by $\mathbf{v}_T \in \mathbb{R}^T$ the restriction of \mathbf{v} onto the coordinates in T .

B_2^n and S^{n-1} denote the unit Euclidean ball and sphere in \mathbb{R}^n , respectively, and B_1^n denotes the unit ball with respect to ℓ_1 norm. The Euclidean and ℓ_1 norms of a vector \mathbf{v} are denoted $\|\mathbf{v}\|_2$ and $\|\mathbf{v}\|_1$, respectively. The number of nonzero entries of \mathbf{v} is denoted $\|\mathbf{v}\|_0$. The operator norm (the largest singular value) of a matrix \mathbf{A} is denoted $\|\mathbf{A}\|$.

II. MEAN WIDTH AND SPARSITY

A. Mean Width

In this section, we explain the geometric meaning of the mean width of a set $K \subset \mathbb{R}^n$ which was defined by the formula (I.5), and discuss its basic properties and examples.

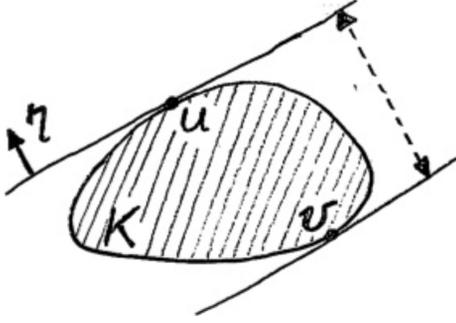


Fig. 1. Width of a set K in the direction of $\boldsymbol{\eta}$ is illustrated by the dashed line.

The notion of mean width plays a significant role in asymptotic convex geometry (see e.g., [9]). The width of K in the direction of $\boldsymbol{\eta} \in S^{n-1}$ is the smallest width of the slab between two parallel hyperplanes with normals $\boldsymbol{\eta}$ that contains K . Analytically, the width can be expressed as

$$\sup_{\mathbf{u} \in K} \langle \boldsymbol{\eta}, \mathbf{u} \rangle - \inf_{\mathbf{v} \in K} \langle \boldsymbol{\eta}, \mathbf{v} \rangle = \sup_{\mathbf{x} \in K-K} \langle \boldsymbol{\eta}, \mathbf{x} \rangle$$

(see Fig. 1). Averaging over $\boldsymbol{\eta}$ uniformly distributed in S^{n-1} , we obtain the *spherical mean width*:

$$\tilde{w}(K) := \mathbb{E} \sup_{\mathbf{x} \in K-K} \langle \boldsymbol{\eta}, \mathbf{x} \rangle.$$

Instead of averaging using $\boldsymbol{\eta} \in S^{n-1}$, it is often more convenient to use a standard Gaussian random vector $\mathbf{g} \in \mathbb{R}^n$. This gives the definition (I.5) of the *Gaussian mean width* of K :

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

In this paper, we shall use the Gaussian mean width, which we call the “mean width” for brevity. Note that the spherical and Gaussian versions of mean width are proportional to each other. Indeed, by rotation invariance, we can realize $\boldsymbol{\eta}$ as $\boldsymbol{\eta} = \mathbf{g}/\|\mathbf{g}\|_2$ and note that $\boldsymbol{\eta}$ is independent of the magnitude factor $\|\mathbf{g}\|_2$. It follows that $w(K) = \mathbb{E}\|\mathbf{g}\|_2 \cdot \tilde{w}(K)$. Further, once can use that $\mathbb{E}\|\mathbf{g}\|_2 \sim \sqrt{n}$ and obtain the useful comparison of Gaussian and spherical versions of mean width:

$$w(K) \sim \sqrt{n} \cdot \tilde{w}(K).$$

Let us record some further simple but useful properties of the mean width.

Proposition 2.1 (Mean Width): The mean width of a subset $K \subset \mathbb{R}^n$ has the following properties.

- 1) The mean width is invariant under orthogonal transformations and translations.
- 2) The mean width is invariant under taking the convex hull, i.e., $w(K) = w(\text{conv}(K))$.
- 3) We have

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

- 4) Denoting the diameter of K in the Euclidean metric by $\text{diam}(K)$, we have

$$\sqrt{\frac{2}{\pi}} \text{diam}(K) \leq w(K) \leq n^{1/2} \text{diam}(K).$$

- 5) We have

$$w(K) \leq 2\mathbb{E} \sup_{\mathbf{x} \in K} \langle \mathbf{g}, \mathbf{x} \rangle \leq 2\mathbb{E} \sup_{\mathbf{x} \in K} |\langle \mathbf{g}, \mathbf{x} \rangle|.$$

For an origin-symmetric set K , both these inequalities become equalities.

- 6) The inequalities in part 5 can be essentially reversed for arbitrary K :

$$w(K) \geq \mathbb{E} \sup_{\mathbf{x} \in K} |\langle \mathbf{g}, \mathbf{x} \rangle| - \sqrt{\frac{2}{\pi}} \text{dist}(0, K).$$

Here, $\text{dist}(0, K) = \inf_{\mathbf{x} \in K} \|\mathbf{x}\|_2$ is the Euclidean distance from the origin to K . In particular, if $0 \in K$ then one has $w(K) \geq \mathbb{E} \sup_{\mathbf{x} \in K} |\langle \mathbf{g}, \mathbf{x} \rangle|$.

Proof: Parts 1, 2, and 5 are obvious by definition; part 3 follows by the symmetry of $K - K$.

To prove part 4, note that for every $\mathbf{x}_0 \in K - K$, one has

$$w(K) \geq \mathbb{E} |\langle \mathbf{g}, \mathbf{x}_0 \rangle| = \sqrt{\frac{2}{\pi}} \|\mathbf{x}_0\|_2. \quad (\text{II.1})$$

The equality here follows because $\langle \mathbf{g}, \mathbf{x}_0 \rangle$ is a normal random variable with variance $\|\mathbf{x}_0\|_2$. This yields the lower bound in part 4. For the upper bound, we can use part 3 along with the bound $|\langle \mathbf{g}, \mathbf{x} \rangle| \leq \|\mathbf{g}\|_2 \|\mathbf{x}\|_2 \leq \|\mathbf{g}\|_2 \cdot \text{diam}(K)$ for all $\mathbf{x} \in K - K$. This gives

$$\begin{aligned} w(K) &\leq \mathbb{E} \|\mathbf{g}\|_2 \cdot \text{diam}(K) \leq (\mathbb{E} \|\mathbf{g}\|_2^2)^{1/2} \text{diam}(K) \\ &= n^{1/2} \text{diam}(K). \end{aligned}$$

To prove part 6, let us start with the special case where $0 \in K$. Then $K - K \supset K \cup (-K)$; thus $w(K) \geq \mathbb{E} \sup_{\mathbf{x} \in K \cup (-K)} \langle \mathbf{g}, \mathbf{x} \rangle = \mathbb{E} \sup_{\mathbf{x} \in K} |\langle \mathbf{g}, \mathbf{x} \rangle|$ as claimed. Next, consider the general case. Fix $\mathbf{x}_0 \in K$ and apply the previous reasoning for the set $K - \mathbf{x}_0 \ni 0$. Using parts 1 and 3, we obtain

$$\begin{aligned} w(K) &= w(K - \mathbf{x}_0) \geq \sup_{\mathbf{x} \in K} |\langle \mathbf{g}, \mathbf{x} - \mathbf{x}_0 \rangle| \\ &\geq \sup_{\mathbf{x} \in K} |\langle \mathbf{g}, \mathbf{x} \rangle| - |\langle \mathbf{g}, \mathbf{x}_0 \rangle|. \end{aligned}$$

Finally, as in (II.1), we note that $\mathbb{E} |\langle \mathbf{g}, \mathbf{x}_0 \rangle| = \sqrt{\frac{2}{\pi}} \|\mathbf{x}_0\|_2$. Minimizing over $\mathbf{x}_0 \in K$ completes the proof. \blacksquare

Example: For illustration, let us evaluate the mean width of some sets $K \subseteq B_2^n$.

- 1) If $K = B_2^n$ or $K = S^{n-1}$ then $w(K) = \mathbb{E} \|\mathbf{g}\|_2 \leq (\mathbb{E} \|\mathbf{g}\|_2^2)^{1/2} = \sqrt{n}$ (and in fact $w(K) \sim \sqrt{n}$).
- 2) If the linear algebraic dimension $\dim(K) = k$, then $w(K) \leq \sqrt{k}$.
- 3) If K is a finite set, then $w(K) \leq C \sqrt{\log |K|}$.

Remark 2.2 (Effective Dimension): The square of the mean width, $w(K)^2$, may be interpreted as the *effective dimension* of a set $K \subseteq B_2^n$. It is always bounded by the linear algebraic dimension (see the example above), but it has the advantage of robustness—a small perturbation of K leads to a small change in $w(K)^2$.

In this light, the invariance of the mean width under taking the convex hull (Proposition 2.1, part 2) is especially useful in compressed sensing, where a usual tactic is to relax the non-convex program to a convex program. It is important that in the course of this relaxation, the “effective dimension” of the signal set K remains the same.

Mean width of a given set K can be computed using several tools from probability in Banach spaces. These include Dudley’s inequality, Sudakov minoration, the Gaussian concentration inequality, Slepian’s inequality, and the sharp technique of majorizing measures and generic chaining [19], [26].

