2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)
Allerton Park and Retreat Center
Monticello, IL, USA, September 24-27, 2019

Sharp Guarantees for Solving Random Equations with One-Bit Information

Hossein Taheri, Ramtin Pedarsani and Christos Thrampoulidis
University of California, Santa Barbara

***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, the sharp nature of our results allows for accurate comparisons between these different 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 $m, n \to +\infty$ and $m/n \to \delta \in (1, +\infty)$. 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.

Electrical and Computer Engineering Department, University of California, Santa Barbara, Santa Barbara, CA 93106, USA. (e-mails: hossein@ucsb.edu, ramtin@ucsb.edu, cthrampo@ucsb.edu.)
This research was supported by NSF Grant CCF-1755808 and Academic Senate Research Grant from UCSB.

B. Contributions

Our goal is to recover $\mathbf{x}_0 \in \mathbb{R}^n$ from measurements $y_i = \text{sign}(\mathbf{a}_i^T \mathbf{x}_0),\ \ i = 1, \ldots, m,$ where $\mathbf{a}_i \in \mathbb{R}^n$ 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 $\hat{\mathbf{x}}_\ell$ that are solutions to the following optimization problem for some convex loss function $\ell(\cdot)$.

$$\widehat{\mathbf{x}}_\ell := \arg\min_{\mathbf{x}} \sum_{i=1}^{m} \ell(y_i \mathbf{a}_i^T \mathbf{x}). \tag{1}$$

When $m, n \to +\infty$ and $m/n \to \delta > \delta_\varepsilon^\star$, we show that the correlation of $\widehat{\mathbf{x}}_\ell$ to the true vector $\mathbf{x}_0$ is sharply predicted by $\sqrt{\frac{1}{1+(\alpha/\mu)^2}}$ where the parameters α and μ 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 α/μ) depends on the loss function $\ell(\cdot)$ through its Moreau envelope function. We prove that $\delta_\varepsilon^\star > 1$ is necessary for the equations to have a bounded solution, but, in general, the value of the threshold $\delta_\varepsilon^\star$ depends both on the noise level ε and on the loss function.

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 and recover the result of [29] (see Section I-C). For the hinge-loss, we show that $\delta_\varepsilon^\star$ is a decreasing function of ε 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 $\delta_\varepsilon^\star$ for general loss functions.

C. 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

$$\widehat{\mathbf{x}} := \arg\min_{\mathbf{x}} \sum_{i=1}^{m} \widetilde{\ell}(y_i - \mathbf{a}_i^T\mathbf{x}) + \lambda R(\mathbf{x}) \qquad (2)$$

Specifically, the loss function $\widetilde{\ell}(\cdot)$ 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[1] than (2). To see this, note that for $y_i \in \pm 1$ and popular symmetric loss functions $\widetilde{\ell}(t) = \widetilde{\ell}(-t)$ (e.g., LS, LAD), (1) results in (2) by choosing $\ell(t) = \widetilde{\ell}(t-1)$ 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 $\ell(\langle \mathbf{a}_i, \mathbf{x}\rangle, y_i)$ 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.1 is comparable to [34] in which the authors have also proposed a method for deriving optimal loss function and measuring its performance. However their results hold for measurements of the form $y_i = \mathbf{a}_i^T\mathbf{x}_0 + \varepsilon_i$, where $\{\varepsilon_i\}_{i=1}^m$ are random errors independent of $\mathbf{a}_i$'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. While we do not account for regularization in this paper, we present results for general loss functions and a different measurement model.

D. 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 result Theorem III.1 in Section III, where we also discuss some of its 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.

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. $\|\mathbf{v}\|_2$ denotes the Euclidean norm of a vector $\mathbf{v}$. 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.

