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
xiaohanw@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 θ_0 ($k \ll d$) *uniformly* up to an error ε with high probability, the best known computationally tractable algorithm requires[1] $m \geq \tilde{\mathcal{O}}(k \log d / \varepsilon^4)$. 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 θ_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 $m = \tilde{\mathcal{O}}(kn \log d / \epsilon^2)$ recovering any $G(x_0)$ uniformly up to an error ε. Moreover, from the lens of computation, we prove that under proper conditions on the ReLU weights, our proposed empirical risk, despite non-convexity, has no stationary point outside of small neighborhoods around the true representation x_0 and its negative multiple. Furthermore, we show that the global minimizer of the empirical risk stays within the neighborhood around x_0 rather than its negative multiple. 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 θ_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

---

[*] Equal Contribution

[1] Here, an algorithm is "computationally tractable" if it has provable convergence guarantees. The notation $\tilde{\mathcal{O}}(\cdot)$ omits a logarithm factor of ε^{-1}.

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.

Specifically, we introduce a new framework for robust dithered one-bit compressed sensing where the structure of target vector θ_0 is represented via a ReLU network $G : \mathbb{R}^k \to \mathbb{R}^d$, i.e. $\theta_0 = G(x_0)$ 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 $G(x_0) \in G(\mathbb{R}^k) \cap \mathbb{B}_2^d(R)$, where $\mathbb{B}_2^d(R)$ is the ball of radius $R > 0$ centered at the origin, the solution $G(\widehat{x}_m)$ to the ERM recovers the true vector $G(x_0)$ up to error ε when the number of samples $m \geq \mathcal{O}(kn \log^4(\varepsilon^{-1})(\log d + \log(\varepsilon^{-1}))/\varepsilon^2)$. 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 $\tilde{\mathcal{O}}(k \log d/\varepsilon^4)$ 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 $\mathcal{O}(\varepsilon^{1/4})$, uniformly for any $x_0 \in \mathbb{B}_2^k(R')$ with $R' = (0.5 + \varepsilon)^{-n/2} R$, when the ReLU network satisfies a weight distribution condition with parameter $\varepsilon > 0$ and $m \geq \mathcal{O}(kn \log^4(\varepsilon^{-1})(\log d + \log(\varepsilon^{-1}))/\varepsilon^2)$. Furthermore, when ε is small enough, one guarantees that the solution $\widehat{x}_m$ 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 $G(\mathbb{R}^k) \cap \mathbb{B}_2^d(R)$ 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 $\mathcal{S}^{d-1}$ and $\mathcal{B}(x, r)$ 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 ψ_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 ψ_1-norm is bounded. A random vector $\mathbf{x} \in \mathbb{R}^d$ is sub-exponential if there exists a a constant $C > 0$ such that $\sup_{t \in \mathcal{S}^{d-1}} \|\langle \mathbf{x}, t \rangle\|_{\psi_1} \leq C$. We use $\|\mathbf{x}\|_{\psi_1}$ to denote the minimal C such that this bound holds. Furthermore, $C, C', c, c_1, c_2, c_3, c_4, c_5$ 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

$$y = \text{sign}(\langle a, G(x_0) \rangle + \xi + \tau), \tag{1}$$

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, τ is a random quantization threshold (i.e. dithering noise), and $x_0 \in \mathbb{R}^k$ is the unknown representation to be recovered. We are interested 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:

$$G(x) = \sigma \circ (W_n \sigma \circ (W_{n-1} \cdots \sigma \circ (W_1 x))), \tag{2}$$

where $\sigma(x) = \max(x, 0)$ and $\sigma \circ (\mathbf{x})$ 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 $d_i \leq d$, $\forall i \in \{1, 2, \cdots, n\}$. 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 $\tau \sim \text{Uni}[-\lambda, +\lambda]$, 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 $G(\hat{x}_m)$ of $G(x_0)$ such that $\|G(\hat{x}_m) - G(x_0)\|_2$ is small.

We propose to solve the following ERM for $\hat{x}_m$:

$$\hat{x}_m \in \arg\min_{x \in \mathbb{R}^k} L(x) := \|G(x)\|_2^2 - \frac{2\lambda}{m} \sum_{i=1}^m y_i \langle a_i, G(x) \rangle, \tag{3}$$

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 ξ 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 $C_{a,\xi,R} = \max\{c_1(R\|a\|_{\psi_1} + \|\xi\|_{\psi_1}), 1\}$, $\lambda \geq 4C_{a,\xi,R} \cdot \log(64C_{\psi,\xi,R} \cdot \varepsilon^{-1})$ and*

$$m \geq c_2\|a\|_{\psi_1}^2 \lambda^2 \log^2(\lambda m)(kn \log(ed) + k \log(2R) + k \log m + u)/\varepsilon^2. \tag{4}$$

