Non-Iterative Recovery from Nonlinear Observations using Generative Models

Jiulong Liu
LSEC, Institute of Computational Mathematics
and Scientific/Engineering Computing,
Academy of Mathematics and System Sciences,
Chinese Academy of Sciences, 100190, China
jiulongliu@lsec.cc.ac.cn

Zhaoqiang Liu *
Department of Computer Science
National University of Singapore
dcslizha@nus.edu.sg

Abstract

In this paper, we aim to estimate the direction of an underlying signal from its nonlinear observations following the semi-parametric single index model (SIM). Unlike for conventional compressed sensing where the signal is assumed to be sparse, we assume that the signal lies in the range of an L-Lipschitz continuous generative model with bounded k-dimensional inputs. This is mainly motivated by the tremendous success of deep generative models in various real applications. Our reconstruction method is non-iterative (though approximating the projection step may require an iterative procedure) and highly efficient, and it is shown to attain the near-optimal statistical rate of order $\sqrt{(k \log L)/m}$, where m is the number of measurements. We consider two specific instances of the SIM, namely noisy 1-bit and cubic measurement models, and perform experiments on image datasets to demonstrate the efficacy of our method. In particular, for the noisy 1-bit measurement model, we show that our non-iterative method significantly outperforms a state-of-the-art iterative method in terms of both accuracy and efficiency.

1. Introduction

The basic insight of compressed sensing (CS) is that a high-dimensional sparse signal can be accurately reconstructed from a small number of measurements [13]. For conventional CS, one aims to recover an s-sparse signal $\mathbf{x} \in \mathbb{R}^n$ from linear measurements of the form:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\eta}, \tag{1}$$

where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m]^T \in \mathbb{R}^{m \times n}$ is the measurement matrix, $\mathbf{y} = [y_1, y_2, \dots, y_m]^T \in \mathbb{R}^m$ is the observed vector, and $\boldsymbol{\eta} = [\eta_1, \eta_2, \dots, \eta_m]^T \in \mathbb{R}^m$ is the noise vector.

The CS problem has been popular over the last 1–2 decades, and its theoretical properties have been investigated in a significant body of works [1, 10, 57, 66]. For example, under i.i.d. random Gaussian measurements, it has been shown that an s-sparse signal can be accurately and efficiently reconstructed using $O(s \log(n/s))$ samples [2, 6, 56, 62].

The conventional CS problem has been extended in a wide variety of directions. Two important ones that we focus on in this paper are (i) considering general nonlinear measurement models, and (ii) assuming that the signal is in the range of a (deep) generative model, instead of being sparse. Both of these settings are practically well-motivated and have attracted sufficient attention in the past years. In the following, we briefly review the background of them.

1.1. Nonlinear Measurement Models

While the linear measurement model used in conventional CS can be a good testbed for illustrating conceptual phenomenon, in many real problems it may not be justifiable, or even plausible. For example, the binary measurement model used in 1-bit CS [5] has been of considerable interests because its hardware implementation is low-cost and efficient, and it is also robust to nonlinear distortions. In addition, 1-bit CS performs even better than conventional CS in certain situations [29]. The limitation of the linear data model motivates the study of general nonlinear measurement models, among which the semi-parametric single index model (SIM) is arguably the most popular one [22]. The SIM models the data as

$$y_i = f_i(\langle \mathbf{a}_i, \mathbf{x} \rangle), \quad i \in \{1, 2, \dots, m\}, \tag{2}$$

where $\mathbf{a}_i$ are i.i.d. realizations of a standard Gaussian vector $\mathbf{a} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n) \in \mathbb{R}^n$, with $\mathbf{a}_i^T$ being the i-th row of the measurement matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$; and $f_i : \mathbb{R} \to \mathbb{R}$ are i.i.d. realizations of an *unknown* random function f, independent of $\mathbf{a}_i$. The goal is to estimate the signal $\mathbf{x}$ using the knowledge of $\mathbf{A}$ and $\mathbf{y}$, despite the unknown nonlinearity f. It is well-known that $\mathbf{x}$ is generally unidentifiable

*Corresponding author.

in the SIM since any scaling of $\mathbf{x}$ can be absorbed into the unknown f. Therefore, it is popular to impose the identifiability constraint $\|\mathbf{x}\|_2 = 1$, and we are only able to estimate the direction of $\mathbf{x}$.

Let $y = f(\langle \mathbf{a}, \mathbf{x} \rangle)$ be the random variable that corresponds to each observation y_i. For an SIM and a standard normal random variable $g \sim \mathcal{N}(0,1)$ that is independent of the nonlinearity f, the following parameters are important to characterize the recovery performance of associated reconstruction algorithms:

$$\mu := \mathbb{E}[y\langle \mathbf{a}, \mathbf{x} \rangle] = \mathbb{E}[f(g)g], \tag{3}$$

$$\xi^2 := \mathbb{E}\left[y^2\right] = \mathbb{E}\left[f(g)^2\right], \tag{4}$$

$$\rho^2 := \mathrm{Var}[y\langle \mathbf{a}, \mathbf{x} \rangle - \mu] = \mathrm{Var}[f(g)g], \tag{5}$$

and

$$\theta^4 := \mathrm{Var}\left[y^2\right] = \mathrm{Var}\left[f(g)^2\right]. \tag{6}$$

Note that the parameters μ, ξ^2, ρ^2 and θ^4 are only used to characterize the recovery performance, and the knowledge of them will never be required for the reconstruction algorithms (since we assume that f is *unknown*). Later we will see (*cf.* Section 3) that we seek to make inference on the importance of the components of $\mathbf{x}$ via estimates of $\mu\mathbf{x}$, and thus we make the following widely-adopted assumption for the SIM [11, 37, 41, 47, 48]:

$$\mu = \mathbb{E}[f(g)g] \neq 0. \tag{7}$$

We highlight that some popular measurement models such as phase retrieval [7, 70] with $f(x) = x^2$ or $f(x) = |x|$ (or the noisy version) are typically beyond the scope of SIM since for these models, $\mu = \mathbb{E}[f(g)g] = 0$.

