Robust One-Bit Recovery via ReLU Generative Networks: Improved Statistical Rate and Global Landscape Analysis

Shuang Qiu*  
University of Michigan  
qiush@umich.edu

Xiaohan Wei*  
University of Southern California  
xiaohamw@usc.edu

Zhuoran Yang  
Princeton University  
zy6@princeton.edu

Abstract

We study the robust one-bit compressed sensing problem whose goal is to design an algorithm that faithfully recovers any sparse target vector $\theta_0 \in \mathbb{R}^d$ uniformly from $m$ quantized noisy measurements. Under the assumption that the measurements are sub-Gaussian, to recover any $k$-sparse $\theta_0$ ($k \ll d$) uniformly up to an error $\varepsilon$ with high probability, the best known computationally tractable algorithm requires $m \geq \tilde{O}(k \log d/\varepsilon^4)$. In this paper, we consider a new framework for the one-bit sensing problem where the sparsity is implicitly enforced via mapping a low dimensional representation $x_0$ through a known $n$-layer ReLU generative network $G : \mathbb{R}^k \to \mathbb{R}^d$. Such a framework poses low-dimensional priors on $\theta_0$ without a known basis. We propose to recover the target $G(x_0)$ via an unconstrained empirical risk minimization (ERM) problem under a much weaker sub-exponential measurement assumption. For such a problem, we establish a joint statistical and computational analysis. In particular, we prove that the ERM estimator in this new framework achieves an improved statistical rate of $m = \tilde{O}(kn \log d/\varepsilon^2)$ recovering any $G(x_0)$ uniformly up to an error $\varepsilon$. Moreover, from the lens of computation, despite non-convexity, we prove that the objective of our ERM problem has no spurious stationary point, that is, any stationary point is equally good for recovering the true target up to scaling with a certain accuracy. Our analysis sheds some light on the possibility of inverting a deep generative model under partial and quantized measurements, complementing the recent success of using deep generative models for inverse problems.

1 Introduction

Quantized compressed sensing investigates how to design the sensing procedure, quantizer and reconstruction algorithm so as to recover a high dimensional vector from a limited number of quantized measurements. The problem of one-bit compressed sensing, which aims at recovering a target vector $\theta_0 \in \mathbb{R}^d$ from single-bit observations $y_i = \text{sign}(\langle a_i, \theta_0 \rangle)$, $i \in \{1, 2, \cdots, m\}$, $m \ll d$ and random sensing vectors $a_i \in \mathbb{R}^d$, is particularly challenging. Previous theoretical successes on this problem (e.g. Jacques et al. (2013); Plan and Vershynin (2013)) mainly rely on two key assumptions: (1) The Gaussianity of the sensing vector $a_i$, (2) The sparsity of the vector $\theta_0$ on a given basis. However, the practical significance of these assumptions are rather limited in the sense that it is difficult to generate Gaussian vectors and high dimensional targets in practice are often distributed near a low-dimensional manifold rather than sparse on some given basis. The goal of this work is to make steps towards addressing these two limitations.

*Equal Contribution

1 Here, an algorithm is “computationally tractable” if it has provable convergence guarantees. The notation $\tilde{O}()$ omits a logarithm factor of $\varepsilon^{-1}$. 
Specifically, we introduce a new framework for robust dithered one-bit compressed sensing where the structure of target vector $b_0$ is represented via a ReLU network $G : \mathbb{R}^k \to \mathbb{R}^d$, i.e. $b_0 = G(x_0)$ for some $x_0 \in \mathbb{R}^k$ and $k \ll d$. Building upon this framework, we propose a new recovery algorithm by solving an unconstrained ERM. We show this algorithm enjoys the following favorable properties:

- Statistically, when taking measurements $a_i$ to be sub-exponential random vectors, with high probability and uniformly for any $G(x_0) \in G(\mathbb{R}^k) \cap \mathbb{B}^2(R)$, where $\mathbb{B}^2(R)$ is the ball of radius $R > 0$ centered at the origin, the solution $G(\tilde{x}_m)$ to the ERM recovers the true vector $G(x_0)$ up to error $\varepsilon$ when the number of samples $m \geq O(kn \log^4(1/\varepsilon) \log d + \log(1/\varepsilon))/\varepsilon^2)$. In particular, our result does not require REC type assumptions adopted in previous analysis of generative signal recovery works and at the same time weakens the known sub-Gaussian assumption adopted in previous one-bit compressed sensing works. When the number of layers $n$ is small, this result meets the minimax optimal rate (up to a logarithm factor) for sparse recovery and simultaneously improves upon the best known $O(k \log d / \varepsilon^4)$ statistical rate for computationally tractable algorithms.