II. Problem Statement

The goal is to recover the unknown vector $\mathbf{x}_0 \in \mathbb{R}^n$ from m noisy signed measurements y_i. Let $\overline{y}_i,\ i \in [m]$ denote the *noiseless* signed measurements $\overline{y}_i = \text{sign}(\mathbf{a}_i^T\mathbf{x}_0),\ i \in [m]$, where $\mathbf{a}_i \in \mathbb{R}^n$ are the measurement vectors. We assume the following noise model in which each measurement is

[1]This is modulo the regularization term $R(\cdot)$ in (2), which is beyond our scope of this paper.

corrupted (i.e., sign flipped) with some constant probability $\varepsilon \in [0, 1/2]$:

$$y_i = \mathrm{BSC}_\varepsilon(\overline{y}_i) := \begin{cases} \mathrm{sign}(\mathbf{a}_i^T\mathbf{x}_0) & , \text{w.p. } \ 1-\varepsilon, \\ -\mathrm{sign}(\mathbf{a}_i^T\mathbf{x}_0) & , \text{w.p. } \ \varepsilon. \end{cases} \tag{3}$$

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 $\widehat{\mathbf{x}}_\ell$ of $\mathbf{x}_0$ that are obtained by solving the following convex optimization program,

$$\widehat{\mathbf{x}}_\ell \in \arg\min_{\mathbf{x}} \sum_{i=1}^m \ell(y_i \mathbf{a}_i^T\mathbf{x}). \tag{4}$$

Here, $\ell : \mathbb{R} \to \mathbb{R}$ is a convex loss function and the subscript ℓ on the estimate $\widehat{\mathbf{x}}_\ell$ emphasizes its dependence on the choice of the loss function in (4). Different choices 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 $\|\mathbf{x}_0\|_2$ of the signal $\mathbf{x}_0$ is lost. Thus, we can only hope to obtain an accurate estimate of the direction of $\mathbf{x}_0$. We measure performance of the estimate $\widehat{\mathbf{x}}_\ell$ by its (absolute) correlation value to $\mathbf{x}_0$, i.e.,

$$\mathrm{corr}\left(\widehat{\mathbf{x}}_\ell\,;\,\mathbf{x}_0\right) := \frac{|\langle \widehat{\mathbf{x}}_\ell, \mathbf{x}_0\rangle|}{\|\widehat{\mathbf{x}}_\ell\|_2\|\mathbf{x}_0\|_2} \in [0,1]. \tag{5}$$

Of course, we seek estimates that maximize correlation.

Our main result characterizes the asymptotic value of $\mathrm{corr}\left(\widehat{\mathbf{x}}_\ell\,;\,\mathbf{x}_0\right)$ in the linear high-dimensional regime in which the problem dimensions m and n grow proportionally to infinity with $m/n \to \delta \in (1,\infty)$. All our results are valid under the assumption that the measurement vectors have entries IID Gaussian.

Assumption 1 (Gaussian measurement vectors)**.** *The vectors* $\mathbf{a}_i,\ i \in [m]$ *have entries IID standard normal* $\mathcal{N}(0,1)$.

We make no further assumptions on the distribution of the true vector $\mathbf{x}_0$.

III. General Result

A. Moreau Envelopes

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

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

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

$$\mathcal{M}_{\ell,1}^{'}\left(x;\tau\right) := \frac{\partial \mathcal{M}_\ell\left(x;\tau\right)}{\partial x},$$
$$\mathcal{M}_{\ell,2}^{'}\left(x;\tau\right) := \frac{\partial \mathcal{M}_\ell\left(x;\tau\right)}{\partial \tau}.$$

The following is a well-known result involves 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*

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

If in addition ℓ *is differentiable and* $\ell^{'}$ *denotes its derivative, then*

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

B. A system of equations

It turns out, that the asymptotic performance of (4) depends on the loss function ℓ via its Moreau envelope. Specifically, let random variables G, S and Y defined as follows (recall the definition of BSC_ε in (3))

$$G, S \overset{\text{iid}}{\sim} \mathcal{N}(0,1) \quad \text{and} \quad Y = \mathrm{BSC}_\varepsilon(\mathrm{sign}(S)), \tag{6}$$

and consider the following system of non-linear equations in three unknowns $(\mu, \alpha > 0, \lambda)$:

$$0 = \mathbb{E}\left[Y\,S \cdot \mathcal{M}_{\ell,1}^{'}\left(\alpha G + \mu SY;\lambda\right)\right], \tag{7a}$$

$$\alpha^2 = \lambda^2\,\delta\,\mathbb{E}\left[\left(\mathcal{M}_{\ell,1}^{'}\left(\alpha G + \mu SY;\lambda\right)\right)^2\right], \tag{7b}$$

$$\alpha = \lambda\,\delta\,\mathbb{E}\left[G \cdot \mathcal{M}_{\ell,1}^{'}\left(\alpha G + \mu SY;\lambda\right)\right]. \tag{7c}$$

The expectations above are with respect to the randomness of the random variables G, S and Y.