For the low-dimensional setting where the number of samples m is larger than the ambient dimension n, the SIM has been studied for a long time, dated back to the last century [18, 31, 59]. In recent years, high-dimensional SIMs have also received much attention, with various papers studying variable selection, estimation and inference mainly under the sparsity assumption [8, 11, 12, 14–16, 39, 41, 44, 45, 47, 48, 50, 64, 69]. In particular, the authors of [48] show that when the signal $\mathbf{x}$ is contained in $\mathcal{K}$ for some closed *star-shaped*[1] set $\mathcal{K} \subseteq \mathbb{R}^n$, and the observations y_i are *sub-Gaussian*[2], then the projection of $\frac{1}{m}\mathbf{A}^T\mathbf{y}$ onto $\mathcal{K}$ gives an accurate estimate of $\mathbf{x}$ with high probability provided that the number of samples is sufficiently large. Based on an idea that the nonlinear measurement model may be transformed into a scaled linear measurement model with an unconventional noise term, the authors of [47] show that the generalized Lasso approach, which minimizes the *linear least-squares* objective over a *convex* set $\mathcal{K}$, is able to return a reliable estimation of the signal in spite of the unknown nonlinearity. However, the range of a Lipschitz continuous generative model (such as a deep neural network), in general, cannot be star-shaped or convex. Moreover, the recovery error bounds in both works [47, 48] generally exhibit the $m^{-1/4}$ scaling, which is weaker than the typical $m^{-1/2}$ scaling.

[1] A set $\mathcal{K}$ is called star-shaped if $\lambda\mathcal{K} \subseteq \mathcal{K}$ for any $0 \le \lambda \le 1$.

[2] A random variable X is said to be sub-Gaussian if $\|X\|_{\psi_2} := \sup_{p\ge 1} p^{-1/2} \left(\mathbb{E}\left[|X|^p\right]\right)^{1/p} < \infty$.

1.2. Inverse Problems using Generative Models

Recently, motivated by enormous advances in deep generative models in an abundance of real applications, a new perspective has emerged in CS, in which the commonly-made sparsity assumption is replaced by the generative modeling assumption. That is, rather than being sparse, the signal is assumed to lie in the range of a (deep) generative model. In the seminal work [4], the authors study CS with generative priors, and characterize the number of random Gaussian linear measurements required for an accurate recovery. They also perform extensive numerical experiments on image datasets showing that to reconstruct the signal up to a given accuracy, compared to the sparse prior, using a pre-trained generative prior can reduce the required number of measurements by a large factor such as 5 to 10. There has been a substantial volume of follow-up works of [4], including [3, 9, 19–21, 24, 25, 32, 33, 40, 42, 43, 53, 61, 67, 68], just to name a few.

In particular, 1-bit CS with generative priors has been studied in [34, 49], for which the nonlinearity is assumed to be *known*. In [34], the authors provide a near-complete analysis for 1-bit CS with generative priors, and propose an iterative algorithm that can be thought of as a generative counterpart to the binary iterative hard thresholding algorithm [23]. The authors of [49] study 1-bit CS with ReLU neural network generative models (with no offsets). They propose an empirical risk minimization algorithm, and show that it can faithfully recover bounded target vectors from quantized noisy measurements. Perhaps closest to our work, near-optimal non-uniform recovery guarantees for CS with SIMs and generative priors have been provided in [37, 65]. The authors of [65] assume that the nonlinear function f is *differentiable* and propose estimators via first- and second-order Steins identity based score functions. The differentiability assumption is not satisfied for certain popular nonlinear measurement models such as 1-bit and other quantized models. The authors of [37] make the assumption of *sub-Gaussian* observations, which encompasses quantized measurement models. They propose a constrained linear least-squares estimator, with the constraint set being the range of a generative model. Both works [37, 65] are primarily theoretical, and neither practical algorithms nor nu-

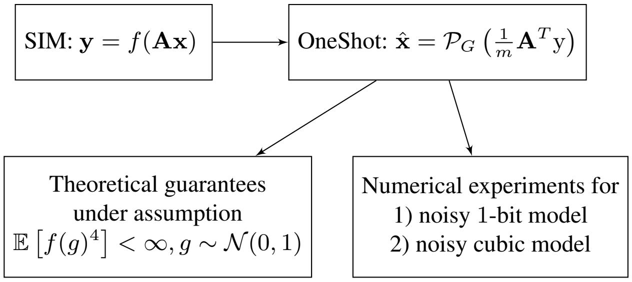


Figure 1. An illustration of main contributions.

merical results are provided in these works, even though attaining the estimators is practically difficult since the corresponding optimization problems are usually highly non-convex.

1.3. Contributions

The main contributions of this work are as follows:

- We propose a highly efficient non-iterative approach for nonlinear CS with SIMs and generative priors.
- We provide near-optimal recovery guarantees for our non-iterative approach. Notably, in our analysis, we do not require the differentiability assumption as in [65] or the assumption about sub-Gaussian observations as in [37].
- To verify the efficacy of our method, we perform extensive numerical experiments for distinct nonlinear measurement models on image datasets. In particular, for noisy 1-bit measurement model, we observe that along with faster computation, our non-iterative approach also leads to more accurate reconstruction compared to the iterative algorithm proposed in [34], which is the state-of-the-art (SOTA) algorithm for 1-bit CS with generative priors. In addition, for the noisy cubic measurement model, we observe that our non-iterative approach significantly outperforms several baselines, and performs on par with an iterative approach.

We also present Figure 1 to highlight the overall contributions. See (8), (9), and Section 4 for more details.

2. Problem Formulation

In this section, we provide some auxiliary results and formally formulate the problem we study. Before proceeding, we summarize notation we use throughout this paper.

2.1. Notation

We use upper and lower case boldface letters to denote matrices and vectors respectively. For any $N \in \mathbb{N}$, we use the shorthand notation $[N] = \{1, 2, \dots, N\}$, and we use $\mathbf{I}_N$ to represent the identity matrix in $\mathbb{R}^{N\times N}$. For a matrix $\mathbf{M}$, let $\|\mathbf{M}\|_{p,q} = \sup_{\|\mathbf{s}\|_p=1} \|\mathbf{M}\mathbf{s}\|_q$. In particular, we have that $\|\mathbf{M}\|_{2,2}$ represents the spectral norm of $\mathbf{M}$. Given two real values a, b, we write $a = O(b)$ if there exists an absolute constant $C < \infty$ such that $a \le Cb$, $a = \Omega(b)$ if there exists an absolute constant $c > 0$ such that $a \ge cb$, and $a = \Theta(b)$ if there exist absolute constants c and C such that $cb \le a \le Cb$. For any $r > 0$, we denote the radius-r ball in $\mathbb{R}^k$ as $B_2^k(r) := \{\mathbf{z} \in \mathbb{R}^k : \|\mathbf{z}\|_2 \le r\}$, and we use $\mathcal{S}^{n-1} := \{\mathbf{s} \in \mathbb{R}^n : \|\mathbf{s}\|_2 = 1\}$ to represent the unit sphere in $\mathbb{R}^n$. A generative model is a function $G : \mathcal{D} \to \mathbb{R}^n$ with latent dimension k, ambient dimension n, and input domain $\mathcal{D} \subseteq \mathbb{R}^k$. For a generative model G and a set $B \subseteq \mathcal{D}$, we write $G(B) = \{G(\mathbf{z}) : \mathbf{z} \in B\}$. Throughout the following, we focus on the setting that $\mathcal{D} = B_2^k(r)$ and $k \ll n$. We use $\mathcal{R}(G)$ to represent the range of G, i.e., $\mathcal{R}(G) = G(B_2^k(r))$.