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*

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

where C' is a large enough absolute constant. This gives the $m = \mathcal{O}(kn \log^4(\varepsilon^{-1})(\log d + \log(\varepsilon^{-1}))/\varepsilon^2)$ 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 $W_{i,+,x} := \text{diag}(W_i W_{i-1,+,x} \cdots W_{1,+,x} x > 0)W_i$, 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 ε_{wdc} if for any nonzero vectors $x_1, x_2 \in \mathbb{R}^{k'}$,*

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

where we have $\theta_{x,z} = \angle(x, z)$ and $M_{\hat{x} \leftrightarrow \hat{z}}$ is the matrix that transforms $\hat{x}$ to $\hat{z}$, $\hat{z}$ to $\hat{x}$, and ϑ to 0 for any $\vartheta \in \text{span}(\{x, z\})^\perp$. 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. Such 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

$$\nabla L(x) = 2\Big(\prod_{j=1}^{n} W_{j,+,x}\Big)^{\top}\Big(\prod_{j=1}^{n} W_{j,+,x}\Big)x - \frac{2\lambda}{m}\sum_{i=1}^{m} y_i \Big(\prod_{j=1}^{n} W_{j,+,x}\Big)^{\top} a_i. \tag{5}$$

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 ε_{wdc} for all $i = 1, \ldots, n$ where $n > 1$. With probability $1 - c_1 \exp(-u)$, for any nonzero x_0 satisfying $\|x_0\|_2 \leq R(1/2 + \varepsilon_{\text{wdc}})^{-n/2}$, if we set $88\pi n^8 \varepsilon_{\text{wdc}}^{1/4} \leq 1$, $\lambda \geq 4C_{a,\xi,R} \cdot \log(64C_{\psi,\xi,R} \cdot \varepsilon_{\text{wdc}}^{-1})$,*

$$m \geq c_2 \|a\|_{\psi_1}^2 \lambda^2 \log^2(\lambda m)(kn\log(ed) + k\log(2R) + k\log m + u)/\varepsilon_{\text{wdc}}^2,$$

for any nonzero x, set $v_x = \lim_{x_N \to x} \nabla L(x_N)$ 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 $\|x_0\|_2^2 > 2^n \varepsilon_{\text{wdc}}$, then*

$$D_{-v_x} L(x) < 0, \forall x \notin \mathcal{B}(x_0, 616n^3 \varepsilon_{\text{wdc}}^{1/4} \|x_0\|_2) \cup \mathcal{B}(-\rho_n x_0, 5500n^{14} \varepsilon_{\text{wdc}}^{1/4} \|x_0\|_2) \cup \{0\},$$
$$D_w L(0) < 0, \forall w \neq 0$$

Case 2: *if $\|x_0\|_2^2 \leq 2^n \varepsilon_{\text{wdc}}$, then*

$$D_{-v_x} L(x) < 0,$$
$$\forall x \notin \mathcal{B}(x_0, 616n^3 \varepsilon_{\text{wdc}}^{1/4} \|x_0\|_2) \cup \mathcal{B}(-\rho_n x_0, 5500n^{14} \varepsilon_{\text{wdc}}^{1/4} \|x_0\|_2) \cup \mathcal{B}(0, 2^{n/2} \varepsilon_{\text{wdc}}^{1/2}).$$

where $\rho_n = \sum_{i=0}^{n-1} \frac{\sin \check{\theta}_i}{\pi} \Big(\prod_{j=i+1}^{n-1} \frac{\pi - \check{\theta}_j}{\pi}\Big)$, $g(\theta) = \cos^{-1}\Big(\frac{(\pi - \theta)\cos\theta + \sin\theta}{\pi}\Big)$, $\check{\theta}_0 = \pi$, and $\check{\theta}_i = g(\check{\theta}_{i-1})$.

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 ε_{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 ε_{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 ε_{wdc} for all $i = 1, \ldots, n$ where $n > 1$. Assume that $c_1 n^3 \varepsilon_{\text{wdc}}^{1/4} \leq 1$, and x_0 is any nonzero vector satisfying $\|x_0\|_2 \leq R(1/2 + \varepsilon_{\text{wdc}})^{-n/2}$. Then, with probability $1 - 2c_4 \exp(-u)$, for any x_0 such that $\|x_0\|_2^2 \geq 2^n \varepsilon_{\text{wdc}}$, setting $\lambda \geq 4C_{a,\xi,R} \cdot \log(64C_{\psi,\xi,R} \cdot \varepsilon_{\text{wdc}}^{-1})$, and*

$$m \geq c_2 \|a\|_{\psi_1}^2 \lambda^2 \log^2(\lambda m)(kn\log(ed) + k\log(2R) + k\log m + u)/\varepsilon_{\text{wdc}}^2,$$

we have

$$L(x) < L(z), \forall x \in \mathcal{B}(\phi_n x_0, c_3 n^{-5} \|x_0\|_2), \text{ and } \forall z \in \mathcal{B}(-\zeta_n x_0, c_3 n^{-5} \|x_0\|_2),$$

where c_1, c_2, c_3, c_4 are absolute constants, ϕ_n, ζ_n are any scalars in $[\rho_n, 1]$. Particularly, we have $c_3 n^{-5} < \min_{n \geq 2} \rho_n$ 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 ε_{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 $\varepsilon_{\text{wdc}} \leq cn^{-76}$ for some constant c, then the global minimum of the proposed ERM (3) is in $\mathcal{B}(\phi_n x_0, c_3 n^{-5} \|x_0\|_2)$. 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

Ai, A., A. Lapanowski, Y. Plan, and R. Vershynin (2014). One-bit compressed sensing with non-gaussian measurements. *Linear Algebra and its Applications 441*, 222–239.