B. Sparse Signal Set

The quintessential signal structure considered in this paper is sparsity. Thus, for given $n \in \mathbb{N}$ and $0 < s \leq n$, we consider the set

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

In words, $S_{n,s}$ consists of s -sparse (or sparser) vectors with length n whose Euclidean norm is bounded by 1.

Although the linear algebraic dimension of $S_{n,s}$ is n (as this set spans \mathbb{R}^n), the dimension of $S_{n,s} \cap \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_0 = s\}$ as a manifold with boundary embedded in \mathbb{R}^n is s .⁶ It turns out that the “effective dimension” of $S_{n,s}$ given by the square of its mean width is much closer to the manifold dimension s than to the linear algebraic dimension n :

2.3 (Mean Width of the Sparse Signal Set): We have

$$cs \log(2n/s) \leq w^2(S_{n,s}) \leq Cs \log(2n/s). \quad (\text{II.2})$$

Proof: Let us prove the upper bound. Without loss of generality, we can assume that $s \in \mathbb{N}$. By representing $S_{n,s}$ as the union of $\binom{n}{s}$ -dimensional unit Euclidean balls, we see that

$$w(S_{n,s}) = \mathbb{E} \max_{|T|=s} \|\mathbf{g}_T\|_2.$$

For each T , the Gaussian concentration inequality (see Theorem 5.2 below) yields

$$\mathbb{P}\{\|\mathbf{g}_T\|_2 \geq \|\mathbf{g}_T\|_2 + t\} \leq \exp(-t^2/2), \quad t > 0.$$

Next, $\mathbb{E} \|\mathbf{g}_T\|_2 \leq (\|\mathbf{g}_T\|_2^2)^{1/2} = \sqrt{s}$. Thus, the union bound gives

$$\mathbb{P}\left\{\max_{|T|=s} \|\mathbf{g}_T\|_2 \geq \sqrt{s} + t\right\} \leq \binom{n}{s} \exp(-t^2/2)$$

⁶Thus, $S_{n,s}$ is the union of $s+1$ manifolds with boundary each of whose dimension is bounded by s .

for $t > 0$. Note that $\binom{n}{s} \leq \exp(s \log(en/s))$; integrating gives the desired upper bound in (II.2).

The lower bound in (II.2) follows from Sudakov minoration (see Theorem 6.1) combined with finding a tight lower bound on the covering number of $S_{n,s}$. Since the lower bound will not be used in this paper, we leave the details to the interested reader. ■

III. APPLICATIONS TO SPARSE SIGNAL MODELS

Our main results stated in the introduction are valid for general signal sets K . Now we specialize to the cases where K encodes *sparsity*. It would be ideal if we could take $K = S_{n,s}$, but this set would not be convex and thus the solver (I.7) would not be known to run in polynomial time. We instead take a *convex relaxation* of $S_{n,s}$, an effective tactic from the sparsity literature. Notice that if $\mathbf{x} \in S_{n,s}$, then $\|\mathbf{x}\|_1 \leq \sqrt{s}$ by Cauchy–Schwarz inequality. This motivates us to consider the convex set

$$K_{n,s} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 \leq 1, \|\mathbf{x}\|_1 \leq \sqrt{s}\} = B_2^n \cap \sqrt{s}B_1^n.$$

$K_{n,s}$ is almost exactly the convex hull of $S_{n,s}$, as shown in [23]:

$$\text{conv}(S_{n,s}) \subset K_{n,s} \subset 2\text{conv}(S_{n,s}). \quad (\text{III.1})$$

$K_{n,s}$ can be thought of as a set of *approximately sparse* or *compressible* vectors.

If the signal is known to be exactly or approximately sparse, i.e., $\mathbf{x} \in K_{n,s}$, we may estimate \mathbf{x} by solving the convex program

$$\begin{aligned} & \max \sum_{i=1}^m y_i \langle \mathbf{a}_i, \mathbf{x}' \rangle \\ & \text{subject to } \|\mathbf{x}'\|_1 \leq \sqrt{s} \quad \text{and} \quad \|\mathbf{x}'\|_2 \leq 1. \end{aligned} \quad (\text{III.2})$$

This is just a restatement of the program (I.7) for the set $K_{n,s}$. In our convex relaxation, we do not require that $\hat{\mathbf{x}} \in S_{n,s}$; this stands in contrast to many previous programs considered in the literature. Nevertheless, the accuracy of the solution $\hat{\mathbf{x}}$ and the fact that $\|\mathbf{x}\|_2 = 1$ implies that $\|\mathbf{x}\|_2 \approx 1$.

Theorems 1.1 and 1.3 are supposed to guarantee that \mathbf{x} can indeed be estimated by a solution to (III.2). But in order to apply these results, we need to know the mean width of $K_{n,s}$. A good bound for it follows from (III.1) and Lemma 2.3, which give

$$w(K_{n,s}) \leq 2w(\text{conv}(S_{n,s})) \leq C\sqrt{s \log(2n/s)}. \quad (\text{III.3})$$

This yields the following version of Corollary 1.2.

Corollary 3.1 (Estimating a Compressible Signal): Let $\mathbf{a}_1, \dots, \mathbf{a}_m$ be independent standard Gaussian random vectors in \mathbb{R}^n , and fix $\mathbf{x} \in K_{n,s}$ satisfying $\|\mathbf{x}\|_2 = 1$. Assume that the measurements y_1, \dots, y_m follow the model from Section I-C.7. Let $\delta > 0$ and suppose that

$$m \geq C\delta^{-2}s \log(2n/s).$$

⁷Specifically, our assumptions were that $\{y_i\}$ are independent random variables that are jointly independent of $\{\mathbf{a}_i\}$, and that (I.3) holds with some function θ satisfying (I.4).

Then, with probability at least $1 - 8 \exp(-c\delta^2 m)$, the solution $\hat{\mathbf{x}}$ to the convex program (III.2) satisfies

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq \delta/\lambda.$$

Remark 3.2: In a similar way, one can also specialize the uniform result, Theorem 1.3, to the approximately sparse case.

A. One-bit Compressed Sensing

Corollary 3.1 can be easily specialized to various specific models of noise. Let us consider some of the interesting models, and compute the correlation coefficient $\lambda = \mathbb{E}\theta(g)g$ in (I.4) for each of them.

Noiseless 1-bit compressed sensing: In the classic noiseless model (I.2), the measurements are given as $y_i = \text{sign}(\langle \mathbf{a}_i, \mathbf{x} \rangle)$ and thus $\theta(z) = \text{sign}(z)$. Thus

$$\lambda = \mathbb{E}|g| = \sqrt{2/\pi}.$$

Therefore, with high probability, we obtain $\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq \delta$ provided that the number of measurements is $m \geq C\delta^{-2}s \log(2n/s)$. This is similar to the results available in [23].

Random bit flips: Assume that each measurement y_i is only correct with probability p ; thus

$$y_i = \xi_i \text{sign}(\langle \mathbf{a}_i, \mathbf{x} \rangle), \quad i = 1, 2, \dots, m$$

where ξ_i are independent $\{-1, 1\}$ valued random variables with $\mathbb{P}\{\xi_i = 1\} = p$, which represent random bit flips. Then, $\theta(z) = \text{sign}(z) \cdot \mathbb{E}\xi_1 = 2\text{sign}(z)(p - 1/2)$ and

$$\lambda = 2(p - 1/2)\mathbb{E}|g| = 2\sqrt{2/\pi}(p - 1/2).$$

Therefore, with high probability, we obtain $\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq \delta$ provided that the number of measurements is $m \geq C\delta^{-2}(p - 1/2)^{-2}s \log(2n/s)$. Thus, we obtain a surprising conclusion:

The signal \mathbf{x} can be estimated even if each measurement is flipped with probability nearly 1/2.

Somewhat surprisingly, the estimation of \mathbf{x} is done by one simple convex program (III.2). Of course, if each measurement is corrupted with probability 1/2, recovery is impossible by any algorithm.

Random noise before quantization: Assume that the measurements are given as

$$y_i = \text{sign}(\langle \mathbf{a}_i, \mathbf{x} \rangle + \nu_i), \quad i = 1, 2, \dots, m$$

where ν_i are iid random variables representing noise added before quantization. This situation is typical in analog-to-digital converters. It is also the latent variable model from statistics.

Assume for simplicity that ν_i have density $f(x)$. Then $\theta(z) = 1 - 2\mathbb{P}\{\nu_i \leq -z\}$, and the correlation coefficient $\lambda = \theta(g)g$ can be evaluated using integration by parts, which gives

$$\lambda = \theta'(g) = 2f(-g) > 0.$$

A specific value of λ is therefore not hard to estimate for concrete densities f . For instance, if ν_i are normal random variables with mean zero and variance σ^2 , then

$$\lambda = \sqrt{\frac{2}{\pi\sigma^2}} \exp(-g^2/2\sigma^2) = \sqrt{\frac{2}{\pi(\sigma^2 + 1)}}.$$

Therefore, with high probability, we obtain $\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq \delta$ provided that the number of measurements is $m \geq C\delta^{-2}(\sigma^2 + 1)s \log(2n/s)$. In other words

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq C \sqrt{\frac{(\sigma^2 + 1)s \log(2n/s)}{m}}. \quad (\text{III.4})$$

Thus, we obtain an unexpected conclusion:

The signal \mathbf{x} can be estimated even when the noise level σ eclipses the magnitude of the linear measurements.

