Sharp Guarantees for Solving Random Equations with One-Bit Information

Hossein Taheri, Ramtin Pedarsani, and Christos Thrampoulidis

Abstract—We study the performance of a wide class of convex optimization-based estimators for recovering a signal from corrupted one-bit measurements in high-dimensions. Our general result predicts sharply the performance of such estimators in the linear asymptotic regime when the measurement vectors have entries IID Gaussian. This includes, as a special case, the previously studied least-squares estimator and various novel results for other popular estimators such as least-absolute deviations, hinge-loss and logistic-loss. Importantly, we exploit the fact that our analysis holds for generic convex loss functions to prove a bound on the best achievable performance across the entire class of estimators. Numerical simulations corroborate our theoretical findings and suggest they are accurate even for relatively small problem dimensions.

I. INTRODUCTION

A. Motivation

Classical statistical signal-processing theory studies estimation problems in which the number of unknown parameters $n$ is small compared to the number of observations $m$. In contrast, modern inference problems are typically high-dimensional, that is $n$ can be of the same order as $m$. Examples are abundant in a wide range of signal-processing applications such as medical imaging, wireless communications, recommendation systems and so on. Classical tools and theories are not applicable in these modern inference problems. As such, over the last two decades or so, the study of high-dimensional estimation problems has received significant attention. Despite the remarkable progress in many directions, several important questions remain to be explored.

This paper studies the fundamental problem of recovering an unknown signal from (possibly corrupted) one-bit measurements in high-dimensions. We focus on a rather rich class of convex optimization-based estimators that includes, for example, least-squares (LS), least-absolute deviations (LAD), logistic regression and hinge-loss as special cases.

For such estimators and Gaussian measurement vectors, we compute their asymptotic performance in the high-dimensional linear regime in which $m/n \rightarrow \delta \in (1, +\infty)$. Importantly, our results are sharp. In contrast to existing related results which are order-wise (i.e., they involve unknown or loose constants) this allows us to accurately compare the relative performance of different methods (e.g., LS vs LAD). It is worth mentioning that while our predictions are asymptotic, our numerical illustrations suggest that they are valid for dimensions $m$ and $n$ that are as small as a few hundreds.

B. Contributions

Our goal is to recover $x_0 \in \mathbb{R}^n$ from measurements $y_i = \text{sign}(a_i^T x_0)$, $i = 1, \ldots, m$, where $a_i \in \mathbb{R}^n$ have entries iid Gaussian. The results account for possible corruptions by allowing each measurement $y_i$ to be sign-flipped with constant probability $\varepsilon \in [0, 1/2]$ (see Section II for details). We study the asymptotic performance of estimators $\hat{x}_\ell$ that are solutions to the following optimization problem for some convex loss function $\ell(\cdot)$.

$$\hat{x}_\ell := \arg\min_{x} \sum_{i=1}^{m} \ell(y_i a_i^T x).$$

When $m, n \to +\infty$ and $m/n \to \delta > \delta^\ast$, we show that the correlation of $\hat{x}_\ell$ to the true vector $x_0$ is sharply predicted by

$$\frac{1}{\sqrt{1+(\alpha/\mu)^2}} \text{ where the parameters } \alpha \text{ and } \mu \text{ are the solutions to a system of three non-linear equations in three unknowns.}$$

We find that the system of equations (and thus, the value of $\alpha/\mu$) depends on the loss function $\ell(\cdot)$ through its Moreau envelope function. We prove that $\delta^\ast > 1$ is necessary for the equations to have a bounded solution, but, in general, the value of the threshold $\delta^\ast$ depends both on the noise level $\varepsilon$ and on the loss function. For the general case where $\varepsilon \in [0, 1/2]$, we propose a method to find the upper bound on correlation which corresponds to the maximum value of correlation that can be achieved for any convex loss function.

Despite the fact that it is not the main focus of this paper, we remark that our results hold for the general case where measurements $y_i$ are determined according to an arbitrary function $f : \mathbb{R} \to [0, 1]$ (see (6)). Moreover as our analysis in Appendix B shows, the system of equations introduced in Section III-B can be extended to address regularized loss functions i.e.,

$$\sum_{i=1}^{m} \ell(y_i a_i^T x) + r \|x\|_2^2.$$

We specialize our general result to specific loss functions such as LS, LAD and hinge-loss. This allows us to numerically compare the performance of these popular estimators by simply evaluating the corresponding theoretical predictions. Our numerical illustrations corroborate our theoretical predictions. For LS, our equations can be solved in closed form.
form and recover the result of [29] (see Section I-C). For
the hinge-loss, we show that $\delta_\varepsilon^\ast$ is a decreasing function of
$\varepsilon$ that approaches $+\infty$ in the noiseless case and 2 when $\varepsilon = 1/2$. We believe that our work opens the possibility for
addressing several important open questions, such as finding
the optimal choice of the loss function in (1) and the value of $\delta_\varepsilon^\ast$ for general loss functions.

\section{Prior work}

As mentioned, over the past two decades there has been a
very long list of works that derive statistical guarantees for
high-dimensional estimation problems. In particular, many
of these are concerned with convex optimization-based
inference methods. Our work is most closely related to the
following two lines of research.

(a) Sharp asymptotic predictions for noisy linear measurements. Most of the results in the literature of high-dimensional statistics are order-wise in nature. Sharp asymptotic predictions have only recently appeared in the literature for the case of noisy linear measurements with Gaussian measurement vectors. There are by now three different approaches that have been used (to different extent each) towards asymptotic analysis of convex regularized estimators: (a) the one that is based on the approximate message passing (AMP) algorithm and its state-evolution analysis; [11], [12], [2], [3], [10] (b) the one that is based on Gaussian process (GP) inequalities, specifically the convex Gaussian min-max Theorem (CGMT); [26], [8], [27], [20], [31], [30] (c) and the “leave-one-out” approach [19], [13]. The three approaches are quite different to each other and each comes with its unique distinguishing features and disadvantages. A detailed comparison is beyond our scope. In this paper, we follow the GP approach and build upon the CGMT. Since concerned with linear measurements, these previous works consider estimators that solve minimization problems of the form

$$\hat{x} := \arg \min_{x} \sum_{i=1}^{m} \ell(y_i - a_i^T x) + rR(x)$$

Specifically, the loss function $\ell(\cdot)$ penalizes the residual. In this paper, we extend the applicability of the CGMT to optimization problems in the form of (1). For our case of signed measurements, (1) is more general than (2). To see this, note that for $y_i \in \pm 1$ and popular symmetric loss functions $\ell(t) = \ell(-t)$ (e.g., LS, LAD), (1) results in (2) by choosing $\ell(t) = \ell(t-1)$ in the former. Moreover, (1) includes several other popular loss functions such as the logistic loss and the hinge-loss which cannot be expressed by (2).