As we show shortly, the solution to these equations is tightly connected to the asymptotic behavior of the optimization in (4).

We remark that the equations are well defined even if the loss function ℓ is not differentiable. If ℓ is differentiable then, using Proposition III.1 the Equations (7) can be equivalently written as follows:

$$0 = \mathbb{E}\left[Y\, S \cdot \ell^{'}\left(\text{prox}_{\ell}\left(\alpha G + \mu SY; \lambda\right)\right)\right] \quad \text{(8a)}$$

$$\alpha^2 = \lambda^2\, \delta\, \mathbb{E}\left[\left(\ell^{'}\left(\text{prox}_{\ell}\left(\alpha G + \mu SY; \lambda\right)\right)\right)^2\right], \quad \text{(8b)}$$

$$\alpha = \lambda\, \delta\, \mathbb{E}\left[G \cdot \ell^{'}\left(\text{prox}_{\ell}\left(\alpha G + \mu SY; \lambda\right)\right)\right]. \quad \text{(8c)}$$

Finally, if ℓ is two times differentiable then applying integration by parts in Equation (8c) results in the following reformulation of (7c):

$$1 = \lambda\, \delta\, \mathbb{E}\left[\frac{\ell^{''}\left(\text{prox}_{\ell}\left(\alpha G + \mu SY; \lambda\right)\right)}{1 + \lambda\, \ell^{''}\left(\text{prox}_{\ell}\left(\alpha G + \mu SY; \lambda\right)\right)}\right]. \quad (9)$$

C. Asymptotic prediction

We are now ready to state the main result of this paper.

Theorem III.1 (General loss function)**.** *Let Assumption 1 hold and fix some $\varepsilon \in [0, 1/2]$ in (3). Assume $\delta > 1$ such that the set of minimizers in (4) is bounded and the system of Equations (7) has a unique solution (μ, α, λ), such that $\mu \neq 0$. Let $\widehat{\mathbf{x}}_\ell$ be as in (4). Then, in the limit of $m, n \to +\infty$, $m/n \to \delta$, it holds with probability one that*

$$\lim_{n\to\infty} \text{corr}\left(\widehat{\mathbf{x}}_\ell\,;\, \mathbf{x}_0\right) = \sqrt{\frac{1}{1 + (\alpha/\mu)^2}}. \quad (10)$$

Moreover,

$$\lim_{n\to\infty} \left\|\widehat{\mathbf{x}}_\ell - \mu \cdot \frac{\mathbf{x}_0}{\|\mathbf{x}_0\|_2}\right\|_2^2 = \alpha^2. \quad (11)$$

Theorem III.1 holds for general loss functions. In Section IV we specialize the result to specific popular choices. We also present numerical simulations that confirm the validity of the predictions of Theorem III.1 (see Figures 2–4). Before that, in Section III-D we present a few remarks on the conditions, interpretation and implications of the theorem.

D. Discussion

Remark 1 (The role of μ and α). The theorem evaluates the asymptotic performance of the estimator $\widehat{\mathbf{x}}_\ell$ for a convex loss function ℓ in (4). According to (10), the prediction for the limiting behavior of the correlation value is given in terms of $\sigma_\ell := \alpha/\mu$, where μ and α are unique solutions of (7). *The smaller the value of σ_ℓ is, the larger becomes the correlation value.* Thus, the correlation value is fully determined by the ratio of the parameters α and μ. Their individual role is clarified in (11). Specifically, according to (11), $\widehat{\mathbf{x}}$ is a biased estimate of the true $\mathbf{x}_0$ and μ represents exactly that bias term. In other words, solving (4) returns an estimator that is close to a μ–scaled version of $\mathbf{x}_0$. When $\mathbf{x}_0$ and $\widehat{\mathbf{x}}$ are scaled appropriately, then the L2 squared norm of their different converges to α^2.