Indeed, the average magnitude of the linear measurements is $\mathbb{E}|\langle \mathbf{a}_i, \mathbf{x} \rangle| = \sqrt{2/\pi}$, while the average noise level σ can be much larger.

Let us also compare to the results available in the standard unquantized compressed sensing model

$$y_i = \langle \mathbf{a}_i, \mathbf{x} \rangle + \nu_i, \quad i = 1, 2, \dots, m$$

where once again we take $\nu_i \sim \mathcal{N}(0, \sigma^2)$. Under the assumption that $\|\mathbf{x}\|_1 \leq \sqrt{s}$, the minimax squared error given by [24, Th. 1] is $\delta = c\sigma\sqrt{\frac{s \log n}{m}}$. A slight variation on their proof yields a minimax squared error under the assumption that $m < n$ and $\mathbf{x} \in K_{n,s} \cap S^{n-1}$ of $\delta = c\sigma\sqrt{\frac{s \log(2n/s)}{m}}$. Up to a constant and when $\sigma > 1$, this matches the upper bound we have just derived in the 1-bit case in (III.4). Thus, we have another surprising result:

The error in recovering the signal \mathbf{x} matches the minimax error for the unquantized compressed sensing problem (up to a constant): Essentially nothing is lost by quantizing to a single bit.

Let us put these results in a perspective of the existing literature on 1-bit compressed sensing. The problem of 1-bit compressed sensing, as introduced by Boufounos and Baraniuk in [5], is the extreme version of quantized compressed sensing; it is particularly beneficial to consider 1-bit measurements in analog-to-digital conversion (see the webpage <http://dsp.rice.edu/1bitCS/>). Several numerical results are available, and there are a few recent theoretical results as well.

Suppose that $\mathbf{x} \in \mathbb{R}^n$ is s -sparse. Gupta *et al.* [13] demonstrate that the support of \mathbf{x} can tractably be recovered from either 1) $O(s \log n)$ nonadaptive measurements assuming a constant dynamic range of \mathbf{x} (i.e., the magnitude of all nonzero entries of \mathbf{x} is assumed to lie between two constants), or 2) $O(s \log n)$ adaptive measurements. Jacques *et al.* [14] demonstrate that any *consistent* estimate of \mathbf{x} will be accurate provided that $m \geq O(s \log n)$. Here, consistent means that the estimate $\hat{\mathbf{x}}$ should have unit norm, be at least as sparse as \mathbf{x} , and agree with the measurements, i.e., $\text{sign}(\langle \mathbf{a}_i, \hat{\mathbf{x}} \rangle) = \text{sign}(\langle \mathbf{a}_i, \mathbf{x} \rangle)$ for all i .

These results of Jacques *et al.* [14] can be extended to handle adversarial bit flips. The difficulty in applying these results is that the first two conditions are nonconvex, and thus, it is unknown whether there is a polynomial-time solver which is guaranteed to return a consistent solution. We note that there are heuristic algorithms, including one in [14] which often provide such a solution in simulations.

In a dual line of research, Gunturk *et al.* [11], [12] analyze sigma-delta quantization. The focus of their results is to achieve an excellent dependence of m on the accuracy δ while minimizing the number of bits per measurement. However, the measurements y_i in sigma-delta quantization are not related to any linear measurements [unlike those in (I.2) and (I.3)] but are allowed to be constructed in a judicious fashion (e.g., iteratively). Furthermore, in [11] and [12], the number of bits per measurement depends on the dynamic range of the nonzero part of \mathbf{x} . Similarly, the recent work of Ardestanizadeh *et al.* [1] requires a finite number of bits per measurement.

The noiseless 1-bit compressed sensing given by the model (I.2) was considered by the present authors in the earlier paper [23], where the following convex program was introduced:

$$\begin{aligned} \min \quad & \|\mathbf{x}'\|_1 \\ \text{subject to} \quad & y_i = \text{sign}(\langle \mathbf{a}_i, \mathbf{x}' \rangle), \quad i = 1, 2, \dots, m \\ \text{and} \quad & \sum_{i=1}^m y_i \langle \mathbf{a}_i, \mathbf{x}' \rangle = m. \end{aligned}$$

This program was shown in [23] to accurately recover an s -sparse vector \mathbf{x} from $m = O(s \log(n/s)^2)$ measurements y_i . This result was the first to propose a polynomial-time solver for 1-bit compressed sensing with provable accuracy guarantees. However, it was unclear how to modify the above convex program to account for possible noise.

This paper proposes to overcome this difficulty by considering the convex program (III.2) [and in the most general case, the optimization problem (I.7)]. One may note that the program (III.2) requires the knowledge of a bound on the (approximate) sparsity level s . In return, it does not need to be adjusted depending on the kind of noise or level of noise.

B. Sparse Logistic Regression

In order to give concrete results accessible to the statistics community, we now specialize Corollary 1.2 to the *logistic regression* model. Further, we drop the assumption that $\|\mathbf{x}\|_2 = 1$ in this section; this will allow easier comparison with the related literature (see below). The simple logistic function is defined as

$$f(z) = \frac{e^z}{e^z + 1}. \quad (\text{III.5})$$

In the logistic regression model, the observations $y_i \in \{-1, 1\}$ are iid random variables satisfying

$$\mathbb{P}\{y_i = 1\} = f(\langle \mathbf{a}_i, \mathbf{x} \rangle), \quad i = 1, 2, \dots, m. \quad (\text{III.6})$$

Note that this is a partial case of the generalized linear model (I.3) with $\theta(z) = \tanh(z/2)$. We thus have the following specialization of Corollary 1.2.

Corollary 3.3 (Sparse Logistic Regression): Let $\mathbf{a}_1, \dots, \mathbf{a}_m$ be independent standard Gaussian random vectors in \mathbb{R}^n , and fix \mathbf{x} satisfying $\mathbf{x}/\|\mathbf{x}\|_2 \in K_{n,s}$. Assume that the observations y_1, \dots, y_m follow the logistic regression model (III.6). Let $\delta > 0$ and suppose that

$$m \geq C\delta^{-2}s \log(2n/s).$$

Then, with probability at least $1 - 8 \exp(-c\delta^2 m)$, the solution $\hat{\mathbf{x}}$ to the convex program (III.2) satisfies

$$\left\| \hat{\mathbf{x}} - \frac{\mathbf{x}}{\|\mathbf{x}\|_2} \right\|_2 \leq \delta \max(\|\mathbf{x}\|_2^{-1}, 1). \quad (\text{III.7})$$

Proof: We begin by reducing to the case when $\|\mathbf{x}\|_2 = 1$ by rescaling the logistic function. Thus, let $\alpha = \|\mathbf{x}\|_2$ and define the scaled logistic function $f_\alpha(\mathbf{x}) = f(\alpha\mathbf{x})$. In particular

$$\mathbb{P}\{y_i = 1\} = f_\alpha(\langle \mathbf{a}_i, \frac{\mathbf{x}}{\|\mathbf{x}\|_2} \rangle).$$

To apply Corollary 3.1, it suffices to compute the correlation coefficient λ in (I.4). First, by rescaling f , we have also rescaled θ , so we consider $\theta(z) = \tanh(\alpha z/2)$. We can now compute λ using integration by parts:

$$\lambda = \mathbb{E}\theta(g)g = \mathbb{E}\theta'(g) = \frac{\alpha}{2} \mathbb{E} \text{sech}^2(\alpha g/2).$$

To further bound this quantity below, we can use the fact that $\text{sech}^2(x)$ is an even and decreasing function for $x \geq 0$. This yields

$$\begin{aligned} \lambda &\geq \frac{\alpha}{2} \mathbb{P}\{|\alpha g/2| \leq 1/2\} \cdot \text{sech}^2(1/2) \\ &\geq \frac{\text{sech}^2(1/2)}{2} \cdot \alpha \cdot \mathbb{P}\{|g| \leq 1/\alpha\} \geq \frac{1}{6} \min(\alpha, 1). \end{aligned}$$

The result follows from Corollary 3.1 since $\alpha = \|\mathbf{x}\|_2$. ■

Remark 3.4: Corollary 3.3 allows one to estimate the projection of \mathbf{x} onto the unit sphere. One may ask whether the norm of \mathbf{x} may be estimated as well. This depends on the assumptions made (see the literature described below). However, note that as $\|\mathbf{x}\|_2$ grows, the logistic regression model quickly approaches the noiseless 1-bit compressed sensing model, in which knowledge of $\|\mathbf{x}\|_2$ is lost in the measurements. Thus, since we do not assume that $\|\mathbf{x}\|_2$ is bounded, recovery of $\|\mathbf{x}\|_2$ becomes impossible.

For concreteness, we specialized to logistic regression. But as mentioned in Section I, the model (I.3) can be interpreted as the generalized linear model, so our results can be readily used for various problems in sparse binomial regression. Some of the recent work in sparse binomial regression includes the papers [2], [6], [15], [20], [21], [25], [27]. Let us point to the most directly comparable results.