(b) One-bit compressed sensing. Our works naturally relates to the literature of one-bit compressed sensing (CS) [5]. The vast majority of performance guarantees for one-bit CS are order-wise in nature, e.g., [18], [22], [21], [23]. To the best of our knowledge, the only existing sharp results are presented in [29] for Gaussian measurement vectors. Specifically, the paper [29] derives the asymptotic performance of regularized LS for generalized nonlinear measurements, which include signed measurements as a special case. Our work can be seen as a direct extension of [29] to loss functions beyond least-squares, such as the hinge-loss. In fact, the result of [29] for our setting is a direct corollary of our main theorem (see Section IV-A). As in [29], our proof technique is based on the CGMT.

There are few works that consider general convex loss functions for estimating a signal from noisy measurements in high dimensions. In [33], the general estimator $\ell((a_i, x), y_i)$ for estimating a structured signal in the non-asymptotic case has been studied. However it is assumed that the loss function satisfies some conditions including restricted strong convexity, continuously differentiability in the first argument and derivative of loss function being Lipschitz-continuous with respect to the second argument. The author furthermore derives some sufficient conditions for the loss function ensuring the restricted strong convexity condition. Our result in Theorem III.2 for achieving the best performance across all loss functions is comparable to [34, Theorem 1] in which they have also proposed a method for deriving optimal loss function and measuring its performance. However their results hold for measurements of the form $y_i = a_i^T x_0 + \varepsilon_i$, where $\{\varepsilon_i\}_{i=1}^{m}$ are random errors independent of $a_i$’s.

Finally, our paper is closely related to [7], [28], in which the authors study the high-dimensional performance of maximum-likelihood (ML) estimation for the logistic model. The ML estimator is a special case of (1) but their measurement model differs from the one considered in this paper. Also, their analysis is based on the AMP. While this paper was being prepared, we became aware of [25], in which the authors extend the results of [28] to regularized ML by using the CGMT. However we present results for general loss functions and a different measurement model.

\section{Organization and notation}

The rest of the paper is organized as follows. Section II formally introduces the problem that this paper is concerned with. We present our main results Theorems III.1 and III.2 in Section III, where we also discuss some of their implications. In Section IV, we specialize the general result of Theorem III.1 to the LS, LAD and hinge-loss estimators. We also present numerical simulations to validate our theoretical predictions. We conclude in Section V with several possible directions for future research. A proof sketch of Theorem III.1 is provided in Appendix A.

The symbols $\mathbb{P}(\cdot)$ and $\mathbb{E}[\cdot]$ denote the probability of an event and the expectation of a random variable, respectively. We use boldface notation for vectors. $\|v\|_2$ denotes the Euclidean norm of a vector $v$. We write $i \in [m]$ for $i = 1, 2, \ldots, m$. When writing $x_* = \arg \min_x f(x)$, we let the operator $\arg \min$ return any one of the possible minimizers of $f$. For all $x \in \mathbb{R}$, Gaussian $Q$-function at $x$ is defined as $Q(x) = \int_{x}^{\infty} \frac{e^{-s^2/2}}{\sqrt{2\pi}} ds$. For a random variable
$H$ with density $p_H(h)$ that has a derivative $p_H'(h)$ for all $h \in \mathbb{R}$, denote its score function $\xi_h := \frac{\partial}{\partial h} \log p(h) = \frac{p_H'(h)}{p_H(h)}$. Fisher information of $H$ (e.g., [1, Sec. 2]) is defined as $\mathcal{I}(H) := \mathbb{E}[\xi_h^2]$.

II. Problem statement

The goal is to recover the unknown vector $x_0 \in \mathbb{R}^n$ from $m$ noisy signed measurements $y_i$. Let $\overline{y}_i$, $i \in [m]$ denote the noiseless signed measurements $\overline{y}_i = \text{sign}(a_i^T x_0)$. $i \in [m]$, where $a_i \in \mathbb{R}^n$ are the measurement vectors. We assume the following noise model in which each measurement is corrupted (i.e., sign flipped) with some constant probability $\varepsilon \in [0, 1/2]$:

$$y_i = \text{BSC}_\varepsilon(\overline{y}_i) := \begin{cases} \text{sign}(a_i^T x_0), & \text{w.p. } 1 - \varepsilon, \\ -\text{sign}(a_i^T x_0), & \text{w.p. } \varepsilon. \end{cases}$$ (3)

We remark that all our results remain valid in the case of (potentially) adversarial noise in which $\varepsilon m$ number of noiseless measurements $\overline{y}_i$ are flipped. Nevertheless, for the rest of the paper, we focus on the measurement model in (3). This paper studies the recovery performance of estimates $\hat{x}_\ell$ of $x_0$ that are obtained by solving the following convex optimization problem:

$$\hat{x}_\ell \in \arg \min_x \frac{1}{m} \sum_{i=1}^{m} \ell(y_i, a_i^T x),$$ (4)

Here, $\ell : \mathbb{R} \to \mathbb{R}$ is a convex loss function and the subscript $\ell$ on the estimate $\hat{x}_\ell$ emphasizes its dependence on the choice of the loss function. Different choices for $\ell(.)$ lead to popular specific estimators. For example, these include the following:

- Least-squares (LS): $\ell(t) = \frac{1}{2}(t - 1)^2$.
- Least-absolute deviations (LAD): $\ell(t) = |t - 1|$.
- Logistic maximum-likelihood: $\ell(t) = \log(1 + e^{-t})$.
- Ada-boost: $\ell(t) = e^{-t}$.
- Hinge-loss: $\ell(t) = \max(1 - t, 0)$.