Remark 2 (On the existence of a solution to (4)). While $\delta > 1$ is a necessary condition for the equations in (4) to have a solution, it is *not* sufficient in general. This depends on the specific choice of the loss function. For example, in Section IV-A, we show that for the squared loss $\ell(t) = (t-1)^2$, the equations have a unique solution iff $\delta > 1$. On the other hand, for logistic regression and hinge-loss, it is argued in Remark 3 that there exists a threshold value $\delta_\varepsilon^\star := \delta^\star(\varepsilon) > 2$ such that the set of minimizers in (4) is unbounded if $\delta > \delta_\varepsilon$. Hence, the theorem does not hold for $\delta < \delta_\varepsilon^\star$. We conjecture that for these choices of loss, the equations (4) are solvable iff $\delta > \delta_\varepsilon$. Justifying this conjecture is an interesting direction for future work. More generally, we leave the study of sufficient and necessary conditions under which the equations (4) admit a unique solution to future work.

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 $\ell(t) \geq 0$ 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 $\delta < \delta_\varepsilon^\star$ for some appropriate $\delta_\varepsilon^\star > 2$. First, note that the set of minimizers is unbounded if the following condition holds:

$$\exists\; \mathbf{x}_s \in \mathbb{R}^p \quad \text{such that} \quad y_i \mathbf{a}_i^T \mathbf{x}_s \geq 1, \quad \forall\; i \in [m]. \quad (12)$$

Indeed, if (12) holds then $\mathbf{x} = c \cdot \mathbf{x}_s$ 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 (12) holds iff [2]

$$\delta \leq \delta_\varepsilon^\star := \min_{c\in\mathbb{R}} \mathbb{E}\left[(G + c\, S\, Y)_-^2\right], \quad (13)$$

where G, S and Y are random variables as in (6) and $(t)_- := \min\{0, t\}$. It can be checked analytically that $\delta^\star(\varepsilon)$ is a decreasing function of ε with $\delta^\star(0^+) = +\infty$ and $\delta^\star(1/2) = 2$. In Figure 1, we have numerically evaluated the threshold value $\delta_\varepsilon^\star$ as a function of the corruption level ε. For $\delta < \delta_\varepsilon^\star$, the set of minimizers of the (4) with logistic or hinge loss is unbounded. An interesting direction for future investigations is to consider regularized versions of (4). The addition of a regularization term (e.g. ridge-regularization) will improve the range of δ's for which the set of minimizers in (4) is bounded and an analogue of theorem III.1 can be established.

Remark 4 (Solving the equations). Evaluating the performance of $\widehat{\mathbf{x}}_\ell$ 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

[2]To be precise, [7] prove the statement for measurements $y_i,\ i \in [m]$ that follow a logistic model. Close inspection of their proof shows that this requirement can be relaxed by appropriately defining the random variable Y in (13).

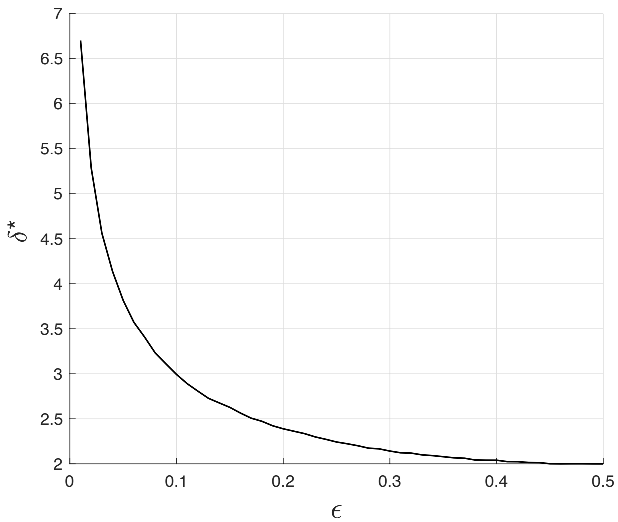

Fig. 1. The value of the threshold $\delta^\star_\varepsilon$ in (13) as a function of probability of error $\varepsilon \in [0, 1/2]$. For logistic and hinge-loss, the set of minimizers in (4) is bounded (as required by Theorem III.1) iff $\delta > \delta^\star_\varepsilon$. See Remark 3 and [7].

(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.

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 $(y_i \mathbf{a}_i^T \mathbf{x} - 1)^2 = (y_i - \mathbf{a}_i^T \mathbf{x})^2$.