In [2], [6], [15], and [21], the authors propose to estimate the coefficient vector (which in our notation is \mathbf{x}) by minimizing the negative log-likelihood plus an extra ℓ_1 regularization term. Bunea [6] considers the logistic regression model. She derives an accuracy bound for the estimate (in the ℓ_1 norm) under a certain *condition stabil* and a lower bound on the magnitude of the entries of \mathbf{x} . Similarly, Bach [2] and Kakade *et al.* [15]

derive accuracy bounds (again in the ℓ_1 norm) under *restrictive eigenvalue conditions*. The most directly comparable result is given by Negahban *et al.* [21]. There, the authors show that if the measurement vectors \mathbf{a}_i have independent sub-Gaussian entries, $\|\mathbf{x}\|_0 \leq s$, and $\|\mathbf{x}\|_2 \leq 1$, then with high probability one has $\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq \delta$, provided that the number of measurements is $m \geq C\delta^{-1}s \log n$. Their results apply to the generalized linear model (I.3) under some assumptions on θ .

One main novelty in this paper is that knowledge of the function θ , which defines the model family, is completely unnecessary when recovering the coefficient vector. Indeed, the optimization problems (I.7) and (III.2) do not need to know θ . This stands in contrast to programs based on maximum likelihood estimation. This may be of interest in nonparametric statistical applications in which it is unclear which binary model to pick—the logistic model may be chosen somewhat arbitrarily.

Another difference between our results and those above is in the conditions required. The above papers allow for more general design matrices than those in this paper, but this necessarily leads to strong assumptions on $\|\mathbf{x}\|_2$. As the inner products between $\langle \mathbf{a}_i, \mathbf{x} \rangle$ grow large, the logistic regression model approaches the 1-bit compressed sensing model. However, as shown in [23], accurate 1-bit compressed sensing is impossible for discrete measurement ensembles (not only is it impossible to recover \mathbf{x} , it is also impossible to recover $\mathbf{x}/\|\mathbf{x}\|_2$). Thus, the above results, all of which *do* allow for discrete measurement ensembles, necessitate rather strong conditions on the magnitude of $\langle \mathbf{a}_i, \mathbf{x} \rangle$, or equivalently, on $\|\mathbf{x}\|_2$; these are made explicitly in [6], [15], and [21] and implicitly in [2]. In contrast, our theoretical bounds on the relative error only improve as the average magnitude of $\langle \mathbf{a}_i, \mathbf{x} \rangle$ increases.

C. Low-Rank Matrix Recovery

We quickly mention that our model applies to single bit measurements of a low-rank matrix. Perhaps the closest practical application is quantum state tomography [10], but still, the requirement of Gaussian measurements is somewhat unrealistic. Thus, the purpose of this section is to give an intuition and a benchmark.

Let $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$ be a matrix of interest with rank r and Frobenius norm $\|\mathbf{X}\|_F = 1$. Consider that we have m single-bit measurements following the model in the introduction so that $n = n_1 \times n_2$. Similarly to sparse vectors, the set of low-rank matrices is not convex, but has a natural convex relaxation as follows. Let

$$K_{n_1, n_2, r} = \{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2} : \|\mathbf{X}\|_* \leq \sqrt{r}, \|\mathbf{X}\|_F \leq 1\}$$

where $\|\mathbf{X}\|_*$ denotes the nuclear norm, i.e., the sum of the singular values of \mathbf{X} .

In order to apply Theorems 1.1 and 1.3, we only need to calculate $w(K_{n_1, n_2, r})$, as follows:

$$w(K_{n_1, n_2, r}) = 2\mathbb{E} \sup_{\mathbf{X} \in K_{n_1, n_2, r}} \langle \mathbf{G}, \mathbf{X} \rangle$$

where \mathbf{G} is a matrix with standard normal entries and the inner product above is the standard entrywise inner product, i.e., $\langle \mathbf{G}, \mathbf{X} \rangle = \sum_{i,j} \mathbf{G}_{i,j} \mathbf{X}_{i,j}$. Since the nuclear norm and operator

norm are dual to each other, we have $\langle \mathbf{G}, \mathbf{X} \rangle \leq \|\mathbf{G}\| \cdot \|\mathbf{X}\|_*$. Further, for each $\mathbf{X} \in K_{n_1, n_2, r}$, $\|\mathbf{X}\|_* \leq \sqrt{r}$, and thus

$$w(K_{n_1, n_2, r}) \leq \sqrt{r} \mathbb{E} \|\mathbf{G}\|.$$

The expected norm of a Gaussian matrix is well studied; one has $\mathbb{E} \|\mathbf{G}\| \leq \sqrt{n_1} + \sqrt{n_2}$ (see, e.g., [28, Th. 5.32]). Thus, $w(K_{n_1, n_2, r}) \leq (\sqrt{n_1} + \sqrt{n_2})\sqrt{r}$. It follows that $O((n_1 + n_2)r)$ noiseless 1-bit measurements are sufficient to guarantee accurate recovery of rank- r matrices. We note that this matches the number of linear (infinite bit precision) measurements required in the low-rank matrix recovery literature (see [7]).

D. Extension to Measurements With Correlated Entries

A commonly used statistical model would take \mathbf{a}_i to be Gaussian vectors with correlated entries, namely $\mathbf{a}_i \sim \mathcal{N}(0, \Sigma)$, where Σ is a given covariance matrix. In this section, we present an extension of our results to allow such correlations. Let $\lambda_{\min} = \lambda_{\min}(\Sigma)$ and $\lambda_{\max} = \lambda_{\max}(\Sigma)$ denote the smallest and largest eigenvalues of Σ ; the condition number of Σ is then $\kappa(\Sigma) = \lambda_{\max}(\Sigma)/\lambda_{\min}(\Sigma)$. It will be convenient to choose the normalization $\|\Sigma^{1/2}\mathbf{x}\|_2 = 1$; as before, this may be done by absorbing a constant into the definition of θ .

We propose the following generalization of the convex program (III.2):

$$\begin{aligned} \max \quad & \sum_{i=1}^m y_i \langle \mathbf{a}_i, \mathbf{x}' \rangle \\ \text{subject to} \quad & \|\mathbf{x}'\|_1 \leq \sqrt{s/\lambda_{\min}} \text{ and } \|\Sigma^{1/2}\mathbf{x}'\|_2 \leq 1. \end{aligned} \quad (\text{III.8})$$

The following result extends Corollary 3.1 to general covariance Σ . For simplicity, we restrict ourselves to exactly sparse signals; however, the proof below allows for a more general signal set.

Corollary 3.5: Let $\mathbf{a}_1, \dots, \mathbf{a}_m$ be independent random vectors with distribution $\mathcal{N}(0, \Sigma)$. Fix \mathbf{x} satisfying $\|\mathbf{x}\|_0 \leq s$ and $\|\Sigma^{1/2}\mathbf{x}\|_2 = 1$. Assume that the measurements y_1, \dots, y_m follow the model from Section I-C. Let $\delta > 0$ and suppose that

$$m \geq C\kappa(\Sigma) \delta^{-2} s \log(2n/s).$$

Then, with probability at least $1 - 8 \exp(-c\delta^2 m)$, the solution $\hat{\mathbf{x}}$ to the convex program (III.8) satisfies

$$\lambda_{\min}(\Sigma) \cdot \|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq \|\Sigma^{1/2}\hat{\mathbf{x}} - \Sigma^{1/2}\mathbf{x}\|_2^2 \leq \delta/\lambda.$$

Remark 3.6: Theorem 1.3 can be generalized in the same way—the number of measurements required is scaled by $\kappa(\Sigma)$ and the error bound is scaled by $\lambda_{\min}(\Sigma)^{-1}$.

Proof of Corollary 3.5: The feasible set in (III.8) is $K := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_1 \leq \sqrt{s/\lambda_{\min}(\Sigma)}, \|\Sigma^{1/2}\mathbf{x}\|_2 \leq 1\}$. Note that the signal \mathbf{x} considered in the statement of the corollary is feasible, since $\|\mathbf{x}\|_1 \leq \|\mathbf{x}\|_2 \sqrt{\|\mathbf{x}\|_0} \leq \|\Sigma^{-1/2}\| \cdot \|\Sigma^{1/2}\mathbf{x}\|_2 \sqrt{s} \leq \sqrt{s/\lambda_{\min}(\Sigma)}$.

Define $\tilde{\mathbf{a}}_i := \Sigma^{-1/2}\mathbf{a}_i$; then $\tilde{\mathbf{a}}_i$ are independent standard normal vectors and $\langle \mathbf{a}_i, \mathbf{x} \rangle = \langle \Sigma^{1/2}\tilde{\mathbf{a}}_i, \mathbf{x} \rangle = \langle \tilde{\mathbf{a}}_i, \Sigma^{1/2}\mathbf{x} \rangle$. Thus, it follows from Corollary 1.2 applied with $\Sigma^{1/2}\mathbf{x}$ replacing \mathbf{x} that if

$$m \geq C\delta^{-2} w(\Sigma^{1/2}K)^2$$

then with probability at least $1 - 8 \exp(-c\delta^2 m)$

$$\|\Sigma^{1/2} \hat{\mathbf{x}} - \Sigma^{1/2} \mathbf{x}\|_2 \leq \delta/\lambda.$$

It remains to bound $w(\Sigma^{1/2} K)$. Since $\Sigma^{1/2}/\|\Sigma^{1/2}\|$ acts as a contraction, Slepian's inequality (see [19, Corollary 3.14]) gives $w(\Sigma^{1/2} K) \leq \|\Sigma^{1/2}\| \cdot w(K) = \lambda_{\max}(\Sigma)^{1/2} \cdot w(K)$. Further, $K \subseteq \lambda_{\min}(\Sigma)^{-1/2} K_{n,s}$. Thus, it follows from (III.3) that $w(\Sigma^{1/2} K)^2 \leq \kappa(\Sigma) w(K_{n,s})^2 \leq C\kappa(\Sigma) s \log(2n/s)$. This completes the proof. \blacksquare

IV. DEDUCING THEOREMS 1.1 AND 1.3 FROM CONCENTRATION INEQUALITIES

In this section, we show how to deduce our main results, Theorems 1.1 and 1.3, from concentration inequalities. These inequalities are stated in Propositions 4.2 and 4.3 below, whose proofs are deferred to Sections V and VI, respectively.