2.2. Setup

Suppose that the generative model $G : B_2^k(r) \to \mathbb{R}^n$ is L-Lipschitz continuous, i.e., $\|G(\mathbf{z}_1) - G(\mathbf{z}_2)\|_2 \le L\|\mathbf{z}_1 - \mathbf{z}_2\|_2$ for any $\mathbf{z}_1, \mathbf{z}_2 \in B_2^k(r)$. The Lipschitzness assumption is naturally satisfied by some popular neural network generative models. For example, it is shown in [4, 34] that any fully-connected neural network generative model with bounded weights and widely-used activation functions (including Sigmoid, ReLU and Hyperbolic tangent functions) is Lipschitz continuous with the Lipschitz constant being $L = n^{\Theta(d)}$, where d is the depth of the neural network.

The nonlinear observations $y_1, y_2, \dots, y_m$ are assumed to be generated according to the SIM in (2), with $\mathbf{a}_i$ being i.i.d. realizations of $\mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ and $\mathbf{x} \in \mathcal{S}^{n-1}$ being the signal to estimate. We further assume that $\mu\mathbf{x} \in \mathcal{R}(G)$, where μ is a parameter depending on the nonlinearity f and is defined in (3), and $\mathcal{R}(G)$ refers to the range of G. Such an assumption is standard for nonlinear CS with generative priors and has also been made in [37, 65]. In this work, for the nonlinear function f, besides the popular assumption $\mu \ne 0$ as in (7), we only additionally assume that

$$\mathbb{E}\left[f(g)^4\right] < \infty, \tag{8}$$

where $g \sim \mathcal{N}(0, 1)$ is a standard normal random variable. Under this assumption, the parameters μ, ξ^2, ρ^2 and θ^4 defined in (3) to (6) are all finite. The condition in (8) holds for quantized measurement models, which do not satisfy the differentiability assumption in [65]. Moreover, it does not require $f(g)$ (corresponds to each observation y_i) to be sub-Gaussian as assumed in [37], thus enables us to deal with more general nonlinear measurement models such as $f(x) = x^3 + \eta$ (the noisy cubic model) or $f(x) = \mathrm{sign}(x)\big(x^2 + 1\big) + \eta$, where η is a zero-mean random Gaussian noise term.

To reconstruct the direction of the signal $\mathbf{x}$ from the knowledge of the measurement matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ and the observed vector $\mathbf{y} = [y_1, y_2, \ldots, y_m]^T \in \mathbb{R}^m$ (despite the unknown nonlinearity f), we set the estimated vector to be

$$\hat{\mathbf{x}} = \mathcal{P}_G\left(\frac{1}{m}\mathbf{A}^T\mathbf{y}\right), \tag{9}$$

where $\mathcal{P}_G(\cdot)$ is the projection operator onto $\mathcal{R}(G)$, i.e., $\mathcal{P}_G(\mathbf{s}) = \arg\min_{\mathbf{w}\in\mathcal{R}(G)} \|\mathbf{w} - \mathbf{s}\|_2$ for any $\mathbf{s} \in \mathbb{R}^n$. We refer to the reconstruction approach corresponding to (9) as *OneShot* to highlight its non-iterative nature, although approximating the projection step may require iterative procedures such as gradient descent.

3. Main Theorem

We have the following theorem concerning the recovery guarantee for OneShot in (9). Recall that μ, ξ^2, ρ^2 and θ^4 are parameters that are dependent only on the nonlinearity f and are defined in (3) to (6).

Theorem 1. *Suppose that the observed vector $\mathbf{y} \in \mathbb{R}^m$ is generated from the SIM in (2) with $\mathbf{a}_i$ being i.i.d. realizations of $\mathcal{N}(\mathbf{0}, \mathbf{I}_n)$, f satisfying (7) and (8), and $\mathbf{x} \in \mathcal{S}^{n-1} \cap \frac{1}{\mu}\mathcal{R}(G)$. Let $\hat{\mathbf{x}}$ be calculated from (9). Then, for any $\delta > 0$ satisfying $Lr = \Omega(\delta n)$ and $\delta = O\left(\xi\sqrt{\frac{k\log\frac{Lr}{\delta}}{m}}\right)$, we have with probability at least $1 - e^{-\Omega\left(k\log\frac{Lr}{\delta}\right)} - \frac{\theta^4}{m\xi^4} - \frac{\rho^2}{\xi^2 k \log\frac{Lr}{\delta}}$ that*

$$\|\hat{\mathbf{x}} - \mu\mathbf{x}\|_2 = O\left(\xi\sqrt{\frac{k\log\frac{Lr}{\delta}}{m}}\right). \tag{10}$$

Since a typical d-layer fully-connected neural network has the Lipschitz constant $L = n^{\Theta(d)}$ and r can be set to be of the same order as L [4], the assumption $Lr = \Omega(\delta n)$ is typically satisfied automatically. In addition, from the assumption (8), ξ is finite. Then, we have that the upper bound in (10) is roughly of order $\sqrt{(k\log L)/m}$, which is naturally conjectured to be near-optimal according to the information-theoretic lower bounds for linear CS with generative priors [26, 38]. Perhaps the major caveat to Theorem 1 is that it assumes the accurate projection. However, this is a standard assumption in relevant works, e.g., see [35, 46, 58], and in practice both gradient- and GAN-based projections have been shown to be highly effective [52, 58].