Since we only observe sign-information, any information about the magnitude $\|x_0\|_2$ of the signal $x_0$ is lost. Thus, we can only hope to obtain an accurate estimate of the direction of $x_0$. We measure performance of the estimate $\hat{x}_\ell$ by its (absolute) correlation value to $x_0$, i.e.,

$$\text{corr} \left( \hat{x}_\ell ; x_0 \right) := \frac{|\langle \hat{x}_\ell, x_0 \rangle|}{\|\hat{x}_\ell\|_2 \|x_0\|_2} \in [0, 1].$$ (5)

Obviously, we seek estimates that maximize correlation.

III. Main results

A. Moreau Envelopes

Before presenting our main result, we need a few definitions. We write

$$\mathcal{M}_\ell(x; \lambda) := \min_v \frac{1}{2\lambda} (x - v)^2 + \ell(v),$$

for the Moreau envelope function of the loss $\ell : \mathbb{R} \to \mathbb{R}$ at $x$ with parameter $\lambda > 0$. Note that the objective function in the minimization above is strongly convex. Thus, for all values of $x$ and $\lambda$, there exists a unique minimizer which we denote by $\text{prox}_\lambda(\ell; x)$. This is known as the proximal operator of $\ell$ at $x$ with parameter $\lambda$. One of the important and useful properties of the Moreau envelope function is that it is continuously differentiable with respect to both $x$ and $\lambda$ [24]. We denote these derivatives as follows

$$\mathcal{M}'_{\ell,1}(x; \lambda) := \frac{\partial \mathcal{M}_\ell(x; \lambda)}{\partial x},$$

$$\mathcal{M}'_{\ell,2}(x; \lambda) := \frac{\partial \mathcal{M}_\ell(x; \lambda)}{\partial \lambda}.$$ (6)

The following is a well-known result that is useful for our purposes.

Proposition III.1 (Derivatives of $\mathcal{M}_\ell$ [24]). For a function $\ell : \mathbb{R} \to \mathbb{R}$ and all $x \in \mathbb{R}$ and $\lambda > 0$, the following properties are true:

- $\mathcal{M}'_{\ell,1}(x; \lambda) = \frac{1}{\lambda} (x - \text{prox}_\lambda(\ell; x))$,
- $\mathcal{M}'_{\ell,2}(x; \lambda) = -\frac{1}{2\lambda^2} (x - \text{prox}_\lambda(\ell; x))^2$.

If in addition $\ell$ is differentiable and $\ell'$ denotes its derivative, then

$$\mathcal{M}'_{\ell,1}(x; \lambda) = \ell'(\text{prox}_\lambda(\ell; x)),$$

$$\mathcal{M}'_{\ell,2}(x; \lambda) = -\frac{1}{2} (\ell'(\text{prox}_\lambda(\ell; x)))^2.$$
B. A system of equations

It turns out that the asymptotic performance of (4) depends on the loss function $\ell$ via its Moreau envelope. Specifically, define random variables $G$, $S$ and $Y$ as follows (recall the definition of BSC in (3))

$$G, S \overset{iid}{\sim} \mathcal{N}(0,1) \text{ and } Y = \text{BSC}_x(\text{sign}(S)),$$  \hspace{1cm} (8)

and consider the following system of non-linear equations in three unknowns $(\mu, \alpha, \lambda)$$:

$$\mathbb{E} \left[ Y \cdot S \cdot \ell'(\alpha G + \mu SY; \lambda) \right] = 0, \hspace{1cm} (9a)$$

$$\lambda^2 \delta \mathbb{E} \left[ \ell^2(\alpha G + \mu SY; \lambda) \right] = \alpha^2, \hspace{1cm} (9b)$$

$$\lambda \delta \mathbb{E} \left[ G \cdot \ell'(\alpha G + \mu SY; \lambda) \right] = \alpha. \hspace{1cm} (9c)$$

The expectations above are with respect to the randomness of the random variables $G$, $S$ and $Y$. As we shall shortly, the solution to these equations is tightly connected to the asymptotic behavior of the optimization in (4).

Remark 3 (Bounded minimizers). Theorem III.1 only holds in regimes for which the set of minimizers of (4) is bounded. As we show here, this is not always the case. Specifically, consider non-negative loss functions $\ell(t)$ with the property $\lim_{t \to +\infty} \ell(t) = 0$. For example, the hinge-loss, Ada-boost and logistic loss all satisfy this property. Now, we show that for such loss functions the set of minimizers is unbounded if $\delta < \delta_c^*$ for some appropriate $\delta^*_c > 2$. First, note that the set of minimizers is unbounded if the following condition holds:

$$\exists \ x_s \in \mathbb{R}^p \text{ such that } y_i a_i^T x_s \geq 1, \quad \forall \ i \in [m].$$  \hspace{1cm} (14)

Indeed, if (14) holds then $x = c \cdot x_s$ with $c \to +\infty$, attains zero cost in (4); thus, it is optimal and the set of minimizers is unbounded. To proceed, we rely on a recent result by
Candes and Sur [7] who prove that (14) holds iff
\[ \delta \leq \delta^*_\varepsilon := \left( \min_{\varepsilon \in \mathbb{R}} \mathbb{E} \left[ (G + \varepsilon SY)^2 \right] \right)^{-1}, \] (15)
where \( G, S \) and \( Y \) are random variables as in (8) and \( \{t\}_{-} := \min\{0, t\} \). It can be checked analytically that \( \delta^*(\varepsilon) \) is a decreasing function of \( \varepsilon \) with \( \delta^*(0^+) = +\infty \) and \( \delta^*(1/2) = 2 \). In Figure 1, we have numerically evaluated the threshold value \( \delta^*_\varepsilon \) as a function of the corruption level \( \varepsilon \). For \( \delta < \delta^*_\varepsilon \), the set of minimizers of the (4) with logistic or hinge loss is unbounded.

Remark 4 (Why \( \delta > 1 \)). The theorem assumes that \( \delta > 1 \) or equivalently \( m > n \). Here, we show that this condition is necessary for the equations (9) to have a bounded solution, or equivalently for the minimizer \( \tilde{x}_\ell \) in (4) to be bounded. To see this, take squares in both sides of (9c) and divide by (9b), to find that
\[ \delta = \frac{\mathbb{E} \left[ \left( \mathcal{M}_{\ell,1} (\alpha G + \mu SY; \lambda) \right)^2 \right]}{\left( \mathbb{E} \left[ G \cdot \mathcal{M}'_{\ell,1} (\alpha G + \mu SY; \lambda) \right] \right)^2} \geq 1. \]
The inequality follows by applying Cauchy-Schwarz and using the fact that \( \mathbb{E}[G^2] = 1 \).