Thus, the estimator $\widehat{\mathbf{x}}$ is minimizing the sum of squares of the residuals:

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \sum (y_i - \mathbf{a}_i^T \mathbf{x})^2. \tag{14}$$

For the choice $\ell(t) = (t-1)^2$, it turns out that we can solve the equations in (7) 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 $\widehat{\mathbf{x}}$ be as in (20). Then, in the limit of $m, n \to +\infty$, $m/n \to \delta$, Equations (10) and (11) hold with probability one with α and μ given as follows:*

$$\mu = (1 - 2\epsilon)\sqrt{\frac{2}{\pi}}, \tag{15}$$

$$\alpha^2 = \left(1 - (1 - 2\epsilon)^2 \, \frac{2}{\pi}\right) \frac{1}{\delta - 1}. \tag{16}$$

Proof. In order to get the values of α and μ as in the statement of the corollary, we show how to simplify Equations (7) for $\ell(t) = (t-1)^2$. In this case, the proximal operator admits a simple expression:

$$\mathrm{prox}_\ell\,(x; \lambda) = (x + 2\lambda)\Big/(1 + 2\lambda).$$

Also, $\ell^{'}(t) = 2(t-1)$. Substituting these in (8a) gives the formula for μ as follows:

$$0 = \mathbb{E}\left[YS(\alpha G + \mu SY - 1)\right] = \mu\,\mathbb{E}[S^2] - \mathbb{E}[YS]$$
$$\implies \mu = \sqrt{\frac{2}{\pi}}(1 - 2\epsilon),$$

where we have also used from (6) that $\mathbb{E}[S^2] = 1$, $\mathbb{E}[YS] = (1 - 2\varepsilon)\sqrt{\frac{2}{\pi}}$ and G is independent of S. Also, since $\ell^{''}(t) = 2$, direct application of (9) gives

$$1 = \lambda\delta\,\frac{2}{1 + 2\lambda} \implies \lambda = \frac{1}{2(\delta - 1)}.$$

Finally, substituting the value of λ in (8b) we obtain the desired value for α as follows

$$\begin{aligned}
\alpha^2 &= 4\lambda^2\delta\,\mathbb{E}\left[(\mathrm{prox}_\ell\,(\alpha G + \mu SY; \lambda) - 1)^2\right] \\
&= \frac{4\lambda^2}{(1 + 2\lambda)^2}\delta\,\mathbb{E}\left[(\alpha G + \mu SY - 1)^2\right] \\
&= \frac{4\lambda^2\delta}{(1 + 2\lambda)^2}(\alpha^2 + 1 - \frac{2}{\pi}(1 - 2\epsilon)^2) \\
&= \frac{1}{\delta}(\alpha^2 + 1 - \frac{2}{\pi}(1 - 2\epsilon)^2) \quad \implies (16).
\end{aligned}$$

□

Remark 5 (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

$$\lim_{n\to\infty} \left\| \widehat{\mathbf{x}}_\ell - \frac{\mu}{\|\mathbf{x}_0\|_2} \cdot \mathbf{x}_0 \right\|_2^2 = \tau^2 \cdot \frac{1}{\delta - 1}, \tag{17}$$

where μ is as in (15) and $\tau^2 := 1 - (1 - 2\varepsilon)\frac{2}{\pi}$. 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., $y_i = \rho \mathbf{a}_i^T \mathbf{x}_0 + \sigma z_i$, $z_i \sim \mathcal{N}(0, 1)$) leads to an estimator that satisfies

$$\lim_{n\to\infty} \|\widehat{\mathbf{x}}_\ell - \rho \cdot \mathbf{x}_0\|_2^2 = \sigma^2 \cdot \frac{1}{\delta - 1}. \tag{18}$$

Direct comparison of (17) to (18) suggests that least-squares with one-bit measurements performs the same as if measurements were linear with scaling factor $\rho = \mu/\|\mathbf{x}_0\|_2$ 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 6 (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 $y_i = f(\mathbf{a}_i^T \mathbf{x}_0)$ with Gaussian vectors $\mathbf{a}_i$. The original work that suggests least-squares as a reasonable estimator in this setting is due

