Sample Complexity Lower Bounds for Compressive Sensing with Generative Models

Zhaoqiang Liu
National University of Singapore
dcslizha@nus.edu.sg

Jonathan Scarlett
National University of Singapore
scarlett@comp.nus.edu.sg

Abstract

The goal of standard compressive sensing is to estimate an unknown vector from linear measurements under the assumption of sparsity in some basis. Recently, it has been shown that significantly fewer measurements may be required if the sparsity assumption is replaced by the assumption that the unknown vector lies near the range of a suitably-chosen generative model. In particular, in (Bora *et al.*, 2017) it was shown that roughly $O(k \log L)$ random Gaussian measurements suffice for accurate recovery when the k-input generative model is bounded and L-Lipschitz, and that $O(kd \log w)$ measurements suffice for k-input ReLU networks with depth d and width w. In this paper, we establish corresponding algorithm-independent lower bounds on the sample complexity using tools from minimax statistical analysis. In accordance with the above upper bounds, our results are summarized as follows: (i) We construct an L-Lipschitz generative model capable of generating group-sparse signals, and show that the resulting necessary number of measurements is $\Omega(k \log L)$; (ii) Using similar ideas, we construct two-layer ReLU networks of high width requiring $\Omega(k \log w)$ measurements, as well as lower-width deep ReLU networks requiring $\Omega(kd)$ measurements. As a result, we establish that the scaling laws derived in (Bora *et al.*, 2017) are optimal or near-optimal in the absence of further assumptions.

1 Introduction

The problem of sparse estimation via linear measurements (commonly referred to as *compressive sensing*) is well-understood, with theoretical developments including sharp performance bounds for both practical algorithms [1, 2, 3, 4] and (potentially intractable) information-theoretically optimal algorithms [5, 6, 7, 8]. Following the tremendous success of deep generative models in a variety of applications [9], a new perspective on compressive sensing was recently introduced, in which the sparsity assumption is replaced by the assumption of the underlying signal being well-modeled by a generative model (typically corresponding to a deep neural network) [10]. This approach was seen to exhibit impressive performance in experiments, with reductions in the number of measurements by large factors such as 5 to 10 compared to sparsity-based methods.

In addition, [10] provided theoretical guarantees on their proposed algorithm, essentially showing that an L-Lipschitz generative model with bounded k-dimensional inputs leads to reliable recovery with $m = O(k \log L)$ random Gaussian measurements (see Section 2 for a precise statement). Moreover, for a ReLU network generative model from $\mathbb{R}^k$ to $\mathbb{R}^n$ with width w and depth d, it suffices to have $m = O(kd \log w)$. A variety of follow-up works provided additional theoretical guarantees (e.g., for more specific optimization algorithms [11, 12], more general models [13], or under random neural network weights [14, 15]) for compressive sensing with generative models, but the main results of [10] are by far the most relevant to ours.

In this paper, we address a prominent gap in the existing literature by establishing *algorithm-independent lower bounds* on the number of measurements needed (e.g., this is explicitly posed as an

open problem in [15]). Using tools from minimax statistical analysis, we show that for generative models satisfying the assumptions of [10], the above-mentioned dependencies $m = O(k \log L)$ and $m = O(kd \log w)$ cannot be improved (or in the latter case, cannot be improved by more than a $\log n$ factor) without further assumptions. Our argument is essentially based on a reduction to compressive sensing with a *group sparsity* model (e.g., see [16]), i.e., forming a neural network that is capable of producing such signals. The proofs are presented in the full paper [17].

2 Bounded Lipschitz-Continuous Generative Models

We begin by stating a simple corollary of a main result of Bora *et al.* [10]. As we show in [17], this is obtained by extending [10, Thm. 1.2] from spherical to rectangular domains, and then converting the high-probability bound to an average one. In [17], we also handle the case of spherical domains.