Remark 5 (Solving the equations). Evaluating the performance of \( \tilde{x}_\ell \) requires solving the system of non-linear equations in (4). We empirically observe (see also [30] for similar observation) that if a solution exists, then it can be efficiently found by the following fixed-point iteration method. Let \( \mathbf{v} := [\mu, \alpha, \lambda]^T \) and \( \mathcal{F} : \mathbb{R}^3 \to \mathbb{R}^3 \) be such that (4) is equivalent to \( \mathbf{v} = \mathcal{F}(\mathbf{v}) \). With this notation, initialize \( \mathbf{v} = \mathbf{v}_0 \) and for \( k \geq 1 \) repeat the iterations \( \mathbf{v}_{k+1} = \mathcal{F}(\mathbf{v}_k) \) until convergence.

D. Performance bounds

In this section, we use the asymptotic prediction of Theorem III.1 to derive an upper bound on the correlation value \( \text{corr} (\tilde{x}_\ell; x_0) \) that holds for all choices of continuously differentiable loss functions. The following theorem is the key intermediate result in this direction. At a high-level, the proof of the theorem involves a careful algebraic manipulation of the system of equations (9), and, leveraging properties of the Moreau envelope function.

Theorem III.2 (Lower bound on \( \sigma_\ell \)). Let \( \ell \) be continuously differentiable and let assumptions and notation of Theorem III.1 hold. Then,
\[ \sigma^2_\ell \mathcal{I}(\sigma_\ell G + SY) \geq \frac{1}{\delta}. \] (16)

Proof. First, applying Gaussian integration by parts in (9c) we find that :
\[ 1 = \lambda \delta \mathbb{E}[\mathcal{M}''_{\ell,1} (\alpha G + \mu SY; \lambda)], \]
where \( \mathcal{M}''_{\ell,1} (x; \lambda) \) denotes the second derivative of the Moreau-envelope function with respect to the first argument (which exists since \( \ell \) is differentiable). Solving for \( \lambda \) and substituting in (9b) gives :
\[ \alpha^2 = \frac{1}{\delta} \cdot \frac{\mathbb{E} \left[ \left( \mathcal{M}_{\ell,1} (\alpha G + \mu SY; \lambda) \right)^2 \right]}{\left( \mathbb{E}[\mathcal{M}'_{\ell,1} (\alpha G + \mu SY; \lambda)] \right)^2}. \] (17)
The rest of the proof follows steps similar to [10, Lem. 3.4] and [30, Rem. 5.3.3]. Call
\[ Z := \mu SY \quad \text{and} \quad W := \alpha G + Z. \]
Note that for \( \alpha \neq 0, W \) has a continuously differentiable function at every \( w \); in fact,
\[ p'_W(w) = \int \phi'(u) p_Z(w - u) \, du, \]
where \( \phi(a) = \frac{1}{\alpha \sqrt{2\pi}} e^{-\frac{a^2}{2\alpha^2}}. \)
With this notation, by Cauchy-Schwartz inequality we have:
\[ \mathbb{E} \left[ \left( \mathcal{M}_{\ell,1} (W; \lambda) \right)^2 \right] \cdot \mathcal{I}(W) \geq \left( \mathbb{E}[\mathcal{M}_{\ell,1} (W; \lambda) \cdot \xi_W] \right)^2 \]
\[ = \left( \mathbb{E}[\mathcal{M}''_{\ell,1} (W; \lambda)] \right)^2, \]
where the last line follows by integration by parts. This, when combined with (17) shows that:
\[ \alpha^2 \geq \frac{1}{\delta} \cdot \frac{1}{\mathcal{I}(W)}. \] (18)
Recall that \( W = \alpha G + \mu SY \) and \( G \sim \mathcal{N}(0, 1) \). Now, since (e.g., [1, Eqn. 2.13])

\[
\mathcal{I}(c; H) = \mathcal{I}(H)/c^2,
\]

we have that

\[
\mathcal{I}(W) = \mathcal{I}\left( \mu \left( \frac{\alpha}{\mu} G + SY \right) \right) = \frac{1}{\mu^2} \mathcal{I}(\sigma_\ell G + SY),
\]

where recall that

\[
\sigma_\ell := \alpha/\mu.
\]

Substituting this in (18) gives the desired inequality in the statement of the theorem.

As it is illustrated in Figure 2, \( \sigma_\ell^2 \mathcal{I}(\sigma_\ell G + SY) \) is an increasing function of \( \sigma_\ell \). Therefore the minimum value of \( \sigma_\ell \) can be derived from the inequality in (16), which corresponds to the optimum \( \sigma_\ell \) that can be achieved for any continuously differentiable loss function. In the following two remarks, this result is extended to achieve the optimum values for correlation.

**Remark 6** (Upper bound on correlation). Using (16) with \( Y = \text{BSC}_\epsilon(\text{sign}(S)) \) and by numerically evaluating \( \mathcal{I}(\sigma_\ell G + SY) \) based on \( \sigma_\ell \), we can find a numerical lower bound on \( \sigma_\ell \). In view of (12) this directly yields an upper bound on \( \text{covr}(X_\ell; x_0) \), i.e., the maximum correlation that can be achieved for any choice of (continuously differentiable) convex loss function. In Figures 3-5, the green lines show the upper bound on correlation for specific values of \( \epsilon \); see Section IV-D for details.

**IV. Special cases**

In this section, we apply the general result of Theorem III.1 to specific popular choices of the loss function.

**A. Least-squares**

By choosing \( \ell(t) = (t-1)^2 \) in (4), we obtain the standard least squares estimate. To see this, note that since \( y_i = \pm 1 \), it holds for all \( i \) that \( (y_i a_i^T x - 1)^2 = (y_i - a_i^T x)^2 \).

Thus, the estimator \( \hat{x} \) is minimizing the sum of squares of the residuals:

\[
\hat{x} = \arg\min_x \sum (y_i - a_i^T x)^2. \tag{19}
\]

For the choice \( \ell(t) = (t-1)^2 \), it turns out that we can solve the equations in (9) in closed form. The final result is summarized in the corollary below.