3.1. Proof Outline of Theorem 1

The proof of Theorem 1 is outlined below, with the full details provided in the supplementary material. Let $\mathcal{E}$ be the event that

$$\frac{1}{m}\sum_{i=1}^m y_i^2 \le 2\xi^2, \tag{11}$$

where ξ^2 is defined in (4). From the Chebyshev's inequality and the definition of θ^4 in (6), we have

$$\mathbb{P}(\mathcal{E}^c) \le \frac{\theta^4}{m\xi^4}. \tag{12}$$

Denoting $\mathbf{P} := \mathbf{x}\mathbf{x}^T$ as the orthogonal projection onto the subspace spanned by $\mathbf{x}$ and $\mathbf{P}^\perp := \mathbf{I}_n - \mathbf{x}\mathbf{x}^T$ as the orthogonal projection onto the orthogonal complement. Based on standard Gaussian concentration [63, Example 2.1], we have the following important lemma, for which the proof is available at the supplementary material.

Lemma 1. *Conditioned on the event $\mathcal{E}$, we have that for any $\varepsilon > 0$ and $\mathbf{s} \in \mathbb{R}^n$, with probability $1 - e^{-\Omega(\varepsilon)}$,*

$$\left|\frac{1}{m}\sum_{i=1}^m y_i\left\langle \mathbf{P}^\perp\mathbf{a}_i, \mathbf{s}\right\rangle\right| \le \frac{\xi\|\mathbf{s}\|_2\sqrt{\varepsilon}}{\sqrt{m}}. \tag{13}$$

Then we move on to present the proof outline.

Proof Outline of Theorem 1. Since $\hat{\mathbf{x}} = \mathcal{P}_G\left(\frac{1}{m}\mathbf{A}^T\mathbf{y}\right)$ and $\mu\mathbf{x} \in \mathcal{R}(G)$, we have $\left\|\frac{1}{m}\mathbf{A}^T\mathbf{y} - \hat{\mathbf{x}}\right\|_2 \le \left\|\frac{1}{m}\mathbf{A}^T\mathbf{y} - \mu\mathbf{x}\right\|_2$. Taking square on both sides, we obtain

$$\|\hat{\mathbf{x}} - \mu\mathbf{x}\|_2^2 \le 2\left\langle \frac{1}{m}\mathbf{A}^T\mathbf{y} - \mu\mathbf{x}, \hat{\mathbf{x}} - \mu\mathbf{x}\right\rangle. \tag{14}$$

Recall that $\mathbf{P}^\perp := \mathbf{I}_n - \mathbf{x}\mathbf{x}^T$. To upper bound the right-hand side of (14), we decompose $\frac{1}{m}\mathbf{A}^T\mathbf{y} - \mu\mathbf{x}$ as

$$\frac{1}{m}\mathbf{A}^T\mathbf{y} - \mu\mathbf{x} = \frac{1}{m}(\mathbf{I}_n - \mathbf{x}\mathbf{x}^T + \mathbf{x}\mathbf{x}^T)\mathbf{A}^T\mathbf{y} - \mu\mathbf{x} \tag{15}$$

$$= \frac{1}{m}\mathbf{P}^\perp\mathbf{A}^T\mathbf{y} + \left(\frac{1}{m}\mathbf{x}^T\mathbf{A}^T\mathbf{y} - \mu\right)\mathbf{x}. \tag{16}$$

Then, we obtain that

- the term $\left|\left\langle\frac{1}{m}\mathbf{P}^\perp\mathbf{A}^T\mathbf{y}, \hat{\mathbf{x}} - \mu\mathbf{x}\right\rangle\right|$ can be controlled using (12), Lemma 1, and a chaining argument [4];
- the term $\left|\left\langle\left(\frac{1}{m}\mathbf{x}^T\mathbf{A}^T\mathbf{y} - \mu\right)\mathbf{x}, \hat{\mathbf{x}} - \mu\mathbf{x}\right\rangle\right|$ can be controlled using the triangle inequality, and the Chebyshev's inequality with the definition of ρ^2 in (5).

Combining the two upper bounds and simplifying terms, we obtain the desired result in Theorem 1. □

3.2. Extensions of Theorem 1

In addition, we have the following corollary extending Theorem 1 in two directions. More specifically, this corollary shows that we can allow for *adversarial noise* that may be dependent on the measurement matrix $\mathbf{A}$ and the existence of *representation error* where $\mu\mathbf{x} \notin \mathcal{R}(G)$. It is worth noting that for the generalized Lasso approach considered in [37,47], handling representation error is not a simple task and is left open. The proof of Corollary 1 is placed in the supplementary material.

Corollary 1. *Suppose that the observed vector* $\mathbf{y} = [y_1, y_2, \ldots, y_m]^T \in \mathbb{R}^m$ *satisfies*

$$\frac{1}{\sqrt{m}} \sqrt{\sum_{i=1}^{m} \left(y_i - f_i(\langle \mathbf{a}_i, \mathbf{x} \rangle)\right)^2} \leq \nu \tag{17}$$

for some $\nu \geq 0$, *with* $\mathbf{a}_i$ *being i.i.d. realizations of* $\mathcal{N}(\mathbf{0}, \mathbf{I}_n)$, f_i *being i.i.d. realizations of* f, f *satisfying* (7) *and* (8), *and* $\mathbf{x} \in \mathcal{S}^{n-1}$. *Let* $\tilde{\mathbf{x}} = \mathcal{P}_G(\mu\mathbf{x})$ *be the vector in* $\mathcal{R}(G)$ *that is closest to* $\mu\mathbf{x}$, *and letting* $\hat{\mathbf{x}}$ *be calculated from* (9). *Then, for any* $\delta > 0$ *satisfying* $Lr = \Omega(\delta n)$ *and* $\delta = O\left(\xi\sqrt{\frac{k \log \frac{Lr}{\delta}}{m}}\right)$, *when* $m = \Omega(k \log \frac{Lr}{\delta})$, *we have with probability at least* $1 - e^{-\Omega\left(k \log \frac{Lr}{\delta}\right)} - \frac{\theta^4}{m\xi^4} - \frac{\rho^2}{\xi^2 k \log \frac{Lr}{\delta}}$ *that*

$$\|\hat{\mathbf{x}} - \mu\mathbf{x}\|_2 = O\left(\xi\sqrt{\frac{k \log \frac{Lr}{\delta}}{m}} + \nu + \|\tilde{\mathbf{x}} - \mu\mathbf{x}\|_2\right). \tag{18}$$

4. Experiments

The proposed method is evaluated on two special cases of the SIM in (2), namely a noisy 1-bit measurement model

$$y_i = \text{sign}(\langle \mathbf{a}_i, \mathbf{x} \rangle + e_i), \quad i \in [m], \tag{19}$$

where e_i are i.i.d. realizations of $\mathcal{N}\left(0, \sigma^2\right)$, and a noisy cubic measurement model

$$y_i = \langle \mathbf{a}_i, \mathbf{x} \rangle^3 + \eta_i, \quad i \in [m], \tag{20}$$

where η_i are i.i.d. realizations $\mathcal{N}\left(0, \sigma^2\right)$. Note that representation error is implicitly allowed in our experiments since the image vectors are not exactly contained in the range of the generative model. For simplicity, throughout this section, we do not consider adversarial noise. Experimental results with adversarial noise and the visualization of samples generated from pre-trained generative models are presented in the supplementary material.