- Computationally, we show that solving the ERM and approximate the true representation $x_0 \in \mathbb{R}^k$ is tractable. More specifically, we prove with high probability, there always exists a descent direction outside two small neighborhoods around $x_0$ and its negative multiple with radius $O(\varepsilon^{1/4})$, uniformly for any $x_0 \in \mathbb{B}^2(R'$) with $R' = (0.5 + \varepsilon)^{-n/2}R$, when the ReLU network satisfies a weight distribution condition with parameter $\varepsilon > 0$ and $m \geq O(kn \log^4(1/\varepsilon) \log d + \log(1/\varepsilon))/\varepsilon^2)$. Furthermore, when $\varepsilon$ is small enough, one guarantees that the solution $\tilde{x}_m$ stays within the neighborhood around $x_0$ (rather than its negative multiple). Our result is achieved without assuming the REC type conditions and under quantization errors, thereby improving upon previously known computational guarantees for ReLU generative signal recovery in linear models with small noise.

From a technical perspective, our proof makes use of the special piecewise linearity property of ReLU network. The merits of such a property in the current scenario are two folds: (1) It allows us to replaces the generic chaining type bounds commonly adopted in previous works (e.g. Dirksen and Mendelson (2018a)) by novel arguments that are “sub-Gaussian free”. (2) From a hyperplane tessellation point of view, we show that for a given accuracy level, a binary embedding of $G(\mathbb{R}^k) \cap \mathbb{B}^2(R)$ into Euclidean space is “easier” in that it requires less random hyperplanes than that of a bounded $k$ sparse set.

**Notations.** Throughout the paper, let $S^{d-1}$ and $B(x, r)$ denotes the unit sphere and the ball of radius $r$ centered at $x$ in $\mathbb{R}^d$, respectively. For a random variable $X \in \mathbb{R}$, the $L_p$-norm ($p \geq 1$) is denoted as $\|X\|_{L_p} := \mathbb{E}[|X|^p]^{1/p}$. The Olicz $\psi_1$-norm is denoted as $\|X\|_{\psi_1} := \sup_{p \geq 1} p^{-1} \|X\|_{L_p}$. We say a random variable is sub-exponential if its $\psi_1$-norm is bounded. A random vector $x \in \mathbb{R}^d$ is sub-exponential if there exists a a constant $C > 0$ such that $\sup_{t \in \mathbb{R}^{d-1}} \| (x, t) \|_{\psi_1} \leq C$. We use $\|x\|_{\psi_1}$ to denote the minimal $C$ such that this bound holds. Furthermore, $C, C', \varepsilon, c_1, c_2, c_3, c_4, c_5$ denote absolute constants, their actual values can be different per appearance.

## 2 Models and Main Results

In this paper, we focus on one-bit recovery model in which one observes quantized measurements of the following form

$$y = \text{sign}(\langle a, G(x_0) \rangle + \xi + \tau),$$

where $a \in \mathbb{R}^d$ is a random measurement vector, $\xi \in \mathbb{R}$ is a random pre-quantization noise with an unknown distribution, $\tau$ is a random quantization threshold (i.e. dithering noise), and $x_0 \in \mathbb{R}^k$ to recover. We are interested in the high-dimensional scenario where the dimension of the representation space $k$ is potentially much less than the ambient dimension $d$. The function $G : \mathbb{R}^k \to \mathbb{R}^d$ is a fixed ReLU neural network of the form:

$$G(x) = \sigma \circ (W_n \sigma \circ (W_{n-1} \cdots \sigma \circ (W_1 x))),$$