**Corollary IV.1** (Least-squares). Let Assumption 1 hold and \( \delta > 1 \). Let \( \ell \) be as in (19). Then, in the limit of \( m, n \to +\infty \), \( m/n \to \delta \), Equations (12) and (13) hold with probability one with \( \alpha \) and \( \mu \) given as follows:

\[
\alpha = (1 - 2\epsilon) \sqrt{\frac{2}{\pi}}, \tag{20}
\]

\[
\mu = \frac{1}{2} \frac{1 - (1 - 2\epsilon)^2}{4(\delta - 1)}. \tag{21}
\]

**Proof.** In order to get the values of \( \alpha \) and \( \mu \) as in the statement of the corollary, we show how to simplify Equations (9) for \( \ell(t) = (t-1)^2 \). In this case, the proximal operator admits a simple expression:

\[
\text{prox}_\ell(x; \lambda) = (x + 2\lambda)/(1 + 2\lambda).
\]

Also, \( \ell'(t) = 2(t-1) \). Substituting these in (10a) gives the formula for \( \mu \) as follows:

\[
0 = \mathbb{E}[Y S(\alpha G + \mu SY - 1)] = \mu \mathbb{E}[S^2] - \mathbb{E}[YS]
\]

\[
\implies \mu = \sqrt{\frac{2}{\pi}}(1 - 2\epsilon),
\]

where we have also used from (8) that \( \mathbb{E}[S^2] = 1 \), \( \mathbb{E}[YS] = (1 - 2\epsilon)\sqrt{\frac{2}{\pi}} \) and \( G \) is independent of \( S \). Also, since \( \ell''(t) = 2 \), direct application of (11) gives

\[
1 = \lambda \delta \frac{2}{1 + 2\lambda} \implies \lambda = \frac{1}{2(\delta - 1)}.
\]

Finally, substituting the value of \( \lambda \) in (10b) we obtain the desired value for \( \alpha \) as follows

\[
\alpha^2 = 4\lambda^2 \delta \mathbb{E}[(\text{prox}_\ell(\alpha G + \mu SY; \lambda) - 1)^2]
\]

\[
= \frac{4\lambda^2}{(1 + 2\lambda)^2} \delta \mathbb{E}[(\alpha G + \mu SY - 1)^2]
\]

\[
= \frac{4\lambda^2 \delta}{(1 + 2\lambda)^2} \left( \alpha^2 + 1 - \frac{2}{\pi}(1 - 2\epsilon)^2 \right)
\]

\[
= \frac{1}{\delta} \left( \alpha^2 + 1 - \frac{2}{\pi}(1 - 2\epsilon)^2 \right) \implies (21).
\]

**Remark 7** (Least-squares: One-bit vs signed measurements). On the one hand, Corollary IV.1 shows that least-squares...
for (noisy) one-bit measurements lead to an estimator that satisfies
\[
\lim_{n \to \infty} \left\| \hat{x} - \frac{\mu}{\|x_0\|_2} \cdot x_0 \right\|_2^2 = \tau^2, \quad \frac{1}{\delta - 1},
\]
(22)
where \(\mu\) is as in (20) and \(\tau^2 := 1 - (1 - 2\epsilon)^2\). On the other hand, it is well-known (e.g., see references in [30, Sec. 5.1]) that least-squares for (scaled) linear measurements with additive Gaussian noise (i.e. \(y_i = \rho a_i^T x_0 + \sigma z_i, \ z_i \sim \mathcal{N}(0, 1)\)) leads to an estimator that satisfies
\[
\lim_{m \to \infty} \frac{1}{m} \left\| \hat{x} - \frac{\mu}{\|x_0\|_2} \cdot x_0 \right\|_2^2 = \frac{\sigma^2}{\delta - 1}.
\]
Direct comparison of (22) to (23) suggests that least-squares with one-bit measurements performs the same as if measurements were linear with scaling factor \(\rho = \mu/\|x_0\|_2\) and noise variance \(\sigma^2 = \tau^2 = \alpha^2(\delta - 1)\). This worth-mentioning conclusion is not new; it was proved in [6], [23], [29]. We elaborate on the relation to this prior work in the following remark.

Remark 8 (Prior work). There is a lot of recent work on the use of least-squares-type estimators for recovering signals from nonlinear measurements of the form \(y_i = f(a_i^T x_0)\) with Gaussian vectors \(a_i\). The original work that suggests least-squares as a reasonable estimator in this setting is due to Brillinger [6]. In his 1982 paper, Brillinger studied the problem in the classical statistics regime (aka \(n\) is fixed not scaling with \(m \to +\infty\)) and he proved for the least-squares solution satisfies
\[
\lim_{m \to +\infty} \frac{1}{m} \left\| \hat{x} - \frac{\mu}{\|x_0\|_2} \cdot x_0 \right\|_2^2 = \tau^2;
\]
where
\[
\mu = \mathbb{E}[S f(S)], \quad S \sim \mathcal{N}(0, 1),
\]
\[
\tau^2 = \mathbb{E}[(f(S) - \mu S)^2].
\]
and the expectations are with respect to \(S\) and possible randomness of \(f\). Evaluating (24) for \(f(S) = \text{BSC}_\epsilon(\text{sign}(S))\) leads to the same values for \(\mu\) and \(\tau^2\) in (22). In other works, (22) for \(\delta \to +\infty\) indeed recovers Brillinger’s result. The extension of Brillinger’s original work to the high-dimensional setting (both \(m, n\) large) was first studied by Plan and Vershynin [23], who derived (non-sharp) non-asymptotic upper bounds on the performance of constrained least-squares (such as the Lasso). Shortly after, [29] extended this result to sharp asymptotic predictions and to regularized least-squares. In particular, Corollary IV.1 is a special case of the main theorem in [29]. Several other interesting extensions of the result by Plan and Vershynin have recently appeared in the literature, e.g., [14], [16], [15], [32]. However, [29] is the only one to give results that are sharp in the flavor of this paper. Our work, extends the result of [29] to general loss functions beyond least-squares. The techniques of [29] that have guided the use of the CGMT in our context have also been recently applied in [9] in the context of phase-retrieval.
D. Numerical simulations

We present numerical simulations that validate the predictions of Theorems III.1 and III.2. For the numerical experiments, we generate random measurements according to (3) and Assumption 1. Without loss of generality (due to rotational invariance of the Gaussian measure) we set \( x_0 = [1, 0, ..., 0]^T \). We then obtain estimates \( \hat{x}_f \) of \( x_0 \) by numerically solving (4) and measuring performance by the correlation value \( \text{corr}(\hat{x}_f; x_0) \). Throughout the experiments, we set \( n = 128 \) and the recorded values of correlation in Figures 3–5 are averages over 25 independent experiments. The theoretical curves for the correlation are computed based on Theorem III.1. We solve the system of equations in (9) by the fixed-point iteration method described in Remark 5. The expectations involved in (9) are evaluated with Monte-Carlo estimation using \( 10^5 \) independent samples. Numerical upper bounds in Figures 3–5 are derived according to Remark 6.