Corollary 1. (Adapted from [10, Thm. 1.2]) *Fix $r > 0$, and let $G : B_\infty^k(r) \to \mathbb{R}^n$ be an L-Lipschitz function, and let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a random measurement matrix whose entries are i.i.d. with distribution $A_{ij} \sim \mathcal{N}(0, \frac{1}{m})$. Given the observed vector $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \boldsymbol{\eta}$ with i.i.d. Gaussian noise $\boldsymbol{\eta} \sim \mathcal{N}(0, \frac{\alpha}{m}\mathbf{I}_m)$ for some $\alpha > 0$, let $\hat{\mathbf{z}}$ be the global minimizer of $\|\mathbf{y} - \mathbf{A}G(\mathbf{z})\|_2$ over $\mathbf{z} \in B_\infty^k(r)$. Then, when $m \geq Ck \log \frac{Lr\sqrt{k}}{\sqrt{\alpha}}$ for a universal constant C, we have for a universal constant C' that*

$$\sup_{\mathbf{x}^* \in \mathrm{Range}(G)} \mathbb{E}\big[\|G(\hat{\mathbf{z}}) - \mathbf{x}^*\|_2^2\big] \leq C'\alpha. \tag{1}$$

In the following, we construct a Lipschitz-continuous generative model that can generate bounded k-group-sparse vectors. Then, by making use of minimax statistical analysis for group-sparse recovery, we provide an information-theoretic lower bounds that matches the upper bound in Corollary 1.

More precisely, we say that a signal in $\mathbb{R}^n$ is *k-group-sparse* if, when divided into k blocks of size $\frac{n}{k}$, each block contains at most one non-zero entry. We define

$$\mathcal{S}_k = \big\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \text{ is } k\text{-group-sparse}\big\}, \tag{2}$$

$$\mathcal{S}_k(x_{\max}) = \big\{\mathbf{x} \in \mathcal{S}_k \,:\, \|\mathbf{x}\|_\infty \leq x_{\max}\big\}. \tag{3}$$

Our construction of the generative function $G : B_\infty^k(r) \to \mathcal{S}_k(x_{\max})$ is given as follows:

- The output $\mathbf{x} \in \mathbb{R}^n$ is divided into k sub-sequences of length $\frac{n}{k}$, denoted $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)} \in \mathbb{R}^{\frac{n}{k}}$.
- A given block $\mathbf{x}^{(i)}$ is only a function of the corresponding input z_i, for $i = 1, \ldots, k$.
- The mapping from z_i to $\mathbf{x}^{(i)}$ is as shown in Figure 1. The interval $[-r, r]$ is divided into $\frac{n}{k}$ intervals of length $\frac{2rk}{n}$, and the j-th entry of $\mathbf{x}^{(i)}$ can only be non-zero if z_i takes a value in the j-th interval. Within that interval, the mapping takes a "double-triangular" shape – the endpoints and midpoint are mapped to zero, the points $\frac{1}{4}$ and $\frac{3}{4}$ of the way into the interval are mapped to $x_{\max}$ and $-x_{\max}$ respectively, and the remaining points are formed via a linear interpolation.

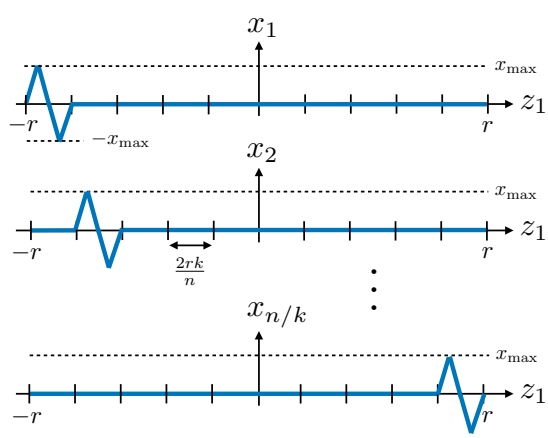

Figure 1: Generative model that produces sparse signals. This figure shows the mapping from $z_1 \to (x_1, \ldots, x_{n/k})$, and the same relation holds for $z_2 \to (x_{n/k+1}, \ldots, x_{2n/k})$, etc. up to $z_k \to (x_{n-k+1}, \ldots, x_n)$.

It is easy to show that the generative model $G : B_\infty^k(r) \to \mathcal{S}_k(x_{\max})$ described above is L Lipschitz continuous with $L = \frac{2nx_{\max}}{kr}$. In addition, using similar steps to the case of k-sparse recovery [7, 18], we are able to obtain a minimax lower bound for k-group-sparse recovery, which holds when $x_{\max}$ is not too small. Based on these results, the following sample complexity lower bound is proved in [17].