Angluin, D. and L. G. Valiant (1979). Fast probabilistic algorithms for Hamiltonian circuits and matchings. *Journal of Computer and system Sciences 18*(2), 155–193.

Arjovsky, M., S. Chintala, and L. Bottou (2017). Wasserstein gan. *arXiv preprint arXiv:1701.07875*.

Bora, A., A. Jalal, E. Price, and A. G. Dimakis (2017). Compressed sensing using generative models. *arXiv preprint arXiv:1703.03208*.

Dirksen, S. and S. Mendelson (2018a). Robust one-bit compressed sensing with non-gaussian measurements. *arXiv preprint arXiv:1805.09409*.

Dirksen, S. and S. Mendelson (2018b). Robust one-bit compressed sensing with partial circulant matrices. *arXiv preprint arXiv:1812.06719*.

Goldstein, L., S. Minsker, and X. Wei (2018). Structured signal recovery from non-linear and heavy-tailed measurements. *IEEE Transactions on Information Theory 64*(8), 5513–5530.

Goldstein, L. and X. Wei (2018, 05). Non-Gaussian observations in nonlinear compressed sensing via Stein discrepancies. *Information and Inference: A Journal of the IMA 8*(1), 125–159.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio (2014). Generative adversarial nets. In *Advances in neural information processing systems*, pp. 2672–2680.

Hammernik, K., T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll (2018). Learning a variational network for reconstruction of accelerated mri data. *Magnetic resonance in medicine 79*(6), 3055–3071.

Hand, P., O. Leong, and V. Voroninski (2018). Phase retrieval under a generative prior. In *Advances in Neural Information Processing Systems*, pp. 9136–9146.

Hand, P. and V. Voroninski (2018). Global guarantees for enforcing deep generative priors by empirical risk. In *Conference On Learning Theory*, pp. 970–978.

Huang, W., P. Hand, R. Heckel, and V. Voroninski (2018). A provably convergent scheme for compressive sensing under random generative priors. *arXiv preprint arXiv:1812.04176*.

Jacques, L., J. N. Laska, P. T. Boufounos, and R. G. Baraniuk (2013). Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. *IEEE Transactions on Information Theory 59*(4), 2082–2102.

Kingma, D. P. and M. Welling (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Ledig, C., L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 105–114. IEEE.

Liu, Z. and J. Scarlett (2019). Information-theoretic lower bounds for compressive sensing with generative models. *arXiv preprint arXiv:1908.10744*.

Plan, Y. and R. Vershynin (2013). Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach. *IEEE Transactions on Information Theory 59*(1), 482–494.

Plan, Y. and R. Vershynin (2014). Dimension reduction by random hyperplane tessellations. *Discrete & Computational Geometry 51*(2), 438–461.

Plan, Y., R. Vershynin, and E. Yudovina (2016). High-dimensional estimation with geometric constraints. *Information and Inference: A Journal of the IMA 6*(1), 1–40.

Rezende, D. J., S. Mohamed, and D. Wierstra (2014). Stochastic backpropagation and approximate inference in deep generative models. *arXiv preprint arXiv:1401.4082*.

Sønderby, C. K., J. Caballero, L. Theis, W. Shi, and F. Huszár (2016). Amortised map inference for image super-resolution. *arXiv preprint arXiv:1610.04490*.

Thrampoulidis, C. and A. S. Rawat (2018). The generalized lasso for sub-gaussian measurements with dithered quantization. *arXiv preprint arXiv:1807.06976*.

Wellner, J. et al. (2013). *Weak convergence and empirical processes: with applications to statistics*. Springer Science & Business Media.

Winder, R. (1966). Partitions of n-space by hyperplanes. *SIAM Journal on Applied Mathematics 14*(4), 811–818.

Xu, C. and L. Jacques (2018). Quantized compressive sensing with rip matrices: The benefit of dithering. *arXiv preprint arXiv:1801.05870*.

Yang, G., S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, et al. (2018). Dagan: Deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction. *IEEE transactions on medical imaging 37*(6), 1310–1321.

Yeh, R. A., C. Chen, T. Yian Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do (2017). Semantic image inpainting with deep generative models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5485–5493.

Zhang, L., J. Yi, and R. Jin (2014). Efficient algorithms for robust one-bit compressive sensing. In *International Conference on Machine Learning*, pp. 820–828.