Comparisons between theoretical and simulated values for LAD and LS estimators along with the upper bound on correlation are illustrated in Figure 3 for the noiseless case. Note that for \( \varepsilon = 0 \), the hinge-loss has an unbounded set of minimizers for all values of \( \delta \) (thus, Theorem III.1 is not applicable). In Figure 4, the probability of error \( \varepsilon \) is increased to 0.1. Note that in this setting hinge-loss estimator exists for \( \delta > \delta_{0,1} \approx 3 \) and that it outperforms LAD and LS estimators for large values of \( \delta \).

In Figure 5 we present similar results for \( \varepsilon = 0.25 \). As it is evident from Figures 4 and 5, the best estimator is varying based on the value of \( \delta \) and \( \varepsilon \). This further emphasizes the impact of studying the accuracy of estimators while we do not restrict ourselves to a specific loss function.

V. Conclusion

This paper derives sharp asymptotic performance guarantees for a wide class of convex optimization based estimators for recovering a signal from corrupted one-bit measurements. Our general result includes as a special case the least-squares estimator that was previously studied in [29]. Beyond that, it applies to other popular estimators such as the LAD, Hinge-loss, logistic loss, etc. One natural and interesting research direction is finding the optimal loss function \( \ell(\cdot) \) in (4). In view of Theorem III.1, this boils down to finding \( \ell(\cdot) \) that minimizes the ratio \( \alpha/\mu \) of the parameters \( \alpha \) and \( \mu \) that solve the system of equations in (9). For this purpose, it might also be important to derive necessary and sufficient conditions that guarantee (9) has a unique solution.

REFERENCES

[1] Andrew R Barron. Monotonic central limit theorem for densities. Department of Statistics, Stanford University, California, Tech. Rep, 50, 1984.
[2] Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. Information Theory, IEEE Transactions on, 57(2):764–785, 2011.
[3] Mohsen Bayati and Andrea Montanari. The lasso risk for gaussian matrices. Information Theory, IEEE Transactions on, 58(4):1997–2017, 2012.
[4] Andrew Mastin and Patrick Jailleth. Log-Quadratic Bounds for the Gaussian Q-function. arXiv preprint arXiv:1304.2488, 2013.
[5] Petros T Boufounos and Richard G Baraniuk. 1-bit compressive sensing. In 2008 42nd Annual Conference on Information Sciences and Systems (CISS), pages 16–21. IEEE, 2008.
[6] David R Brillinger. A generalized linear model with" gaussian" regressor variables. A Festschrift For Erich L. Lehmann, page 97, 1982.
[7] Emmanuel J Candès and Pragya Sur. The phase transition for the existence of the maximum likelihood estimate in high-dimensional logistic regression. arXiv preprint arXiv:1804.06733, 2018.
[8] Venkat Chandrasekaran, Benjamin Recht, Pablo A Parrilo, and Alan S Willsky. The convex geometry of linear inverse problems. Foundations of Computational Mathematics, 12(6):805–849, 2012.
In this section we provide a proof sketch of Theorem III.1. The main technical tool that facilitates our analysis is the convex Gaussian min-max theorem (CGMT), which is an extension of Gordon’s Gaussian min-max inequality (GMT). We introduce the necessary background on the CGMT in A.

A. Technical tool: CGMT

1) Gordon’s Min-Max Theorem (GMT): The Gordon’s Gaussian comparison inequality [17] compares the min-max value of two doubly indexed Gaussian processes based on how their autocorrelation functions compare. The inequality is quite general (see [17]), but for our purposes we only need its application to the following two Gaussian processes:

\[
X_{w,u} := u^T G w + \psi(w,u),
\]

\[
Y_{w,u} := \|w\|_2 g^T u + \|u\|_2 h^T w + \psi(w,u),
\]

where: \(G \in \mathbb{R}^{m \times n}, g \in \mathbb{R}^m, h \in \mathbb{R}^n\), they all have entries iid Gaussian; the sets \(S_w \subset \mathbb{R}^n\) and \(S_u \subset \mathbb{R}^m\) are compact; and, \(\psi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}\). For these two processes, define the following (random) min-max optimization programs, which (following [30]) we refer to as the primary optimization (PO) problem and the auxiliary optimization (AO) – for purposes that will soon become clear.

\[
\Phi(G) = \min_{w \in S_w} \max_{u \in S_u} X_{w,u},
\]

\[
\phi(g,h) = \min_{w \in S_w} \max_{u \in S_u} Y_{w,u}.
\]

According to Gordon’s comparison inequality, for any \(c \in \mathbb{R}\), it holds:

\[
\mathbb{P}(\Phi(G) < c) \leq 2\mathbb{P}(\phi(g,h) < c).
\]

In other words, a high-probability lower bound on the AO is a high-probability lower bound on the PO. The premise is that it is often much simpler to lower bound the AO rather than the PO. To be precise, (29) is a slight reformulation of Gordon’s original result proved in [31] (see therein for details).

2) Convex Gaussian Min-Max Theorem (CGMT): The proof of Theorem III.1 builds on the CGMT [31]. For ease of reference we summarize here the essential ideas of the framework following the presentation in [30]; please see [30, Section 6] for the formal statement of the theorem and further details. The CGMT is an extension of the GMT and it asserts that the AO in (28b) can be used to tightly infer properties of the original (PO) in (28a), including the optimal cost and the optimal solution. According to the CGMT [30, Theorem 6.1], if the sets \(S_w\) and \(S_u\) are convex and \(\psi\) is
continuous convex-concave on $\mathcal{S}_w \times \mathcal{S}_u$, then, for any $\nu \in \mathbb{R}$ and $t > 0$, it holds
\[
\mathbb{P}(\|\Phi(G) - \nu\| > t) \leq 2\mathbb{P}(\|\phi(g, h) - \nu\| > t).
\] (30)
In words, concentration of the optimal cost of the AO problem around $\mu$ implies concentration of the optimal cost of the corresponding PO problem around the same value $\mu$. Moreover, starting from (30) and under strict convexity conditions, the CGMT shows that concentration of the optimal solution of the AO problem implies concentration of the optimal solution of the PO to the same value. For example, if minimizers of (28b) satisfy $\|\mathbf{w}^*(g, h)\|_2 \to \zeta^*$ for some $\zeta^* > 0$, then, the same holds true for the minimizers of (28a): $\|\mathbf{w}^*(G)\|_2 \to \zeta^*$ [30, Theorem 6.1(iii)]. Thus, one can analyze the AO to infer corresponding properties of the PO, the premise being of course that the former is simpler to handle than the latter.