Theorem 1. *Consider the problem of compressive sensing with L-Lipschitz generative models with input domain $B_\infty^k(r)$, and i.i.d. $N\big(0, \frac{\alpha}{m}\big)$ noise. Let $C_1 > 0$ and $C_{\mathbf{A}} > 0$ be fixed constants, and assume that $L \geq \Omega\big(\frac{1}{r}\sqrt{\frac{\alpha}{k}}\big)$ with a sufficiently large implied constant. Then there exists an L-Lipschitz generative model $G : B_\infty^k(r) \to \mathbb{R}^n$ (and associated output dimension n) such that, for any $\mathbf{A} \in \mathbb{R}^{m\times n}$ satisfying $\|\mathbf{A}\|_{\mathrm{F}}^2 = C_{\mathbf{A}} n$, any algorithm that produces some $\hat{\mathbf{x}}$ satisfying*

$$\sup_{\mathbf{x}^* \in \mathrm{Range}(G)} \mathbb{E}\big[\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2\big] \leq C_1 \alpha \tag{4}$$

must also have $m = \Omega\big(k \log \frac{Lr\sqrt{k}}{\sqrt{\alpha}}\big)$.

Theorem 1 not only shows that the scaling laws of Corollary 1 cannot be improved under i.i.d. $N\big(0, \frac{1}{m}\big)$ measurements (in which case $\|\mathbf{A}\|_{\mathrm{F}}^2$ is close to n with high probability), but also that no further improvements (beyond constant factors) are possible even for general measurement matrices having a similar Frobenius norm. The result holds under the assumption that $L \geq \Omega\big(\frac{1}{r}\sqrt{\frac{\alpha}{k}}\big)$ with a sufficiently large implied constant, which is a very mild assumption since for fixed r and α, the right-hand side tends to zero as k grows large (whereas typical Lipschitz constants are at least equal to one, if not much higher).

3 Generative Models Based on ReLU Networks

In this section, as opposed to considering general Lipschitz-continuous generative models, we focus on generative models given by neural networks with ReLU activations. Similar to the derivation of Corollary 1, we have the following corollary for ReLU-based networks from [10, Thm. 1.1].

Corollary 2. (Adapted from [10, Thm. 1.1]) *Let $G : \mathbb{R}^k \to \mathbb{R}^n$ be a generative model from a d-layer neural network with ReLU activations and at most w nodes per layer. The linear measurements model is the same as that in Corollary 1 with the domain of $G(\cdot)$ being $\mathbb{R}^k$ and the measurement matrix $\mathbf{A}$ being deterministic. Then, when $m \geq Ckd \log w$ for a universal constant C, there exists some $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $\frac{n}{2} \leq \|\mathbf{A}\|_{\mathrm{F}}^2 \leq \frac{3n}{2}$ such that for a universal constant C',*

$$\sup_{\mathbf{x}^* \in \mathrm{Range}(G)} \mathbb{E}\big[\|G(\hat{\mathbf{z}}) - \mathbf{x}^*\|_2^2\big] \leq C'\alpha. \tag{5}$$

Note that this result holds even when the domain $\mathcal{D} = \mathbb{R}^k$, so we do not need to distinguish between the rectangular and spherical domains. Moreover, this result makes no assumptions about the neural network weights (nor domain size), but rather, only the input size, width, and depth.

Thus far, we have considered forming a generative model $G : \mathbb{R}^k \to \mathbb{R}^n$ capable of producing k-group-sparse signals, which leads to a lower bound of $m = \Omega(k \log n)$. While this precise approach does not appear to be suited to properly understanding the dependence on width and depth in Corollary 2, we now show that a simple variant indeed suffices: We form a wide and/or deep ReLU network $G : \mathbb{R}^k \to \mathbb{R}^n$ capable of producing all (kk_0)-group-sparse signals having non-zero entries $\pm\xi$, where k_0 is a certain positive integer that may be much larger than one. The idea of the construction is illustrated in Figure 2, which shows the mappings for $k = 1$ (the general case simply repeats this structure in parallel to get an output dimension $n = n_0 k$). Note also that we need to replace the rectangular shapes by trapeziums (with high-gradient diagonals) to make them implementable with a ReLU network.

Again using the minimax lower bound for group-sparse recovery and a suitable choice of ξ, the following is proved in [17].