where $\sigma(x) = \max(x, 0)$ and $\sigma \circ (x)$ denotes the entry-wise application of $\sigma(\cdot)$. We consider a scenario where the number of layers $n$ is smaller than $d$, $W_i \in \mathbb{R}^{d_i \times d_{i+1}}$, $d_{n+1} = d$ and $d_i \leq d_i \forall i \in \{1, 2, \ldots, n\}$. Throughout the paper, we assume that $G(x_0)$ is bounded, i.e. there exists an $R \geq 1$ such that $\|G(x_0)\|_2 \leq R$, and we take $\tau \sim \text{Unif}[\lambda - \lambda, \lambda + \lambda]$, i.e. a uniform distribution bounded by a chosen parameter $\lambda > 0$. Let $((a_i, y_i))_{i=1}^m$ be i.i.d. copies of $(a, y)$. Our goal is to compute an estimator $G(\tilde{x}_m)$ of $G(x_0)$ such that $\|G(\tilde{x}_m) - G(x_0)\|_2$ is small.
We propose to solve the following ERM for $\hat{x}_m$:

$$
\hat{x}_m \in \arg\min_{x \in \mathbb{R}^k} L(x) := \|G(x)\|_2^2 = \frac{2\lambda}{m} \sum_{i=1}^{m} y_i \langle a_i, G(x) \rangle,
$$

(3)

where $y_i = \text{sign}(\langle a_i, G(x_0) \rangle + \xi_i + \tau_i)$. It is worth mentioning, in general, there is no guarantee that the minimizer of $L(x)$ is unique. Nevertheless, in Section §2.2, we will show that any solution $\hat{x}_m$ to this problem must stay inside small neighborhoods around the true signal $x_0$ and its negative multiple with high probability.

### 2.1 Statistical Guarantee

Our statistical guarantee relies on the following assumption on the measurement vector and noise:

**Assumption 2.1.** The measurement vector $a \in \mathbb{R}^d$ is mean 0, isotropic and sub-exponential. The noise $\xi$ is also a sub-exponential random variable.

Under this assumption, we have the following main statistical performance theorem:

**Theorem 2.1.** Suppose Assumption 2.1 holds and consider any $\varepsilon \in (0, 1)$. Set the constant $C_{a,\xi,R} = \max\{c_1(R)\|\varphi_i\|_1, 1\}$, $\lambda \geq 4C_{a,\xi,R} \cdot \log(64C_{\varphi,\xi,R} \cdot \varepsilon^{-1})$ and

$$
m \geq c_2\|a\|_2^2 \lambda^2 \log^2(\lambda m)(k_1 \log(\varepsilon d) + k \log(2R) + k \log(m + u))/\varepsilon^2.
$$

(4)

Then, with probability at least $1 - c_3 \exp(-u)$, $\forall u \geq 0$, any solution $\hat{x}_m$ to (3) satisfies

$$
\|G(\hat{x}_m) - G(x_0)\|_2 \leq \varepsilon
$$

for all $x_0$ such that $\|G(x_0)\|_2 \leq R$, where $c_1$, $c_2$, $c_3 \geq 1$ are absolute constants.

**Remark 2.1.** It is easy to verify that the sample complexity enforced by (4) holds when

$$
m \geq C' \cdot C_{a,\xi,R}^2 \log^4(C_{a,\xi,R} \cdot \varepsilon^{-1})(k_1 \log(\varepsilon d) + k \log(2R) + k \log(\varepsilon^{-1}) + u)/\varepsilon^2
$$

where $C'$ is a large enough absolute constant. This gives the $m = \mathcal{O}(k_1 \log^4(\varepsilon^{-1})(\log d + \log(\varepsilon^{-1}))/\varepsilon^2)$ statistical rate. The question whether or not the dependency on $n$ is redundant (comparing to that of sparse recovery guarantees) warrants further studies.

**Remark 2.2.** Note that our result is a uniform recovery result in the sense that the bound $\|G(\hat{x}_m) - G(x_0)\|_2 \leq \varepsilon$ holds with high probability uniformly for any target $x_0 \in \mathbb{R}^k$ such that $\|G(x_0)\|_2 \leq R$. This should be distinguished from known bounds (Plan and Vershynin (2013); Zhang et al. (2014); Goldstein et al. (2018); Thrampoulidis and Rawat (2018)) on sparse one-bit sensing which hold only for a fixed sparse vector. Furthermore, though assuming boundedness of $G(x_0)$, our recovery algorithm solves for the minimizer without knowing this bound, which is favorable for practice.