B. Applying the CGMT to (7)

In this section, we show how to apply the CGMT to (7). For convenience, we drop the subscript $\ell$ from $\hat{x}_\ell$ and simply write
\[
\hat{x} = \arg\min_{\mathbf{x}} \frac{1}{m} \sum_{i=1}^{m} \ell(y_i a_i^T \mathbf{x}) + \frac{1}{m} \sum_{i=1}^{m} \beta_i u_i - \frac{1}{m} \sum_{i=1}^{m} \beta_i y_i a_i^T \mathbf{x} + r \|\mathbf{x}\|_2^2.
\] (31)
where the measurements $y_i$, $i \in [m]$ follow (3). By rotational invariance of the Gaussian distribution of the measurement vectors $a_i$, $i \in [m]$, we assume without loss of generality that $x_0 = [1, 0, \ldots, 0]^T$. Denoting $y_i a_i^T \mathbf{x}$ by $u_i$, (31) is equivalent to the following min-max optimization:
\[
\min_{\mathbf{u}, \mathbf{x}} \max_\beta \frac{1}{m} \sum_{i=1}^{m} \ell(u_i) + \frac{1}{m} \sum_{i=1}^{m} \beta_i u_i - \frac{1}{m} \sum_{i=1}^{m} \beta_i y_i a_i^T \mathbf{x} + r \|\mathbf{x}\|_2^2.
\] (32)
Now, let us define
\[
a_i = [s_i; \tilde{a}_i], \quad i \in [m] \quad \text{and} \quad \mathbf{x} = [x_1; \tilde{\mathbf{x}}],
\]
such that $s_i$ and $x_1$ are the first entries of $a_i$ and $\mathbf{x}$, respectively. Note that in this new notation (6) becomes:
\[
y_i = \begin{cases} 1 & \text{, w.p. } f(s), \\ -1 & \text{, w.p. } 1 - f(s), \end{cases}
\] (33)
and
\[
corr(\mathbf{x}; x_0) = \sqrt{\frac{\tilde{x}_1}{\tilde{x}_1^2 + \|\tilde{\mathbf{x}}\|_2^2}}.
\] (34)
where we denote $\tilde{\mathbf{x}} = [\tilde{x}_1; \tilde{\mathbf{x}}]$. Also, (32) is written as
\[
\min_{\mathbf{u}, \mathbf{x}} \max_\beta \frac{1}{m} \sum_{i=1}^{m} \ell(u_i) + \frac{1}{m} \sum_{i=1}^{m} \beta_i u_i + \frac{1}{m} \sum_{i=1}^{m} \beta_i y_i a_i^T \mathbf{x} - \frac{1}{m} \sum_{i=1}^{m} \beta_i y_i s_i x_1 + r x_1^2 + r \|\tilde{\mathbf{x}}\|_2^2.
\]
or, in matrix form:
\[
\min_{\mathbf{u}, \mathbf{x}} \max_\beta \frac{1}{m} \beta^T D_y \hat{A} \hat{x} + \frac{1}{m} x_1 \beta^T D_y s - \frac{1}{m} \sum_{i=1}^{m} \ell(u_i),
\] (35)
where $D_y := \text{diag}(y_1, y_2, \ldots, y_m)$ is a diagonal matrix with $y_1, y_2, \ldots, y_m$ on the diagonal, $s = [s_1, \ldots, s_m]^T$ and $\hat{A}$ is an $m \times (n - 1)$ matrix with rows $\hat{a}_1, i \in [m]$.

In (35) we recognize that the first term has the bilinear form required by the GMT in (28a). The rest of the terms form the function $\psi$ in (28a); they are independent of $\hat{A}$ and convex-concave as desired by the CGMT. Therefore, we have expressed (31) in the desired form of a PO and for the rest of the proof we will analyze the probabilistically equivalent AO problem. In view of (28b), this is given as follows,
\[
\min_{\mathbf{u}, \mathbf{x}} \max_\beta \frac{1}{m} \mathbf{x}^T D_y \beta + \frac{1}{m} \|D_y \beta\|_2 h^T \hat{x} - \frac{1}{m} x_1 \beta^T D_y s - \frac{1}{m} \beta^T \mathbf{u} + r x_1^2 + r \|\hat{x}\|_2^2 + \frac{1}{m} \sum_{i=1}^{m} \ell(u_i),
\] (36)
where as in (28b) $g \sim \mathcal{N}(0, I_m)$ and $h \sim \mathcal{N}(0, I_{m-1})$.

C. Analysis of the Auxiliary Optimization

Here, we show how to analyze the AO in (36) \footnote{There are several technical details in the proof that we omit from this proof sketch. This includes: boundedness of the constraint sets in (36); changing the order of min-max when optimizing over the direction of $\hat{x}$ in (37); and, uniform convergence in going from (39) to (40.).}. The steps are similar and follow the recipe prescribed in [31], [30]. To begin with, note that $y_i \in \pm 1$, therefore $D_y g \sim \mathcal{N}(0, I_m)$ and $\|D_y \beta\|_2 = \|\beta\|_2$. Let us denote
\[
\alpha := \|\hat{x}\|_2 \quad \text{and} \quad \mu := x_1.
\]
We can simplify (36) by optimizing over the direction of $\hat{x}$. This leads to the following:
\[
\min_{\alpha \geq 0, \mu, \mathbf{u}} \max_\beta \frac{1}{m} \alpha g^T \beta - \frac{\alpha}{m} \|\beta\|_2 \|h\|_2 - \frac{1}{m} s^T D_y \beta + \frac{1}{m} \beta^T u + r \mu^2 + r \alpha^2 + \frac{1}{m} \sum_{i=1}^{m} \ell(u_i).
\] (37)
Next, let $\gamma := \|\beta\|_2/\sqrt{m}$ and optimize over the direction of $\beta$ to yield
\[
\min_{\alpha \geq 0, \mu, \mathbf{u}} \max_\gamma \frac{\gamma}{\sqrt{m}} \|ag - \mu D_y s + u\|_2 - \frac{\alpha}{\sqrt{m}} \|h\|_2 + r \mu^2 + r \alpha^2 + \frac{1}{m} \sum_{i=1}^{m} \ell(u_i).
\] (38)
To continue, we utilize the fact that for all $x \in \mathbb{R}$, $\min_{r \geq 0} \frac{x^2}{2r} + \frac{r^2}{2m} = \frac{x^2}{2m}$. With this trick, the optimization
over \( u \) becomes separable over its coordinates and (38) can be rewritten as