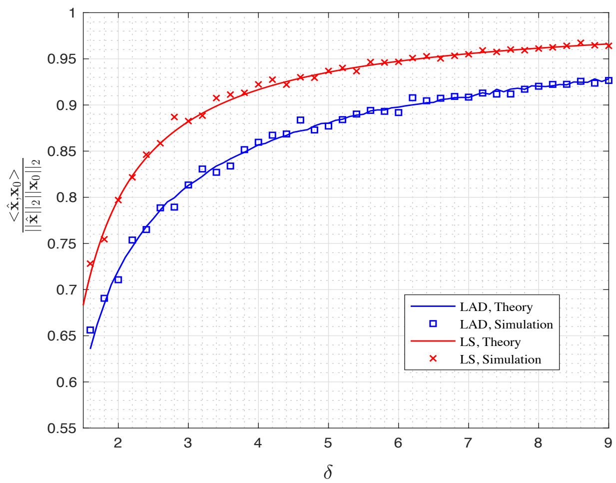

Fig. 2. Comparisons between theoretical and simulated results for the least-squares (LS) and least-absolute deviations (LAD) estimators as a function of δ, for noiseless measurements ($\epsilon = 0$). The LS estimator significantly outperforms the LAD for all values of δ.

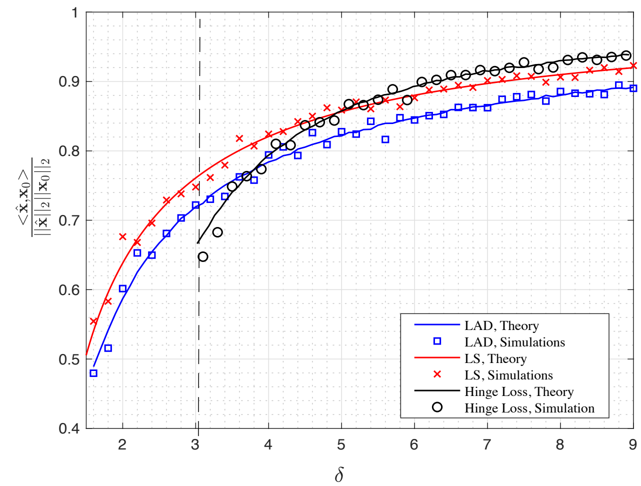

Fig. 3. Comparison between theoretical and simulated results for LAD, LS and Hinge-Loss estimators as a function of δ for probability of error $\epsilon = 0.1$. The dashed line represents the value of the threshold $\delta_\varepsilon^\star$ for $\epsilon = 0.1$ (see Figure 1). For small values of δ LS outperforms the other two estimators, but the hinge-loss becomes better as δ in increases.

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

$$\lim_{m\to+\infty} \frac{1}{m}\|\widehat{\mathbf{x}}_\ell - \frac{\mu}{\|\mathbf{x}_0\|_2}\cdot \mathbf{x}_0\|_2^2 = \tau^2,$$

where

$$\mu = \mathbb{E}[Sf(S)], \qquad S \sim \mathcal{N}(0,1),$$
$$\tau^2 = \mathbb{E}[(f(S)-\mu S)^2]. \tag{19}$$

and the expectations are with respect to S and possible randomness of f. Evaluating (19) for $f(S) = \mathrm{BSC}_\varepsilon(\mathrm{sign}(S))$ leads to the same values for μ and τ^2 in (17). In other works, (17) 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* asymtpotic 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.

B. Least-absolute deviations

By choosing $\ell(t) = |t-1|$ in (4), we obtain a least-absolute deviations estimate. Again, since $y_i = \pm 1$, it holds for all i that $|y_i\mathbf{a}_i^T\mathbf{x} - 1| = |y_i - \mathbf{a}_i^T\mathbf{x}|$. Thus, this choice of loss function leads to residuals:

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \sum |y_i - \mathbf{a}_i^T\mathbf{x}|. \tag{20}$$

As in Section IV-A, for $\ell(t) = |t-1|$ the proximal operator admits a simple expression, as follows:

$$\mathrm{prox}_\ell\,(x;\lambda) = 1 + \mathcal{H}\,(x-1;\lambda) \tag{21}$$

where

$$\mathcal{H}\,(t;\lambda) = \begin{cases} t-\lambda, & \text{if } t>\lambda,\\ t+\lambda, & \text{if } t<-\lambda,\\ 0, & \text{otherwise.}\end{cases}$$

is the standard soft-thresholding function.