A. Proof of Theorem 1.1

Consider the rescaled objective function from the program (I.7):

$$f_{\mathbf{x}}(\mathbf{x}') = \frac{1}{m} \sum_{i=1}^m y_i \langle \mathbf{a}_i, \mathbf{x}' \rangle. \quad (\text{IV.1})$$

Here, the subscript x indicates that f is a random function whose distribution depends on x through y_i . Note that the solution $\hat{\mathbf{x}}$ to the program (I.7) satisfies $f_{\mathbf{x}}(\hat{\mathbf{x}}) \geq f_{\mathbf{x}}(\mathbf{x})$, since \mathbf{x} is feasible. We claim that for any $\mathbf{x}' \in K$ which is far away from \mathbf{x} , the value $f_{\mathbf{x}}(\mathbf{x}')$ is small with high probability. Thus, $\hat{\mathbf{x}}$ must be near to \mathbf{x} .

To begin to substantiate this claim, let us calculate $\mathbb{E}f_{\mathbf{x}}(\mathbf{x}')$ for a fixed vector $\mathbf{x}' \in K$.

Lemma 4.1 (Expectation): Fix $\mathbf{x}' \in \mathbb{R}^n$. Then

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

and thus

$$\mathbb{E}[f_{\mathbf{x}}(\mathbf{x}) - f_{\mathbf{x}}(\mathbf{x}')] = \lambda(1 - \langle \mathbf{x}, \mathbf{x}' \rangle) \geq \frac{\lambda}{2} \|\mathbf{x} - \mathbf{x}'\|_2^2.$$

Proof: We have

$$\mathbb{E}f_{\mathbf{x}}(\mathbf{x}') = \frac{1}{m} \sum_{i=1}^m \mathbb{E}y_i \langle \mathbf{a}_i, \mathbf{x}' \rangle = \mathbb{E}y_1 \langle \mathbf{a}_1, \mathbf{x}' \rangle.$$

Now we condition on \mathbf{a}_1 to give

$$\begin{aligned} \mathbb{E}y_1 \langle \mathbf{a}_1, \mathbf{x}' \rangle &= \mathbb{E}\mathbb{E}[y_1 \langle \mathbf{a}_1, \mathbf{x}' \rangle | \mathbf{a}_1] \\ &= \mathbb{E}\theta(\langle \mathbf{a}_1, \mathbf{x} \rangle) \langle \mathbf{a}_1, \mathbf{x}' \rangle. \end{aligned}$$

Note that $\langle \mathbf{a}_1, \mathbf{x} \rangle$ and $\langle \mathbf{a}_1, \mathbf{x}' \rangle$ are a pair of normal random variables with covariance $\langle \mathbf{x}, \mathbf{x}' \rangle$. Thus, by taking $g, h \in \mathcal{N}(0, 1)$ to be independent, we may rewrite the above expectation as

$$\begin{aligned} \theta(g) (\langle \mathbf{x}, \mathbf{x}' \rangle g + (\|\mathbf{x}'\|_2^2 - \langle \mathbf{x}, \mathbf{x}' \rangle)^{1/2} h) \\ = \langle \mathbf{x}, \mathbf{x}' \rangle \mathbb{E}\theta(g) g = \lambda \langle \mathbf{x}, \mathbf{x}' \rangle \end{aligned}$$

where the last equality follows from (I.4). Lemma 4.1 is proved. \blacksquare

Next we show that $f(\mathbf{x}')$ does not deviate far from its expectation *uniformly* for all $\mathbf{x}' \in K$.

Proposition 4.2 (Concentration): For each $t > 0$, we have

$$\left\{ \mathbb{P} \left\{ \sup_{\mathbf{z} \in K-K} |f_{\mathbf{x}}(\mathbf{z}) - \mathbb{E}f_{\mathbf{x}}(\mathbf{z})| \geq 4w(K)/\sqrt{m} + t \right\} \right\} \leq 4 \exp(-mt^2/8).$$

This result is proved in Section V using standard techniques of probability in Banach spaces.

Theorem 1.1 is a direct consequence of this proposition.

Proof of Theorem 1.1: Let $t > 0$. By Proposition 4.2, the following event occurs with probability at least $1 - 4 \exp(-mt^2/8)$:

$$\sup_{\mathbf{z} \in K} |f_{\mathbf{x}}(\mathbf{z}) - f_{\mathbf{x}}(\mathbf{z})| \leq 4w(K)/\sqrt{m} + t.$$

Suppose the above event indeed occurs. Let us apply this inequality for $\mathbf{z} = \hat{\mathbf{x}} - \mathbf{x} \in K - K$. By definition of $\hat{\mathbf{x}}$, we have $f_{\mathbf{x}}(\hat{\mathbf{x}}) \geq f_{\mathbf{x}}(\mathbf{x})$. Noting that the function $f_{\mathbf{x}}(\mathbf{z})$ is linear in \mathbf{z} , we obtain

$$\begin{aligned} 0 &\leq f_{\mathbf{x}}(\hat{\mathbf{x}}) - f_{\mathbf{x}}(\mathbf{x}) = f_{\mathbf{x}}(\hat{\mathbf{x}} - \mathbf{x}) \\ &\leq \mathbb{E}[f_{\mathbf{x}}(\hat{\mathbf{x}} - \mathbf{x})] + 4w(K)/\sqrt{m} + t \\ &\leq -\frac{\lambda}{2} \|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 + 4w(K)/\sqrt{m} + t. \end{aligned} \quad (\text{IV.2})$$

The last inequality follows from Lemma 4.1. Finally, we choose $t = 4\beta/\sqrt{m}$ and rearrange terms to complete the proof of Theorem 1.1. \blacksquare

B. Proof of Theorem 1.3

The argument is similar to that of Theorem 1.1 given above. We consider the rescaled objective functions with corrupted and uncorrupted measurements:

$$\begin{aligned} f_{\mathbf{x}}(\mathbf{x}') &= \frac{1}{m} \sum_{i=1}^m y_i \langle \mathbf{a}_i, \mathbf{x}' \rangle \\ \tilde{f}_{\mathbf{x}}(\mathbf{x}') &= \frac{1}{m} \sum_{i=1}^m \tilde{y}_i \langle \mathbf{a}_i, \mathbf{x}' \rangle \\ &= \frac{1}{m} \sum_{i=1}^m \text{sign}(\langle \mathbf{a}_i, \mathbf{x} \rangle) \langle \mathbf{a}_i, \mathbf{x}' \rangle. \end{aligned} \quad (\text{IV.3})$$

Arguing as in Lemma 4.1 (with $\theta(z) = \text{sign}(z)$), we have

$$\tilde{f}_{\mathbf{x}}(\mathbf{x}') = \lambda \langle \mathbf{x}, \mathbf{x}' \rangle = |g| \cdot \langle \mathbf{x}, \mathbf{x}' \rangle = \sqrt{2/\pi} \langle \mathbf{x}, \mathbf{x}' \rangle. \quad (\text{IV.4})$$

Similarly to the proof of Theorem 1.1, we now need to show that $f_{\mathbf{x}}(\mathbf{x}')$ does not deviate far from the expectation of $\tilde{f}_{\mathbf{x}}(\mathbf{x}')$; but this time, the result should hold uniformly over not only $\mathbf{x}' \in K - K$ but also $\mathbf{x} \in K$ and \mathbf{y} with small Hamming distance to $\tilde{\mathbf{y}}$. This is the content of the following proposition.

Proposition 4.3 (Uniform Concentration): Let $\delta > 0$ and suppose that

$$m \geq C\delta^{-6}w(K)^2.$$

Then, with probability at least $1 - 8\exp(-c\delta^2m)$, we have

$$\sup_{\mathbf{x}, \mathbf{z}, \mathbf{y}} |f_{\mathbf{x}}(\mathbf{z}) - \mathbb{E}f_{\mathbf{x}}(\mathbf{z})| \leq \delta\sqrt{\log(e/\delta)} + 4\tau\sqrt{\log(e/\tau)} \quad (\text{IV.5})$$

where the supremum is taken over $\mathbf{x} \in K \cap S^{n-1}$, $\mathbf{z} \in K - K$ and \mathbf{y} satisfying $d_H(\mathbf{y}, \tilde{\mathbf{y}}) \leq \tau m$.

This result is significantly deeper than Proposition 4.2. It is based on a recent geometric result from [22] on random tessellations of sets on the sphere. The proof is given in Section VI.