4.1. Implementation Details

The experiments are performed on the MNIST [30] and CelebA [36] datasets. The MNIST dataset consists of $60,000$ images of handwritten digits. The size of each image in the MNIST dataset is 28×28, and thus the ambient dimension is $n = 784$. The CelebA dataset contains more than $200,000$ face images of celebrities. Each input image was cropped to a 64×64 RGB image, giving $n = 64 \times 64 \times 3 = 12288$ inputs per image. The generative model G for the MNIST dataset is set to be a pre-trained variational autoencoder (VAE) model with latent dimension $k = 20$. The encoder and decoder are both fully connected neural networks with two hidden layers, with the architecture being $20 - 500 - 500 - 784$. The VAE is trained by the Adam optimizer with a mini-batch size of 100 and a learning rate of 0.001 using the original training set of MNIST.

For the CelebA dataset, we choose the DCGAN [28, 51] for the generative model G. The architecture of the DCGAN follows that in [28] and the dimension of the input vector is set to be $k = 100$, with each entry being independently drawn from the standard normal distribution. We use the same training setup as in [4] to train the DCGAN on the training set of CelebA. Images that are selected from the testing sets (unseen by the pre-trained generative models) of the MNIST and CelebA datasets are used to generate 1-bit and cubic observations based on (19) and (20) respectively, with more details being listed in Table 1.

To approximate the projection step $\mathcal{P}_G(\cdot)$, we utilize the following two methods: 1) A gradient descent method that is performed using the Adam optimizer with 100 steps and a learning rate of 0.1. Such an *iterative* method is also used in [34, 35, 46, 58]. 2) A GAN-based projection method [52] that is *non-iterative* and much faster, for which we follow the settings in [52] to train the GAN models used in our experiments.

Table 1. The parameters of the measurement models

1-bit measurements		
Dataset	σ	m
MNIST	0.1, 0.5, 1, 5	25, 50, 100, 200, 400
CelebA	0.01,0.05,0.1,0.5	4000,6000,10000,15000
cubic measurements		
Dataset	σ	m
MNIST	0.1, 0.5, 1, 5	25, 50, 100, 200, 400
CelebA	0.01,0.05,0.1,0.5	4000,6000,10000,15000

More specifically, we perform the recovery tasks for the two nonlinear measurement models described in (19) and (20) using our proposed non-iterative method as in (9) (denoted by `OneShot` when using iterative projection, and `OneShotF` when using faster non-iterative projection), with comparison to some sparsity-based methods, and the method proposed in [4] (denoted by `CSGM`), as well as some generative model based projected iterative methods. For the sparse recovery with MNIST, we use Lasso [60] on the images in image domain with the shrinkage parameter setting to be 0.1 (denoted by `Lasso`). For the sparse recovery with CelebA, we use Lasso on the images in wavelet domain using 2D Daubechies-1 Wavelet Transform with the shrinkage parameter setting to be 0.00001 (denoted by `Lasso-W`). For the projected iterative method with 1-bit measurements, we use the method proposed in [34] (denoted by `BIPG` when using iterative projection, and `BIFPG` when using faster non-iterative projection), which is the SOTA method for 1-bit CS with generative priors, using the same pre-trained generative model as those described above. The corresponding formula is as follows:

$$\mathbf{x}^{(t+1)} = \mathcal{P}_G\left(\mathbf{x}^{(t)} + \lambda \mathbf{A}^T \left(\mathbf{y} - \operatorname{sign}\left(\mathbf{A}\mathbf{x}^{(t)}\right)\right)\right). \quad (21)$$

For the projected iterative method with cubic measurements, since there is no existing method specifically designed for this case, we compare with the method proposed in [46,58] (denoted by `PGD` when using iterative projection, and `FPGD` when using faster non-iterative projection), although it is initially designed for *linear* CS with generative priors. We also use the same pre-trained generative model as those described above. The corresponding formula is as follows:

$$\mathbf{x}^{(t+1)} = \mathcal{P}_G\left(\mathbf{x}^{(t)} + \lambda \mathbf{A}^T \left(\mathbf{y} - \mathbf{A}\mathbf{x}^{(t)}\right)\right). \quad (22)$$

For `BIPG` (or `BIFPG`) and `PGD` (or `FPGD`), we set the step size as $\lambda = 1/m$, the initial vector as $\mathbf{x}^{(0)} = \mathbf{0}$, and the total number of iterations as $T = 30$. All experiments are run using Python 3.6 and TensorFlow 1.5.0, with a NVIDIA GeForce GTX 1080 Ti 11GB GPU. To reduce the impact of local minima, we perform 10 random restarts, and choose the best among these. The cosine similarity refers to the inner product between the signal $\mathbf{x}$ and the normalized output vector of each recovery method, and it is averaged over both the testing images and these 10 random restarts.

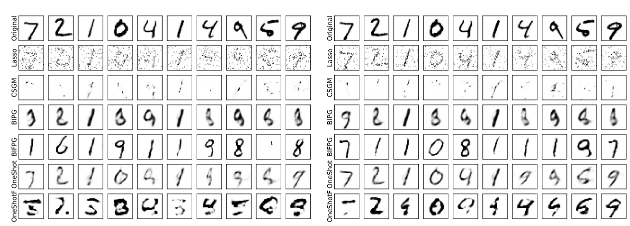
4.2. Recovery Results from 1-bit Measurements

The reconstructed images of the MNIST dataset from 1-bit measurements are shown in Figure 2, where we consider two settings with $\sigma = 1.0, m = 200$ and $\sigma = 0.1, m = 400$. In addition, we provide quantitative comparisons according to cosine similarity. To illustrate the effect of the sample size m, the cosine similarity in terms of $m \in \{25, 50, 100, 200, 400\}$ for MNIST reconstruction is plotted in Figure 3a, with fixing $\sigma = 1$. In addition, to illustrate the effect of the noise level σ, the cosine similarity in terms of $\sigma \in \{0.1, 0.5, 1, 5\}$ for MNIST reconstruction is plotted in Figure 3b, with fixing $m = 200$. We observe the following from Figures 2 and 3:

- `Lasso` and `CSGM` attain poor reconstructions.
- `OneShot` and `BIPG` significantly outperform all other methods, with the reconstruction performance of `OneShot` being slightly better than `BIPG`, though `OneShot` is non-iterative and performs much faster than `BIPG` (*cf.* Table 2).
- `OneShot` outperforms `OneShotF` and `BIPG` outperforms `BIFPG`, which amount to showing that at least for the MNIST dataset, the faster computation of the GAN-based projection step comes at the price of worse reconstruction.