Theorem 2. *Fix $C_1, C_{\mathbf{A}} > 0$, and consider the problem of compressive sensing with generative models under i.i.d. $N\big(0, \frac{\alpha}{m}\big)$ noise, a measurement matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ satisfying $\|\mathbf{A}\|_{\mathrm{F}}^2 = C_{\mathbf{A}} n$, and the above-described generative model $G : \mathbb{R}^k \to \mathbb{R}^n$ with parameters k, k_0, n_0, and ξ. Then, if $n_0 \geq C_0 k_0$ for an absolute constant C_0, then there exists a constant $C_2 = \Theta(1)$ such that the choice $\xi = \sqrt{\frac{C_2 \alpha}{k}}$ yields the following:*

- *Any algorithm producing some* $\hat{\mathbf{x}}$ *satisfying* $\sup_{\mathbf{x}^* \in \mathrm{Range}(G)} \mathbb{E}\big[\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2\big] \leq C_1 \alpha$ *must also have* $m = \Omega\big(kk_0 \log \frac{n}{kk_0}\big)$ *(or equivalently* $m = \Omega\big(kk_0 \log \frac{n_0}{k_0}\big)$*, since* $n = n_0 k$*).*
- *The generative function* G *can be implemented as a ReLU network with a single hidden layer (i.e.,* $d = 2$*) of width at most* $w = O(k(\frac{n_0}{k_0})^{k_0})$*.*
- *Alternatively, if* $\frac{n_0}{k_0}$ *is an integer power of two, the generative function* G *can be implemented as a ReLU network with depth* $d = O\big(k_0 \log \frac{n_0}{k_0}\big)$ *and width* $w = O(n)$*.*

In the settings described in the second and third dot points, the sample complexity from Corollary 2 behaves as $kd \log w = O\big(kk_0 \log \frac{n_0}{k_0} + k \log k\big)$ *and* $kd \log w = O\big(kk_0 \cdot \log \frac{n_0}{k_0} \cdot \log n\big)$ *respectively.*

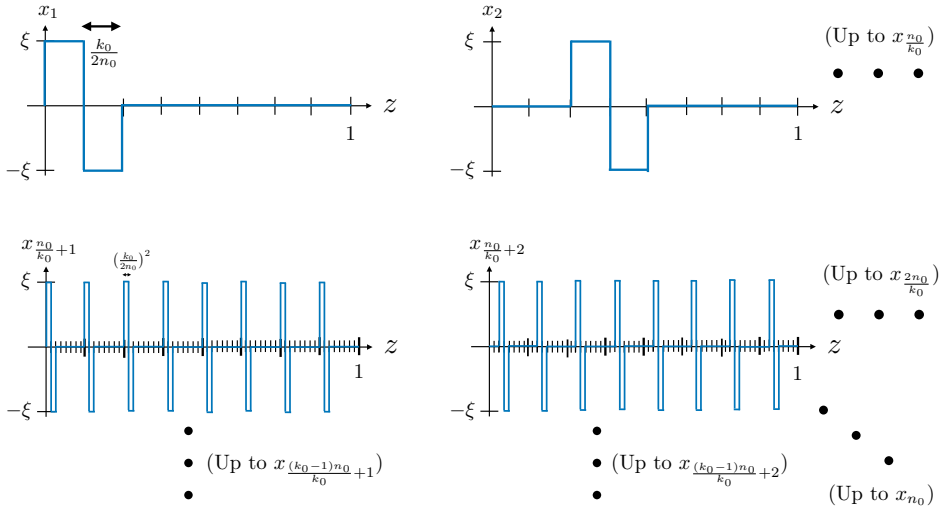

Figure 2: Recursively defined generative model mapping a single input z to $\mathbb{R}^{n_0}$. Note that this figure depicts an idealized version in which the mapping is discontinuous; the final generative model used replaces each vertical line by a sharp (but finite-gradient) transition, creating pulses that are trapezoidal rather than rectangular.

While we do not claim a lower bound for every possible combination of depth and width, the final statement of Theorem 2 reveals that the upper and lower bounds match up to a constant factor (high-width case with $\log k \leq O\big(k_0 \log \frac{n_0}{k_0}\big)$) or up to a $\log n$ factor (high-depth case).