Theorem 1.3 now follows from the same steps as in the proof of Theorem 1.1 above. ■

Remark 4.4 (Random Noise): The adversarial bit flips allowed in Theorem 1.3 can be combined with random noise. We considered two models of random noise in Section III-A. One was *random bit flips* where one would take $\tilde{y}_i = \xi_i \text{sign}(\langle \mathbf{a}_i, \mathbf{x} \rangle)$; here ξ_i are iid Bernoulli random variables satisfying $\mathbb{P}\{\xi_i = 1\} = p$. The proof of Theorem 1.3 would remain unchanged under this model, aside from the calculation

$$\mathbb{E}f_{\mathbf{x}}(\mathbf{x}') = \sqrt{2/\pi}(2p - 1).$$

The end result is that the error bound in (I.10) would be divided by $2p - 1$.

Another model considered in Section III-A was *random noise before quantization*. Thus, we let

$$\tilde{y}_i = \text{sign}(\langle \mathbf{a}_i, \mathbf{x} \rangle + g_i) \quad (\text{IV.6})$$

where $g_i \sim \mathcal{N}(0, \sigma^2)$ are iid. Once again, a slight modification of the proof of Theorem 1.3 allows the incorporation of such noise. Note that the above model is equivalent to $\tilde{y}_i = \text{sign}(\langle \tilde{\mathbf{a}}_i, \tilde{\mathbf{x}} \rangle)$ where $\tilde{\mathbf{a}}_i = (\mathbf{a}_i, g_i)$ and $\tilde{\mathbf{x}} = (\mathbf{x}, \sigma)$ (we have concatenated an extra entry onto \mathbf{a}_i and \mathbf{x}). Thus, by slightly adjusting the set K , we are back in our original model.

V. CONCENTRATION: PROOF OF PROPOSITION 4.2

Here, we prove the concentration inequality given by Proposition 4.2.

A. Tools: Symmetrization and Gaussian Concentration

The argument is based on the standard techniques of probability in Banach spaces—symmetrization and the Gaussian concentration inequality. Let us recall both these tools.

Lemma 5.1 (Symmetrization): Let $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ be independent Rademacher random variables.⁸ Then

$$\begin{aligned} \mu &:= \mathbb{E} \sup_{\mathbf{z} \in K-K} |f_{\mathbf{x}}(\mathbf{z}) - \mathbb{E}f_{\mathbf{x}}(\mathbf{z})| \\ &\leq 2\mathbb{E} \sup_{\mathbf{z} \in K-K} \frac{1}{m} \left| \sum_{i=1}^m \varepsilon_i y_i \langle \mathbf{a}_i, \mathbf{z} \rangle \right|. \end{aligned} \quad (\text{V.1})$$

⁸This means that $\mathbb{P}\{\varepsilon_i = 1\} = \mathbb{P}\{\varepsilon_i = -1\} = 1/2$ for each i . The random variables ε_i are assumed to be independent of each other and of any other random variables in question, namely \mathbf{a}_i and y_i .

Furthermore, we have the deviation inequality

$$\begin{aligned} &\mathbb{P} \left\{ \sup_{\mathbf{z} \in K-K} |f(\mathbf{z}) - \mathbb{E}f(\mathbf{z})| \geq 2\mu + t \right\} \\ &\leq 4 \left\{ \sup_{\mathbf{z} \in K-K} \left| \sum_{i=1}^m \varepsilon_i y_i \langle \mathbf{a}_i, \mathbf{z} \rangle \right| > t/2 \right\}. \end{aligned} \quad (\text{V.2})$$

Inequality (V.1) follows, e.g., from the proof of [19, Lemma 6.3]. The proof of inequality (V.2) is contained in [19, Chapter 6.1]. ■

Theorem 5.2 (Gaussian Concentration Inequality): Let $(G_{\mathbf{x}})_{\mathbf{x} \in T}$ be a centered Gaussian process indexed by a finite set T . Then, for every $r > 0$, one has

$$\mathbb{P} \left\{ \sup_{\mathbf{x} \in T} G_{\mathbf{x}} \geq \sup_{\mathbf{x} \in T} G_{\mathbf{x}} + r \right\} \leq \exp(-r^2/2\sigma^2)$$

where $\sigma^2 = \sup_{\mathbf{x} \in T} \mathbb{E}G_{\mathbf{x}}^2 < \infty$.

A proof of this result can be found e.g., in [18, Th. 7.1]. ■

This theorem can be extended to separable sets T in metric spaces by an approximation argument. In particular, given a set $K \subseteq B_2^n$ and $r > 0$, the standard Gaussian random vector \mathbf{g} in \mathbb{R}^n satisfies

$$\mathbb{P} \left\{ \sup_{\mathbf{x} \in K-K} \langle \mathbf{g}, \mathbf{x} \rangle \geq w(K) + r \right\} \leq \exp(-r^2/2). \quad (\text{V.3})$$

B. Proof of Proposition 4.2

We apply the first part (V.1) of Symmetrization Lemma 5.1. Note that since \mathbf{a}_i have symmetric distributions and $y_i \in \{-1, 1\}$, the random vectors $\varepsilon_i y_i \mathbf{a}_i$ has the same (iid) distribution as \mathbf{a}_i . Using the rotational invariance and the symmetry of the Gaussian distribution, we can represent the right-hand side of (V.1) as

$$\begin{aligned} &\sup_{\mathbf{z} \in K-K} \frac{1}{m} \left| \sum_{i=1}^m \varepsilon_i y_i \langle \mathbf{a}_i, \mathbf{z} \rangle \right| \stackrel{\text{dist}}{=} \sup_{\mathbf{z} \in K-K} \frac{1}{m} \left| \sum_{i=1}^m \langle \mathbf{a}_i, \mathbf{z} \rangle \right| \\ &\stackrel{\text{dist}}{=} \frac{1}{\sqrt{m}} \sup_{\mathbf{z} \in K-K} |\langle \mathbf{g}, \mathbf{z} \rangle| = \frac{1}{\sqrt{m}} \sup_{\mathbf{z} \in K-K} \langle \mathbf{g}, \mathbf{z} \rangle \end{aligned} \quad (\text{V.4})$$

where $\stackrel{\text{dist}}{=}$ signifies the equality in distribution. So taking the expectation in (V.1), we obtain

$$\begin{aligned} \mathbb{E} \sup_{\mathbf{z} \in K-K} |f_{\mathbf{x}}(\mathbf{z}) - \mathbb{E}f_{\mathbf{x}}(\mathbf{z})| &\leq \frac{2}{\sqrt{m}} \mathbb{E} \sup_{\mathbf{z} \in K-K} \langle \mathbf{g}, \mathbf{z} \rangle \\ &= \frac{2w(K)}{\sqrt{m}}. \end{aligned} \quad (\text{V.5})$$

To supplement this expectation bound with a deviation inequality, we use the second part (V.2) of Symmetrization Lemma 2.3 along with (V.5) and (V.4). This yields

$$\begin{aligned} &\mathbb{P} \left\{ \sup_{\mathbf{z} \in K-K} |f_{\mathbf{x}}(\mathbf{z}) - \mathbb{E}f_{\mathbf{x}}(\mathbf{z})| \geq \frac{4w(K)}{\sqrt{m}} + t \right\} \\ &\leq 4\mathbb{P} \left\{ \frac{1}{\sqrt{m}} \sup_{\mathbf{z} \in K-K} \langle \mathbf{g}, \mathbf{z} \rangle > t/2 \right\}. \end{aligned}$$

Now it remains to use the Gaussian concentration inequality (V.3) with $r = t\sqrt{m}/2$. The proof of Proposition 4.2 is complete. ■

VI. CONCENTRATION: PROOF OF PROPOSITION 4.3

Here, we prove the uniform concentration inequality given by Proposition 4.3. Besides standard tools in geometric functional analysis such as Sudakov minoration for covering numbers, our argument is based on the recent work [22] on random hyperplane tessellations. Let us first recall the relevant tools.

A. Tools: Covering Numbers, Almost Isometries, and Random Tessellations

Consider a set $T \subset \mathbb{R}^n$ and a number $\varepsilon > 0$. Recall that an ε -net of T (in the Euclidean norm) is a set $N_\varepsilon \subset T$ which has the following property: for every $\mathbf{x} \in T$, there exists $\tilde{\mathbf{x}} \in N_\varepsilon$ satisfying $\|\mathbf{x} - \tilde{\mathbf{x}}\|_2 \leq \varepsilon$. The covering number of T to precision ε , which we call $N(T, \varepsilon)$, is the minimal cardinality of an ε -net of T . The covering numbers are closely related to the mean width, as shown by the following well-known inequality:

Theorem 6.1 (Sudakov Minoration): Given a set $T \subset \mathbb{R}^n$ and a number $\varepsilon > 0$, one has

$$\log N(K, \varepsilon) \leq \log N(K - K, \varepsilon) \leq C\varepsilon^{-2}w(K)^2.$$

A proof of this theorem can be found e.g., in [19, Th. 3.18]. ■

We will also need two results from the recent work [22]. To state them conveniently, let \mathbf{A} denote the $m \times n$ matrix with rows \mathbf{a}_i . Thus, \mathbf{A} is standard Gaussian matrix with iid standard normal entries. The first (simple) result guarantees that \mathbf{A} acts as a almost isometric embedding from $(K, \|\cdot\|_2)$ into $(\mathbb{R}^m, \|\cdot\|_1)$.