The reconstructed images of the CelebA dataset from 1-bit measurements are shown in Figure 4, where we consider two settings $\sigma = 0.01, m = 4000$ and $\sigma = 0.05, m = 10000$. To illustrate the effect of the sample size m, the cosine similarity in terms of $m \in \{4000, 6000, 10000, 15000\}$ for CelebA reconstruction[3] is plotted in Figure 5a, with fixing $\sigma = 0.01$. In addition, to illustrate the effect of the noise level σ, the cosine similarity in terms of $\sigma \in \{0.01, 0.05, 0.1, 0.5\}$ for CelebA reconstruction is plotted in Figure 5b, with fixing $m = 4000$. We observe the following from Figures 4 and 5:

- `Lasso-W` and `CSGM` almost fail to recover the images, although the cosine similarities corresponding to `Lasso-W` are unusually high.
- The recovered images by `BIPG` and `BIFPG` are not plausible.
- Our proposed methods `OneShot` and `OneShotF` obtain much better reconstruction compared to all other methods. It is worth noting that the non-iterative approach `OneShot` (or `OneShotF`) significantly outperforms the projected iterative approach `BIPG` (or `BIFPG`). While this is a bit counter-intuitive, similar results concerning sparse priors have been reported in [71, Figures 1 to 3], showing that for synthetic data, a non-iterative approach leads to better recovery performance when compared with a sparse counterpart to `BIPG`.



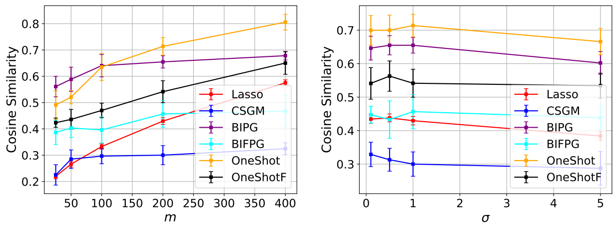
(a) $\sigma = 1.0$ and $m = 200$ (b) $\sigma = 0.1$ and $m = 400$

Figure 2. Examples of reconstructed images from 1-bit measurements on the MNIST images.

4.3. Recovery Results from Cubic Measurements

The reconstructed results from cubic measurements for the MNIST dataset are shown in Figure 6, and quantitative comparsions in terms of cosine similarity are presented in Figure 7. From Figures 6 and 7, we observe that `OneShot` outperforms all other competing methods (including `OneShotF` and `PGD`) by a large margin. The reconstructed results from cubic measurements for the

[3] Note that for 1-bit CS, it is very practical to set $m > n = 12288$ since 1-bit measurements can be taken at extremely high rates [71].



(a) Fixing $\sigma = 1.0$ and varying m

(b) Fixing $m = 200$ and varying σ

Figure 3. Quantitative comparisons according to the cosine similarity for 1-bit measurements on MNIST images.

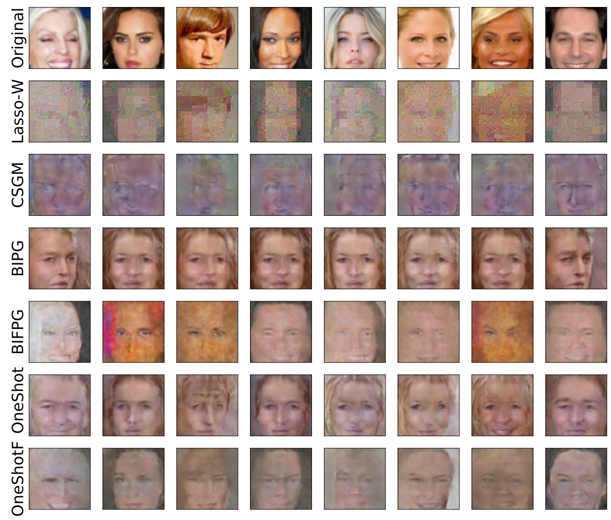
CelebA dataset are shown in Figure 8, and quantitative comparsions in terms of cosine similarity are presented in Figure 9. From these two figures, we observe that `CSGM` almost totally fails to recover the images, and `OneShot` and `OneShotF` can still obtain high-quality reconstructed images that are much better than those of `Lasso-W`. In particular, we observe that `OneShot` performs on par with `PGD`, for which the first iterative step reduces to `OneShot` when setting the initial vector $\mathbf{x}^{(0)} = \mathbf{0}$ and the step size $\lambda = 1/m$. This reveals that for `PGD`, one iterative step may be sufficient, and subsequent iterations will not lead to significant better reconstruction.

4.4. Running Time

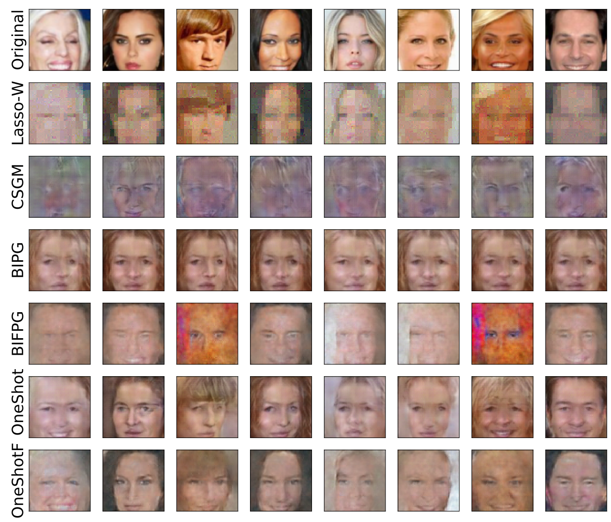
The running times shown in Table 2 illustrate that compared to using gradient-based iterative projection, using GAN-based non-iterative projection leads to much faster computation. `OneShot` is slower than sparsity-based recovery methods because approximating the projection step is time-consuming. However, since `OneShot` only requires one projection, it is much faster than the generative model based projected iterative methods `BIPG` and `PGD`. Furthermore, recall that from the numerical results, we observe that `OneShot` mostly achieves better reconstruction performance compared to that of `BIPG` and `PGD`, as well as that of `CSGM` and sparsity-based recovery methods.