### 2.2 Global Landscape of the Proposed ERM

Before presenting the results on the global landscape, we first introduce some notations used in the rest of this paper. For any fixed $x$, we define $W_{+,x} := \text{diag}(Wx > 0)W$, in which we set the rows of $W$ having negative product with $x$ to be zeros. We further define $W_{i,+,x} := \text{diag}(W_iW_{i-1,+,x} \cdots W_{1,+,x} > 0)W_i$, where only active rows of $W_i$ are kept. Thus, we can represent the RuLU network by $G(x) = (\Pi_{i=1}^{n} W_{i,+,x})x := W_{n,+,x}W_{n-1,+,x} \cdots W_{1,+,x}x$.

**Definition 2.1** (Weighted Distribution Condition (WDC) (Hand and Voroninski, 2018; Hand et al., 2018)). The matrix $W \in \mathbb{R}^{d \times k'}$ satisfies the Weighted Distribution Condition with constant $\varepsilon_{\text{wdc}}$ if for any nonzero vectors $x_1, x_2 \in \mathbb{R}^{k'}$,

$$
\|W_{+,x_1}^TW_{+,x_2} - Q_{x_1,z}\|_2 \leq \varepsilon_{\text{wdc}} \text{, with } Q_{x,z} := \frac{\pi - \theta_{x,z}}{2\pi} I_{k'} + \frac{\sin \theta_{x,z}}{2\pi} M_{x+\hat{z}}.
$$

where we have $\theta_{x,z} = \angle(x, z)$ and $M_{x+\hat{z}}$ is the matrix that transforms $\hat{x}$ to $\hat{z}$, $\hat{z}$ to $\hat{x}$, and $\hat{0}$ to 0 for any $\hat{\theta} \in \text{span}\{\{x, z\}\}$. Here we define $\hat{x} := x/\|x\|_2$, $\hat{z} := z/\|z\|_2$. 3
Before presenting Theorem 2.2 and 2.3, we define the directional derivative along non-zero $z$ as
\[
D_z L(x) = \begin{cases} 
\langle \nabla L(x), z/\|z\|_2 \rangle, & \text{if } L(x) \text{ is differentiable at } x \\
\lim_{N \to +\infty} \langle \nabla L(x_N), z/\|z\|_2 \rangle, & \text{otherwise}
\end{cases}
\]
where \{x_N\} is a sequence such that $x_N \to x$ and $L(x)$ is differentiable at any $x_N$. Each sequence must exist due to the piecewise linearity of $G(x)$. For any $x$ such that $L(x)$ is differentiable, the gradient of $L(x)$ is can be easily computed as
\[
\nabla L(x) = 2 \left( \prod_{j=1}^{n} W_{j,+} x \right)^\top \left( \prod_{j=1}^{n} W_{j,+} x - \frac{2\lambda}{m} \sum_{i=1}^{m} y_i \left( \prod_{j=1}^{n} W_{j,+} x \right)^\top a_i \right) \tag{5}
\]

Next, we will present Theorem 2.2 to show that under certain conditions, local minimum can only lie in small neighborhoods of two points $x_0$ and its negative multiple $-\rho_n x_0$.

**Theorem 2.2.** Suppose that $G$ is a ReLU network with $W_i$ satisfying WDC with error $\varepsilon_{wdc}$ for all $i = 1, \ldots, n$ where $n > 1$. With probability $1 - c_1 \exp(-u)$, for any nonzero $x_0$ satisfying
\[
\|x_0\|_2 \leq R(1/2 + \varepsilon_{wdc})^{-n/2},
\]
if we set $88\pi n^4 \varepsilon_{wdc} \leq 1$, $\lambda \geq 4C_{a,\xi,R} \log(64C_{\psi,\xi,R} \cdot \varepsilon_{wdc})$, $m \geq c_2\|a\|_2^2 \lambda^2 \log^2(\lambda m)(kn \log(ed) + k \log(2R) + k \log m + u)/\varepsilon_{wdc}$, for any nonzero $x$, set $v_x = \lim_{x \to x} \nabla L(x_N)$ where \{x_N\} is the sequence such that $\nabla L(x_N)$ exists for all $x_N$ (and $v_x = \nabla L(x)$ if $L(x)$ is differentiable at $x$), then, there exists a constant $\rho_n \leq 1$ such that the directional derivative satisfies:

**Case 1:** if $\|x_0\|_2^2 > 2^n \varepsilon_{wdc}$, then
\[
D_{-v_x} L(x) < 0, \forall x \notin B(x_0, 616n^3 \varepsilon_{wdc} \|x_0\|_2) \cup B(-\rho_n x_0, 5500n^{14} \varepsilon_{wdc} \|x_0\|_2) \cup \{0\},
\]
\[
D_w L(0) < 0, \forall w \neq 0
\]

**Case 2:** if $\|x_0\|_2^2 \leq 2^n \varepsilon_{wdc}$, then
\[
D_{-v_x} L(x) < 0, \forall x \notin B(x_0, 616n^3 \varepsilon_{wdc} \|x_0\|_2) \cup B(-\rho_n x_0, 5500n^{14} \varepsilon_{wdc} \|x_0\|_2) \cup B(0, 2^n / 2^{1/2} \varepsilon_{wdc}).
\]

Note that in the above theorem, case 1 indicates that the when the magnitude of the true representation $\|x_0\|_2^2$ is larger than the accuracy level $\varepsilon_{wdc}$, the global minimum lies in small neighborhoods around $x_0$ and its scalar multiple $-\rho_n x_0$, while for any point outside the neighborhoods of $x_0$ and $-\rho_n x_0$, one can always find a direction with a negative directional derivative. Note that $x = 0$ is a local maximum due to $D_w L(0) < 0$ along any non-zero directions $w$. One the other hand, case 2 implies that when $\|x_0\|_2^2$ is smaller than $\varepsilon_{wdc}$, the global minimum lies in the neighborhood around 0 (and thus around $x_0$). Moreover, in the following theorem, we further pin down the global minimum of the loss function for case 1.

**Theorem 2.3.** Suppose that $G$ is a ReLU network with $W_i$ satisfying WDC with error $\varepsilon_{wdc}$ for all $i = 1, \ldots, n$ where $n > 1$. Assume that $c_1 n^{1/4} \varepsilon_{wdc} \leq 1$, and $x_0$ is any nonzero vector satisfying $\|x_0\|_2 \leq R(1/2 + \varepsilon_{wdc})^{-n/2}$. Then, with probability $1 - 2c_4 \exp(-u)$, for any $x_0$ such that $\|x_0\|_2^2 \geq 2^n \varepsilon_{wdc}$, setting $\lambda \geq 4C_{a,\xi,R} \log(64C_{\psi,\xi,R} \cdot \varepsilon_{wdc})$, and $m \geq c_2\|a\|_2^2 \lambda^2 \log^2(\lambda m)(kn \log(ed) + k \log(2R) + k \log m + u)/\varepsilon_{wdc}$, we have
\[
L(x) < L(z), \forall z \in B(\phi_N x_0, c_3 n^{-5} \|x_0\|_2) \text{ and } \forall z \in B(-\zeta_N x_0, c_2 n^{-5} \|x_0\|_2),
\]
where $c_1, c_2, c_3, c_4$ are absolute constants, $\phi_N, \zeta_N$ are any scalars in $[\rho_n, 1]$. Particularly, we have $c_3 n^{-5} < \min_{n > 2} \rho_n$ such that the radius $c_3 n^{-5} \|x_0\|_2 < \rho_n \|x_0\|_2$ for any $n$.

**Remark 2.3.** The significance of Theorem 2.3 are two folds: first, it shows that the value of ERM is always smaller around $x_0$ compared to its negative multiple $-\rho_n x_0$; second, when the accuracy level $\varepsilon_{wdc}$ is small, one can guarantee that the global minimum of $L(x)$ stays around $x_0$. In particular, by Theorem 2.2 and 2.3, our theory implies that if $\varepsilon_{wdc} \leq cn^{-76}$ for some constant $c$, then the global minimum of the proposed ERM (3) is in $B(\phi_N x_0, c_3 n^{-5} \|x_0\|_2)$. Since we do not focus on optimizing the order of $n$ here, further improvement of such a dependency will be one of our future works.
References