\[
\min_{\alpha \geq 0, \tau > 0, u, \mu, \gamma \geq 0} \max_{m, n} \frac{\gamma \tau}{2} + \frac{\gamma}{2 \tau m} \| \alpha g - \mu D_y s + u \|_2^2 + r \mu^2 + r \alpha^2 + \frac{1}{m} \sum_{i=1}^m \ell(u_i) - \frac{\alpha}{\sqrt{m}} \gamma \| h \|_2,
\]

Equivalently, we express this in the following convenient form:

\[
\min_{\mu, \alpha \geq 0, \tau > 0} \max_{\gamma \geq 0} \frac{\gamma \tau}{2} - \frac{\alpha}{\sqrt{m}} \gamma \| h \|_2 + r \mu^2 + r \alpha^2 + \frac{1}{m} \sum_{i=1}^m M_\ell \left( \alpha G + \mu Y S; \frac{\tau}{\gamma} \right),
\]

where recall the definition of the Moreau envelope:

\[
M_\ell \left( \alpha g_i + \mu y_i s_i; \frac{\tau}{\gamma} \right) = \min_{u_i} \frac{\gamma}{2\tau} (\alpha g_i + \mu y_i s_i - u_i)^2 + \ell(u_i).
\]

As to now, we have reduced the AO into a random minimax optimization over only four scalar variables in (39). For fixed \( \mu, \alpha, \tau, \gamma \), direct application of the weak law of large numbers, shows that the objective function of (39) converges in probability to the following as \( m, n \to \infty \):

\[
\frac{\gamma \tau}{2} - \frac{\alpha \gamma}{\sqrt{\delta}} + r \mu^2 + r \alpha^2 + \mathbb{E} \left[ M_\ell \left( \alpha G + \mu Y S; \frac{\tau}{\gamma} \right) \right],
\]

where \( G, S \sim \mathcal{N}(0, 1) \) and \( Y \sim 2\text{Bernoulli}(f(S)) - 1 \) (in view of (33)). Based on that, it can be shown (see [30], [9] for similar arguments; details are deferred to the long version of the paper) that the random optimizers \( \alpha_n \) and \( \mu_n \) of (39) converge to the deterministic optimizers \( \alpha \) and \( \mu \) of the following (deterministic) optimization problem (whenever these are bounded as the statement of the theorem requires):

\[
\min_{\alpha \geq 0, \mu, \tau > 0} \max_{\gamma \geq 0} \frac{\gamma \tau}{2} - \frac{\alpha \gamma}{\sqrt{\delta}} + r \mu^2 + r \alpha^2 + \mathbb{E} \left[ M_\ell \left( \alpha G + \mu Y S; \frac{\tau}{\gamma} \right) \right].
\]

D. First-order optimality conditions

By direct differentiation, the first order optimality conditions of the min-max optimization in (40) are as follows:

\[
2r \mu + \mathbb{E} \left[ Y S M_{\ell,1}' \left( \alpha G + \mu Y S; \frac{\tau}{\gamma} \right) \right] = 0, \tag{41}
\]

\[
2r \alpha + \mathbb{E} \left[ G M_{\ell,1}' \left( \alpha G + \mu Y S; \frac{\tau}{\gamma} \right) \right] = \frac{\gamma}{\sqrt{\delta}}, \tag{42}
\]

\[
\frac{\gamma}{2} + \frac{1}{\gamma} \mathbb{E} \left[ M_{\ell,2}' \left( \alpha G + \mu Y S; \frac{\tau}{\gamma} \right) \right] = 0, \tag{43}
\]

\[
- \frac{\alpha}{\sqrt{\delta}} - \frac{\tau}{\gamma^2} \mathbb{E} \left[ M_{\ell,2}' \left( \alpha G + \mu Y S; \frac{\tau}{\gamma} \right) \right] + \frac{\tau}{2} = 0. \tag{44}
\]

Next, we show how these equations simplify to the following system of equation:

\[
\mathbb{E} \left[ Y S \cdot M_{\ell,1}' \left( \alpha G + \mu Y S; \lambda \right) \right] = -2r \mu, \tag{45a}
\]

\[
\lambda^2 \delta \mathbb{E} \left[ \left( M_{\ell,1}' \left( \alpha G + \mu Y S; \lambda \right) \right)^2 \right] = \alpha^2, \tag{45b}
\]

\[
\lambda \delta \mathbb{E} \left[ G \cdot M_{\ell,1}' \left( \alpha G + \mu Y S; \lambda \right) \right] = \alpha (1 - 2r \lambda \delta). \tag{45c}
\]

Denote \( \lambda := \frac{\tau}{\gamma} \). First, (45a) is followed by equation (41). Second, substituting \( \gamma \) from (43) in (44) yields \( \tau = \frac{\alpha}{\sqrt{\delta}} \) or \( \gamma = \frac{\alpha}{\lambda \sqrt{\delta}} \), which together with (42) leads to (45c). Finally, (45b) can be obtained by substituting \( \gamma = \frac{\alpha}{\lambda \sqrt{\delta}} \) in (43) and using the relation:

\[
M_{\ell,2}' \left( \alpha G + \mu Y S; \lambda \right) = \frac{1}{2} (M_{\ell,1}' \left( \alpha G + \mu Y S; \lambda \right))^2.
\]

As already mentioned in Section II, we focus on the non-regularized loss functions in this paper. Setting \( r = 0 \) in (45) will give the desired system of equations in (9)