Lemma 6.2 (See [22, Lemma 2.1]): Consider a subset $K \subset B_2^n$. Then, for every $u > 0$, one has

$$\mathbb{P} \left\{ \sup_{\mathbf{x} \in K-K} \left| \frac{1}{m} \|\mathbf{A}\mathbf{x}\|_1 - \sqrt{\frac{2}{\pi}} \|\mathbf{x}\|_2 \right| \geq \frac{4w(K)}{\sqrt{m}} + u \right\} \leq 2 \exp \left(-\frac{mu^2}{2} \right).$$

The second (not simple) result demonstrates that the discrete map $\mathbf{x} \mapsto \text{sign}(\mathbf{A}\mathbf{x})$ acts as an almost isometric embedding from $(K \cap S^{n-1}, d_G)$ into the Hamming cube $(\{-1, 1\}^m, d_H)$. Here, d_G and d_H denote the geodesic and Hamming metrics, respectively; the sign function is applied to all entries of $\mathbf{A}\mathbf{x}$; thus, $\text{sign}(\mathbf{A}\mathbf{x})$ is the vector with entries $\text{sign}(\langle \mathbf{a}_i, \mathbf{x} \rangle)$, $i = 1, \dots, m$.

Theorem 6.3 (See [22, Th. 1.5]): Consider a subset $K \subset B_2^n$ and let $\delta > 0$. Let

$$m \geq C\delta^{-6}w(K)^2.$$

Then, with probability at least $1 - 2 \exp(-c\delta^2 m)$, the following holds for all $\mathbf{x}, \mathbf{x}' \in K \cap S^{n-1}$:

$$\left| \frac{1}{\pi} d_G(\mathbf{x}, \mathbf{x}') - \frac{1}{m} d_H(\text{sign}(\mathbf{A}\mathbf{x}), \text{sign}(\mathbf{A}\mathbf{x}')) \right| \leq \delta.$$

B. Proof of Proposition 4.3

Let us first assume that $\tau = 0$ for simplicity; the general case will be discussed at the end of the proof. With this assumption, (IV.3) becomes

$$f_{\mathbf{x}}(z) = \tilde{f}_{\mathbf{x}}(z) = \frac{1}{m} \sum_{i=1}^m \text{sign}(\langle \mathbf{a}_i, \mathbf{x} \rangle) \langle \mathbf{a}_i, z \rangle. \quad (\text{VI.1})$$

To be able to approximate \mathbf{x} by a net, we use Sudakov minoration (Theorem 6.1) and find a δ -net N_δ of $K \cap S^{n-1}$ whose cardinality satisfies

$$\log |N_\delta| \leq C\delta^{-2}w(K)^2. \quad (\text{VI.2})$$

Lemma 6.4: Let $\delta > 0$ and assume that $m \geq C\delta^{-6}w(K)^2$. Then, we have the following.

- 1) (Bound on the net.) With probability at least $1 - 4 \exp(-cm\delta^2)$ we have

$$\sup_{\mathbf{x}_0, z} |f_{\mathbf{x}_0}(z) - \mathbb{E} f_{\mathbf{x}_0}(z)| \leq \delta \quad (\text{VI.3})$$

where the supremum is taken over all $\mathbf{x}_0 \in N_\delta$ and all $z \in K - K$.

- 2) (Deviation of sign patterns.) For $\mathbf{x}, \mathbf{x}_0 \in K \cap S^{n-1}$, consider the set

$$T(\mathbf{x}, \mathbf{x}_0) := \{i \in [m] : \text{sign}(\langle \mathbf{a}_i, \mathbf{x} \rangle) \neq \text{sign}(\langle \mathbf{a}_i, \mathbf{x}_0 \rangle)\}.$$

Then, with probability at least $1 - 2 \exp(-cm\delta^2)$, we have

$$\sup_{\mathbf{x}, \mathbf{x}_0} |T(\mathbf{x}, \mathbf{x}_0)| \leq 2m\delta \quad (\text{VI.4})$$

where the supremum is taken over all $\mathbf{x}, \mathbf{x}_0 \in K \cap S^{n-1}$ satisfying $\|\mathbf{x} - \mathbf{x}_0\|_2 \leq \delta$.

- 3) (Deviation of sums.) Let s be a natural number. With probability at least $1 - 2 \exp(-s \log(em/s)/2)$, we have

$$\begin{aligned} & \sup_{z, T} \sum_{i \in T} |\langle \mathbf{a}_i, z \rangle| \\ & \leq s \left(\sqrt{\frac{8}{\pi}} + \frac{4w(K)}{\sqrt{s}} + 2\sqrt{\log(em/s)} \right) \end{aligned} \quad (\text{VI.5})$$

where the supremum is taken over all $z \in K - K$ and all subsets $T \subset [n]$ with cardinality $|T| \leq s$.

Proof of Proposition 4.3: Let us apply Lemma 6.4, where in part 3 we take $s = 2m\delta$ rounded down to the next smallest integer. Then, all of the events (VI.3), (VI.4), (VI.5) hold simultaneously with probability at least $1 - 8 \exp(-cm\delta^2)$.

Recall that our goal is to bound the deviation of $f_{\mathbf{x}}(z)$ from its expectation uniformly over $\mathbf{x} \in K \cap S^{n-1}$ and $z \in K - K$. To this end, given $\mathbf{x} \in K \cap S^{n-1}$ we choose $\mathbf{x}_0 \in N_\delta$ so that $\|\mathbf{x} - \mathbf{x}_0\|_2 \leq \delta$. By (VI.1) and definition of the set $T(\mathbf{x}, \mathbf{x}_0)$ in Lemma 6.4, we can approximate $f_{\mathbf{x}}(z)$ by $f_{\mathbf{x}_0}(z)$ as follows:

$$|f_{\mathbf{x}}(z) - f_{\mathbf{x}_0}(z)| \leq \frac{2}{m} \sum_{i \in T(\mathbf{x}, \mathbf{x}_0)} |\langle \mathbf{a}_i, z \rangle|. \quad (\text{VI.6})$$

Furthermore, (VI.4) guarantees that $|T(\mathbf{x}, \mathbf{x}_0)| \leq 2m\delta$. It then follows from (VI.5), our choice $s = 2m\delta$ and the assumption on m that

$$\sum_{i \in T(\mathbf{x}, \mathbf{x}_0)} |\langle \mathbf{a}_i, \mathbf{z} \rangle| \leq Cm\delta \sqrt{\log(e/\delta)}.$$

Thus

$$|f_{\mathbf{x}}(\mathbf{z}) - f_{\mathbf{x}_0}(\mathbf{z})| \leq 2C\delta \sqrt{\log(e/\delta)}.$$

Combining this with (VI.3), we obtain

$$|f_{\mathbf{x}}(\mathbf{z}) - \mathbb{E}f_{\mathbf{x}_0}(\mathbf{z})| \leq C_1\delta \sqrt{\log(e/\delta)}. \quad (\text{VI.7})$$

Further, recall from (VI.4) that $\mathbb{E}f_{\mathbf{x}_0}(\mathbf{z}) = \sqrt{2/\pi} \langle \mathbf{x}, \mathbf{z} \rangle$ and thus

$$\begin{aligned} |\mathbb{E}f_{\mathbf{x}_0}(\mathbf{z}) - \mathbb{E}f_{\mathbf{x}}(\mathbf{z})| &= \sqrt{2/\pi} |\langle \mathbf{x}_0 - \mathbf{x}, \mathbf{z} \rangle| \\ &\leq 2\sqrt{2/\pi} \|\mathbf{x}_0 - \mathbf{x}\|_2 \leq 2\sqrt{2/\pi} \delta. \end{aligned} \quad (\text{VI.8})$$

The last two inequalities in this line follow since $\mathbf{z} \in K - K \subset 2B_2^n$ and $\|\mathbf{x}_0 - \mathbf{x}\|_2 \leq \delta$. Finally, we combine inequalities (VI.7) and (VI.8) to give

$$|f_{\mathbf{x}}(\mathbf{z}) - \mathbb{E}f_{\mathbf{x}}(\mathbf{z})| \leq C_2\delta \sqrt{\log(e/\delta)}.$$

Note that we can absorb the constant C_2 into the requirement $m \geq C\delta^{-6}w(K)^2$. This completes the proof in the case where $\tau = 0$.

In the general case, we only need to tweak the above argument by increasing the size of s considered in (VI.5). Specifically, it is enough to choose s to be $\tau m + 2m\delta$ rounded down to the next smallest integer. This allows one to account for arbitrary τm bit flips of the numbers $\text{sign}(\langle \mathbf{a}_i, \mathbf{x} \rangle)$, which produce the difference between $f_{\mathbf{x}}(\mathbf{z})$ and $f_{\mathbf{x}_0}(\mathbf{z})$. The proof of Proposition 4.3 is complete. ■

It remains to prove Lemma 6.4 that was used in the argument above.

Proof of Lemma 6.4: To prove part 1, we can use Proposition 4.2 combined with the union bound over the net N_δ . Using the bound (VI.2) on the cardinality of N_δ , we obtain

$$\begin{aligned} \mathbb{P} \left\{ \sup_{\mathbf{x}_0, \mathbf{z}} |f_{\mathbf{x}_0}(\mathbf{z}) - \mathbb{E}f_{\mathbf{x}_0}(\mathbf{z})| \geq 4w(K)/\sqrt{m} + t \right\} \\ \leq |N_\delta| 4 \exp(-mt^2/8) \\ \leq 4 \exp(-mt^2/8 + C\delta^{-2}w(K)^2) \end{aligned}$$

where the supremum is taken over all $\mathbf{x}_0 \in N_\delta$ and $\mathbf{z} \in K - K$. It remains to choose $t = \delta/2$ and recall that $m \geq C\delta^{-6}w(K)^2$ to finish the proof.