5. Conclusion and Future Work

In this paper, we propose a non-iterative approach for nonlinear CS with SIMs and generative priors. We make the assumption (8) for the nonlinear function f, and this enables us to study both the noisy 1-bit (this cannot be handled by [65] due to the differentiability assumption) and cubic (this cannot be handled by [37] due to the sub-Gaussianity assumption for the observations) measurement models. We show that our approach attains the near-optimal statistical rate $O(\sqrt{(k \log L)/m})$. We also show via extensive numerical experiments that our approach is efficient and leads



(a) $\sigma = 0.01$ and $m = 4000$



(b) $\sigma = 0.05$ and $m = 10000$

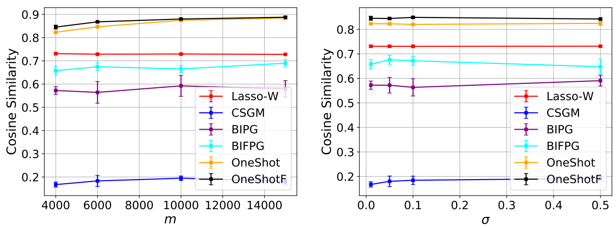
Figure 4. Examples of reconstructed images from 1-bit measurements on CelebA images.

to better reconstruction compared to various baselines.

Possible extensions include 1) providing a matching information-theoretic lower bound for CS with SIMs and generative priors; 2) extending to the case that the measurement vectors $\mathbf{a}_i$ are i.i.d. realizations of $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ with an unknown covariance matrix $\mathbf{\Sigma}$, instead of the current assumption that $\mathbf{a}_i$ are i.i.d. realizations of $\mathcal{N}(\mathbf{0}, \mathbf{I}_n)$; 3) training generative models [17,27,54,55] that are more advanced compared to the DCGAN.

6. Acknowledgment

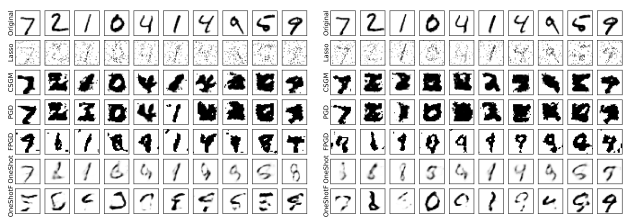
J. Liu was partially supported by the Fund of the Youth Innovation Promotion Association, CAS (2022002).



(a) Fixing $\sigma = 0.01$ and varying m

(b) Fixing $m = 4000$ and varying σ

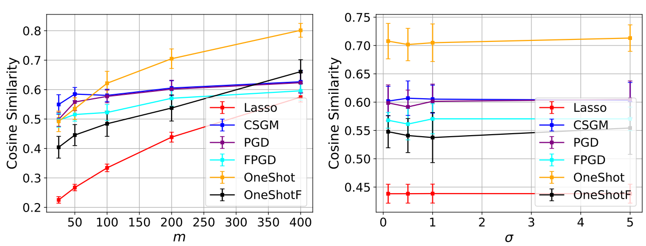
Figure 5. Quantitative comparisons according to the cosine similarity for 1-bit measurements on CelebA images.



(a) $\sigma = 1$ and $m = 200$

(b) $\sigma = 0.1$ and $m = 400$

Figure 6. Examples of reconstructed images from cubic measurements on MNIST images.



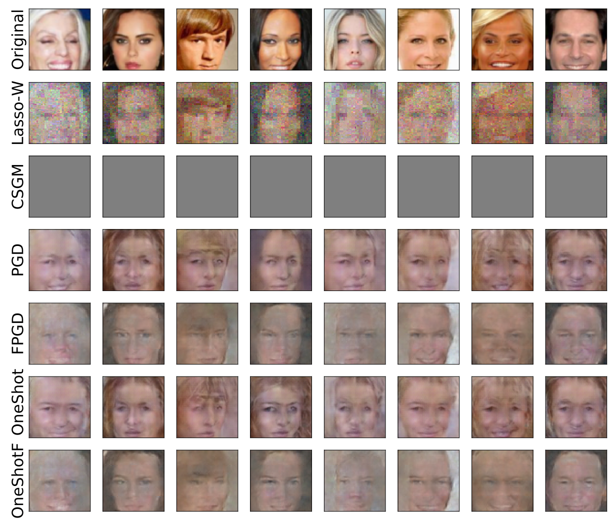
(a) Fixing $\sigma = 1$, varying m

(b) Fixing $m = 200$, varying σ

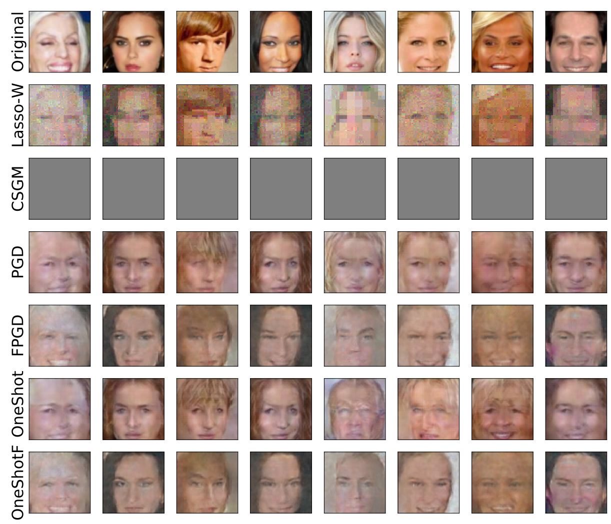
Figure 7. Quantitative comparisons according to the cosine similarity for cubic measurements on MNIST images.

Table 2. The averaged time cost (secs) per reconstruction.