C. Hinge-loss

We obtain the hinge-loss estimator in by setting $\ell(t) = \max(1-t, 0)$ in (4). Similar to Section IV-B, the proximal operator of the hinge-loss can be expressed in terms of the soft-thresholding function as follows:

$$\mathrm{prox}_\ell\,(x;\lambda) = 1 + \mathcal{H}\left(x + \frac{\lambda}{2} - 1; \frac{\lambda}{2}\right).$$

As already mentioned in Remark 3, the set of minimizers of the hinge-loss is bounded (required by Theorem III.1) only for $\delta > \delta_\varepsilon^\star$ where $\delta_\varepsilon^\star$ is the value of the threshold in (13). Our numerical simulations in Figures 3 and 4 suggest that hinge-loss is robust to measurement corruptions, as for moderate to large values of δ it outperforms the LS and the LAD estimators. Theorem III.1 opens the way to analytically confirm such conclusions, which is an interesting future direction.

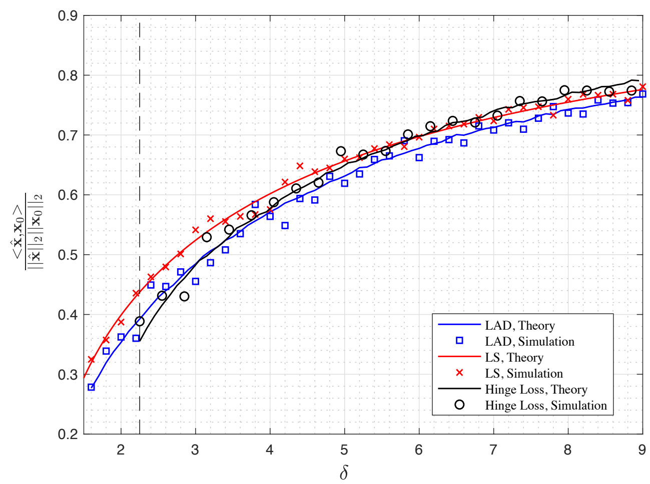

Fig. 4. Comparison between theoretical and simulated results for LAD, LS and Hinge-Loss estimators as a function of δ for probability of error $\epsilon = 0.25$. As in Figure 3, the dashed line represents the value of the threshold δ^*_ϵ for $\epsilon = 0.25$.

D. Numerical simulations

We present numerical simulations that validate the predictions of Theorem III.1. 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 $\mathbf{x}_0 = [1, 0, ..., 0]^T$. We then obtain estimates $\widehat{\mathbf{x}}_\ell$ of $\mathbf{x}_0$ by numerically solving (4). We measure performance by the correlation value $\text{corr}\left(\hat{\mathbf{x}}_\ell\,;\,\mathbf{x}_0\right)$. Throughout the experiments, we set $n = 128$ and the recorded values of correlation in Figures 2–4 are averages over 25 independent realizations. The theoretical curves for the correlation are computed based on Theorem III.1. We solve the system of equations in (7) by the fixed-point iteration method described in Remark 4. The expectations involved in (7) are evaluated with Monte-Carlo estimation using 10^5 independent samples.

Comparisons between theoretical and simulated values for LAD and LS estimators are presented in Figure 2 for the noiseless case. Note that for $\varepsilon = 0$, the hinge-loss has an unbounded set of minimizers for all values of δ (thus, Theorem III.1 is not applicable). In Figure 3, the probability of error ϵ 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 δ.
In Figure 4 we present similar results for $\epsilon = 0.25$. As it is evident from Figures 3 and 4, the best estimator is varying based on the value of δ and ε. 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 α/μ of the parameters α and μ that solve the system of equations in (7). For this purpose, it might also be important to derive necessary and sufficient conditions that guarantee (7) has a unique solution. Finally, it is possible to extend the results of this paper to (i) other measurement models (as it is indicated in the proof of Theorem III.1); (ii) structured signal recovery (e.g., sparse signals by including ℓ_1-regularization in (4)).

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] N Blachman. The convolution inequality for entropy powers. *IEEE Transactions on Information Theory*, 11(2):267–271, 1965.

[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:1804.09753*, 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.

[9] Oussama Dhifallah, Christos Thrampoulidis, and Yue M Lu. Phase retrieval via polytope optimization: Geometry, phase transitions, and new algorithms. *arXiv preprint arXiv:1805.09555*, 2018.