We now turn to part 2. First, note that $|T(\mathbf{x}, \mathbf{x}_0)| = d_H(\text{sign}(\mathbf{A}\mathbf{x}), \text{sign}(\mathbf{A}\mathbf{x}_0))$. Theorem 6.3 demonstrates that this Hamming distance is almost isometric to the geodesic distance, which itself satisfies $\frac{1}{\pi} d_G(\mathbf{x}, \mathbf{x}_0) \leq \|\mathbf{x} - \mathbf{x}_0\|_2 \leq \delta$. Specifically, Theorem 6.3 yields that under our assumption that $m \geq C\delta^{-6}w(K)^2$, with probability at least $1 - 2 \exp(-c\delta^2 m)$, one has

$$|T(\mathbf{x}, \mathbf{x}_0)| \leq 2m\delta \quad (\text{VI.9})$$

for all $\mathbf{x}, \mathbf{x}_0 \in K \cap S^{n-1}$ satisfying $\|\mathbf{x} - \mathbf{x}_0\|_2 \leq \delta$. This proves part 2.

In order to prove part 3, we may consider the subsets T satisfying $|T| = s$; there are $\binom{m}{s} \leq \exp(s \log(em/s))$ of them. Now we apply Lemma 6.2 for the $T \times n$ matrix $\mathbf{P}_T \mathbf{A}$ where \mathbf{P}_T denotes the coordinate restriction in \mathbb{R}^m onto \mathbb{R}^T ; so in the statement of Lemma 6.2, we replace m by $|T| = s$. Combined with the union bound over all T , this gives

$$\begin{aligned} \mathbb{P} \left\{ \sup_{\mathbf{z}, T} \frac{1}{s} \sum_{i \in T} |\langle \mathbf{a}_i, \mathbf{z} \rangle| \geq \sqrt{\frac{2}{\pi}} \|\mathbf{z}\|_2 + \frac{4w(K)}{\sqrt{s}} + u \right\} \\ \leq 2 \exp \left(s \log(em/s) - \frac{su^2}{2} \right). \end{aligned}$$

Recall that $\|\mathbf{z}\|_2 \leq 2$ since $\mathbf{z} \in K - K \subset 2B_2^n$. Finally, we take $u^2 = 4 \log(em/s)$ to complete the proof. ■

VII. DISCUSSION

Unlike traditional compressed sensing, which has already enjoyed an extraordinary wave of theoretical results, 1-bit compressed sensing is in its early stages. In this paper, we proposed a polynomial-time solver (given by a convex program) for noisy 1-bit compressed sensing, and we gave theoretical guarantees on its performance. The discontinuity inherent in 1-bit measurements led to some unique mathematical challenges. We also demonstrated the connection to sparse binomial regression and derived novel results for this problem as well.

The problem setup in 1-bit compressed sensing (as first defined in [5]) is quite elegant, allowing for a theoretical approach. On the other hand, there are many compressed sensing results assuming substantially finer quantization. It would be of interest to build a bridge between the two regimes; for example, 2-bit compressed sensing would already open up new questions.

REFERENCES

- [1] E. Ardestanizadeh, M. Cheraghchi, and A. Shokrollahi, "Bit precision analysis for compressed sensing," in *Proc. IEEE Int. Symp. Inf. Theory*, 2009, pp. 1–5.
- [2] F. Bach, "Self-concordant analysis for logistic regression," *Electron. J. Statist.*, vol. 4, pp. 384–414, 2010.
- [3] P. Bartlett and S. Mendelson, "Rademacher and Gaussian complexities: Risk bounds and structural results," *J. Mach. Learn. Res.*, vol. 3, pp. 463–482, 2003.
- [4] P. Boufounos, "Reconstruction of sparse signals from distorted randomized measurements," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2010, pp. 3998–4001.
- [5] P. T. Boufounos and R. G. Baraniuk, "1-Bit compressive sensing," presented at the 42nd Annu. Conf. Inf. Sci. Syst., 2008.
- [6] F. Bunea, "Honest variable selection in linear and logistic regression models via ℓ_1 and $\ell_1 + \ell_2$ penalization," *Electron. J. Statist.*, vol. 2, pp. 1153–1194, 2008.
- [7] E. Candès and Y. Plan, "Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2342–2359, Apr. 2011.
- [8] C. Eldar and G. Kutyniok, Eds., *Compressed Sensing*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
- [9] A. A. Giannopoulos and V. D. Milman, "Asymptotic Convex Geometry: Short Overview," in *Different faces of geometry*, I. M. Series, Ed. Norwell, MA: Kluwer/Plenum, 2004, pp. 87–162.
- [10] D. Gross, Y. Liu, S. Flammia, S. Becker, and J. Eisert, "Quantum state tomography via compressed sensing," *Phys. Rev. Lett.*, vol. 105, no. 15, pp. 150401-1–150401-4, 2010.

- [11] C. Gunturk, M. Lammers, A. Powell, R. Saab, and O. Yilmaz, "Sigma delta quantization for compressed sensing," presented at the IEEE 44th Annu. Conf. Inf. Sci. Syst., 2010.
- [12] C. Gunturk, M. Lammers, A. Powell, R. Saab, and O. Yilmaz, Sobolev duals for random frames and sigma-delta quantization of compressed sensing measurements [Online]. Available: <http://arxiv.org/abs/1002.0182>
- [13] A. Gupta, R. Nowak, and B. Recht, "Sample complexity for 1-bit compressed sensing and sparse classification," presented at the IEEE Int. Symp. Inf. Theory, 2010.
- [14] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors [Online]. Available: <http://arxiv.org/abs/1104.3160>
- [15] O. Shamir, K. Sridharan, A. Tewari, and S. Kakade, Learning exponential families in high-dimensions: Strong convexity and sparsity [Online]. Available: <http://arxiv.org/abs/0911.0054>
- [16] J. Laska, "Regime change: Sampling rate vs. bit-depth in compressive sensing," Ph.D. dissertation, Rice Univ., Houston, TX, 2011.
- [17] J. Laska and R. Baraniuk, Regime change: Bit-depth versus measurement-rate in compressive sensing [Online]. Available: <http://arxiv.org/abs/1110.3450>
- [18] M. Ledoux, *The Concentration of Measure Phenomenon*. Providence, RI: Amer. Math. Soc., 2001.
- [19] M. Ledoux and M. Talagrand, *Probability in Banach Spaces: Isoperimetry and Processes*. Berlin, Germany: Springer-Verlag, 1991.
- [20] L. Meier, V. Van der Geer, and P. Bühlmann, "The group lasso for logistic regression," *J. R. Statist. Soc.: Series B (Statist. Methodol.)*, vol. 70, no. 1, pp. 53–71, 2008.
- [21] S. Negahban, P. Ravikumar, M. Wainwright, and B. Yu, A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers [Online]. Available: <http://arxiv.org/abs/1010.2731>
- [22] Y. Plan and R. Vershynin, Dimension reduction by random hyperplane tessellations [Online]. Available: <http://arxiv.org/abs/1111.4452>
- [23] Y. Plan and R. Vershynin, One-bit compressed sensing by linear programming [Online]. Available: <http://arxiv.org/abs/1109.4299>
- [24] G. Raskutti, M. Wainwright, and B. Yu, "Minimax rates of convergence for high-dimensional regression under ℓ_1 -ball sparsity," in *Proc. 47th Annu. Allert. Conf. IEEE Commun., Control, Comput.*, 2009, pp. 251–257.
- [25] P. Ravikumar, M. Wainwright, and J. Lafferty, "High-dimensional Ising model selection using ℓ_1 -regularized logistic regression," *Ann. Statist.*, vol. 38, no. 3, pp. 1287–1319, 2010.
- [26] M. Talagrand, "Majorizing measures: The generic chaining," *Ann. Probabil.*, vol. 24, no. 3, pp. 1049–1103, 1996.
- [27] S. Van der Geer, "High-dimensional generalized linear models and the lasso," *Ann. Statist.*, vol. 36, no. 2, pp. 614–645, 2008.
- [28] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," in *Compressed Sensing: Theory and Applications*, Y. Eldar and G. Kutyniok, Eds. Cambridge, U.K.: Cambridge Univ. Press.

Yaniv Plan received his B.A. from the University of California, Berkeley in 2004 where he double majored in Applied Mathematics and Physics. He earned his Ph.D. in Applied and Computational Mathematics at Caltech in 2011. He was also a visiting researcher at Stanford in 2010–2011. He is currently an Assistant Professor of Mathematics at the University of Michigan. His main areas of interest are compressive sensing, sparse reconstruction, low-rank matrix recovery, non-asymptotic random matrix theory, and probability in Banach spaces.

Roman Vershynin received his M.S. in Mathematics at Kharkiv National University, Ukraine in 1996. He earned his Ph.D. degree in Mathematics at University of Missouri, Columbia in 2000. After holding postdoctoral positions at Weizmann Institute of Science, Israel and University of Alberta, Canada, he was a faculty at the University of California, Davis, U.S.A. during 2003–2008. He is now a Professor of Mathematics at the University of Michigan, U.S.A. His core interests are in geometric functional analysis and probability theory, and most recently in non-asymptotic random matrix theory. He has worked on various applications of these areas in theoretical computer science, numerical analysis, statistics and information theory.