The proof of the claim for the high-width case is based on the fact that in Figure 2, upon replacing the rectangles by trapeziums, each mapping is piecewise linear, and at the ℓ-th scale the number of pieces is $O\big(\big(\frac{n_0}{k_0}\big)^{\ell-1}\big)$, which we sum over $\ell = 1, \ldots, k_0$ to get the overall width. In the high-depth case, we exploit the periodic nature of the signals in Figure 2, and use the fact that depth-d neural networks can be used to produce periodic signals with $O(2^d)$ repetitions [19]. In our case, the maximum number of repetitions is $O\big(\big(\frac{n_0}{k_0}\big)^{k_0}$ (at the finest scale in Figure 2).

4 Conclusion

We have established, to our knowledge, the first lower bounds on the sample complexity for compressive sensing with generative models. To achieve these, we constructed generative models capable of producing group-sparse signals, and then applied a minimax lower bound for group-sparse recovery. For bounded Lipschitz-continuous generative models we matched the $O(k \log L)$ scaling law derived in [10], and for ReLU-based generative models, we showed that the dependence of the $O(kd \log w)$ bound from [10] has an optimal or near-optimal dependence on both the width and depth. A possible direction for future research is to understand what additional assumptions could be placed on the generative model to further reduce the sample complexity.

Acknowledgment. This work was supported by the Singapore National Research Foundation (NRF) under grant number R-252-000-A74-281.

References

[1] M. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso)," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2183–2202, May 2009.

[2] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7434–7464, Nov. 2013.

[3] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp, "Living on the edge: Phase transitions in convex programs with random data," *Information and Inference*, vol. 3, no. 3, pp. 224–294, 2014.

[4] J. Wen, Z. Zhou, J. Wang, X. Tang, and Q. Mo, "A sharp condition for exact support recovery with orthogonal matching pursuit," *IEEE Trans. Sig. Proc.*, vol. 65, no. 6, pp. 1370–1382, 2016.

[5] M. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5728–5741, Dec. 2009.

[6] E. Arias-Castro, E. J. Candes, and M. A. Davenport, "On the fundamental limits of adaptive sensing," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 472–481, Jan. 2013.

[7] E. J. Candes and M. A. Davenport, "How well can we estimate a sparse vector?" *Appl. Comp. Harm. Analysis*, vol. 34, no. 2, pp. 317–323, 2013.

[8] J. Scarlett and V. Cevher, "Limits on support recovery with probabilistic models: An information-theoretic framework," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 593–620, 2017.

[9] D. Foster, *Generative Deep Learning : Teaching Machines to Paint, Write, Compose and Play*. O'Reilly Media, Inc, USA, 2019.

[10] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, "Compressed sensing using generative models," in *Int. Conf. Mach. Learn. (ICML)*, 2017, pp. 537–546.

[11] V. Shah and C. Hegde, "Solving linear inverse problems using GAN priors: An algorithm with provable guarantees," in *IEEE Int. Conf. Acoust. Sp. Sig. Proc. (ICASSP)*, 2018, pp. 4609–4613.

[12] S. Jalali and X. Yuan, "Using auto-encoders for solving ill-posed linear inverse problems," *arXiv preprint arXiv:1901.05045*, 2019.

[13] M. Dhar, A. Grover, and S. Ermon, "Modeling sparse deviations for compressed sensing using generative models," in *Int. Conf. Mach. Learn. (ICML)*, 2018.

[14] P. Hand and V. Voroninski, "Global guarantees for enforcing deep generative priors by empirical risk," in *Conf. Learn. Theory (COLT)*, S. Bubeck, V. Perchet, and P. Rigollet, Eds., 2018, pp. 970–978.

[15] X. Wei, Z. Yang, and Z. Wang, "On the statistical rate of nonlinear recovery in generative models with heavy-tailed data," in *Int. Conf. Mach. Learn. (ICML)*, 2019.

[16] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, April 2010.

[17] Z. Liu and J. Scarlett, "Information-theoretic lower bounds for compressive sensing with generative models," 2019, https://arxiv.org/abs/1908.10744.

[18] J. C. Duchi and M. J. Wainwright, "Distance-based and continuum Fano inequalities with applications to statistical estimation," 2013, http://arxiv.org/abs/1311.2669.

[19] M. Telgarsky, "Benefits of depth in neural networks," in *Conf. Learn. Theory (COLT)*, 2016.