[10] David Donoho and Andrea Montanari. High dimensional robust m-estimation: Asymptotic variance via approximate message passing. *Probability Theory and Related Fields*, 166(3-4):935–969, 2016.

[11] David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

[12] David L Donoho, Arian Maleki, and Andrea Montanari. The noise-sensitivity phase transition in compressed sensing. *Information Theory, IEEE Transactions on*, 57(10):6920–6941, 2011.

[13] Noureddine El Karoui. On the impact of predictor geometry on the performance on high-dimensional ridge-regularized generalized robust regression estimators. 2015.

[14] Martin Genzel. High-dimensional estimation of structured signals from non-linear observations with general convex loss functions. *IEEE Transactions on Information Theory*, 63(3):1601–1619, 2017.

[15] Martin Genzel and Peter Jung. Recovering structured data from superimposed non-linear measurements. *arXiv preprint arXiv:1708.07451*, 2017.

[16] Larry Goldstein, Stanislav Minsker, and Xiaohan Wei. Structured signal recovery from non-linear and heavy-tailed measurements. *IEEE Transactions on Information Theory*, 64(8):5513–5530, 2018.

[17] Yehoram Gordon. *On Milman's inequality and random subspaces which escape through a mesh in* $\mathbb{R}^n$. Springer, 1988.

[18] Laurent Jacques, Jason N Laska, Petros T Boufounos, and Richard G Baraniuk. Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. *IEEE Transactions on Information Theory*, 59(4):2082–2102, 2013.

[19] Noureddine El Karoui. Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: rigorous results. *arXiv preprint arXiv:1311.2445*, 2013.

[20] Samet Oymak, Christos Thrampoulidis, and Babak Hassibi. The squared-error of generalized lasso: A precise analysis. *arXiv preprint arXiv:1311.0830*, 2013.

[21] Yaniv Plan and Roman Vershynin. Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach. *IEEE Transactions on Information Theory*, 59(1):482–494, 2012.

[22] Yaniv Plan and Roman Vershynin. One-bit compressed sensing by linear programming. *Communications on Pure and Applied Mathematics*, 66(8):1275–1297, 2013.

[23] Yaniv Plan and Roman Vershynin. The generalized lasso with non-linear observations. *IEEE Transactions on information theory*, 62(3):1528–1537, 2016.

[24] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[25] Fariborz Salehi, Ehsan Abbasi, and Babak Hassibi. The impact of regularization on high-dimensional logistic regression. *arXiv preprint arXiv:1906.03761*, 2019.

[26] Mihailo Stojnic. Various thresholds for ℓ_1-optimization in compressed sensing. *arXiv preprint arXiv:0907.3666*, 2009.

[27] Mihailo Stojnic. A framework to characterize performance of lasso algorithms. *arXiv preprint arXiv:1303.7291*, 2013.

[28] Pragya Sur and Emmanuel J Candès. A modern maximum-likelihood theory for high-dimensional logistic regression. *Proceedings of the National Academy of Sciences*, page 201810420, 2019.

[29] Christos Thrampoulidis, Ehsan Abbasi, and Babak Hassibi. Lasso with non-linear measurements is equivalent to one with linear measurements. In *Advances in Neural Information Processing Systems*, pages 3420–3428, 2015.

[30] Christos Thrampoulidis, Ehsan Abbasi, and Babak Hassibi. Precise error analysis of regularized m-estimators in high dimensions. *IEEE Transactions on Information Theory*, 64(8):5592–5628, 2018.

[31] Christos Thrampoulidis, Samet Oymak, and Babak Hassibi. Regularized linear regression: A precise analysis of the estimation error. In *Proceedings of The 28th Conference on Learning Theory*, pages 1683–1709, 2015.

[32] Christos Thrampoulidis and Ankit Singh Rawat. The generalized lasso for sub-gaussian measurements with dithered quantization. *arXiv preprint arXiv:1807.06976*, 2018.

[33] Martin Genzel. High-Dimensional Estimation of Structured Signals from Non-Linear Observations with General Convex Loss Functions. *arXiv preprint arXiv:1602.03436*, 2016.

[34] Derek Bean, Peter Bickel, Noureddine El Karoui and Bin Yu. Optimal M-estimation in high-dimensional regression. In *Proceedings of the National Academy of Sciences*,110(36):14563–14568, 2013.