	1-bit		cubic	
	MNIST $m = 200$	CelebA $m = 4000$	MNIST $m = 200$	CelebA $m = 4000$
`Lasso`	0.12	/	0.11	/
`Lasso-W`	/	12.81	/	16.27
`CSGM`	0.63	185.26	0.79	186.54
`BIPG`	14.30	2589.11	/	/
`BIFPG`	0.54	8.93	/	/
`PGD`	/	/	12.74	2539.24
`FPGD`	/	/	0.83	8.09
`OneShot`	0.34	130.63	0.61	128.89
`OneShotF`	0.03	0.61	0.03	0.66

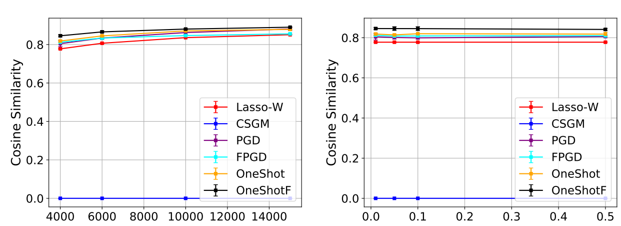


(a) $\sigma = 0.01$ and $m = 4000$



(b) $\sigma = 0.05$ and $m = 10000$

Figure 8. Examples of reconstructed images from cubic measurements on CelebA images.



(a) Fixing $\sigma = 0.01$ and varying m

(b) Fixing $m = 4000$ and varying σ

Figure 9. Quantitative comparisons according to the cosine similarity for cubic measurements on CelebA images.

References

[1] Dennis Amelunxen, Martin Lotz, Michael B McCoy, and Joel A Tropp. Living on the edge: Phase transitions in convex programs with random data. *Inf. Inference*, 3(3):224–294, 2014. 1

[2] Ery Arias-Castro, Emmanuel J Candes, and Mark A Davenport. On the fundamental limits of adaptive sensing. *IEEE Trans. Inf. Theory*, 59(1):472–481, 2012. 1

[3] Muhammad Asim, Max Daniels, Oscar Leong, Ali Ahmed, and Paul Hand. Invertible generative models for inverse problems: Mitigating representation error and dataset bias. In *Int. Conf. Mach. Learn. (ICML)*, pages 399–409. PMLR, 2020. 2

[4] Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G Dimakis. Compressed sensing using generative models. In *Int. Conf. Mach. Learn. (ICML)*, pages 537–546, 2017. 2, 3, 4, 5

[5] Petros T Boufounos and Richard G Baraniuk. 1-bit compressive sensing. In *Conf. Inf. Sci. Syst. (CISS)*, pages 16–21. IEEE, 2008. 1

[6] Emmanuel J Candes and Mark A Davenport. How well can we estimate a sparse vector? *Appl. Comp. Harm. Analysis*, 34(2):317–323, 2013. 1

[7] Emmanuel J Candès, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Trans. Inf. Theory*, 61(4):1985–2007, 2015. 2

[8] Longjie Cheng, Peng Zeng, and Yu Zhu. BS-SIM: An effective variable selection method for high-dimensional single index model. *Electron. J. Stat.*, 11(2):3522–3548, 2017. 2

[9] Manik Dhar, Aditya Grover, and Stefano Ermon. Modeling sparse deviations for compressed sensing using generative models. In *Int. Conf. Mach. Learn. (ICML)*, pages 1214–1223. PMLR, 2018. 2

[10] David L Donoho, Adel Javanmard, and Andrea Montanari. Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing. *IEEE Trans. Inf. Theory*, 59(11):7434–7464, 2013. 1

[11] Hamid Eftekhari, Moulinath Banerjee, and Yaacov Ritov. Inference in high-dimensional single-index models under symmetric designs. *J. Mach. Learn. Res.*, 22:27–1, 2021. 2

[12] Jared C Foster, Jeremy MG Taylor, and Bin Nan. Variable selection in monotone single-index models via the adaptive Lasso. *Stat. Med.*, 32(22):3944–3954, 2013. 2

[13] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Springer New York, 2013. 1

[14] Ravi Ganti, Nikhil Rao, Rebecca M Willett, and Robert Nowak. Learning single index models in high dimensions. *https://arxiv.org/1506.08910*, 2015. 2

[15] Martin Genzel. High-dimensional estimation of structured signals from non-linear observations with general convex loss functions. *IEEE Trans. Inf. Theory*, 63(3):1601–1619, 2016. 2

[16] Larry Goldstein, Stanislav Minsker, and Xiaohan Wei. Structured signal recovery from non-linear and heavy-tailed measurements. *IEEE Trans. Inf. Theory*, 64(8):5513–5530, 2018. 2

[17] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. *Conf. Neur. Inf. Proc. Sys. (NeurIPS)*, 30, 2017. 7

[18] Aaron K Han. Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator. *J. Econom.*, 35(2-3):303–316, 1987. 2

[19] Paul Hand, Oscar Leong, and Vladislav Voroninski. Phase retrieval under a generative prior. In *Conf. Neur. Inf. Proc. Sys. (NeurIPS)*, pages 9154–9164, 2018. 2

[20] Paul Hand and Vladislav Voroninski. Global guarantees for enforcing deep generative priors by empirical risk. In *Conf. Learn. Theory (COLT)*, pages 970–978. PMLR, 2018. 2

[21] Reinhard Heckel and Paul Hand. Deep decoder: Concise image representations from untrained non-convolutional networks. In *Int. Conf. Learn. Repr. (ICLR)*, 2019. 2

[22] Joel L Horowitz. *Semiparametric and nonparametric methods in econometrics*, volume 12. Springer, 2009. 1

[23] 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 Trans. Inf. Theory*, 59(4):2082–2102, 2013. 2

[24] A Jalal, S Karmalkar, A Dimakis, and E Price. Instance-optimal compressed sensing via posterior sampling. In *Int. Conf. Mach. Learn. (ICML)*, 2021. 2

[25] Ajil Jalal, Liu Liu, Alexandros G Dimakis, and Constantine Caramanis. Robust compressed sensing using generative models. *Conf. Neur. Inf. Proc. Sys. (NeurIPS)*, 2020. 2

[26] Akshay Kamath, Sushrut Karmalkar, and Eric Price. On the power of compressed sensing with generative models. In *Int. Conf. Mach. Learn. (ICML)*, pages 5101–5109, 2020. 4

[27] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks. *Conf. Neur. Inf. Proc. Sys. (NeurIPS)*, 34, 2021. 7

[28] Taehoon Kim. A tensorflow implementation of deep convolutional generative adversarial networks. *Software available at https://github. com/carpedm20/DCGAN-tensorflow*, 2017. 5

[29] Jason N Laska and Richard G Baraniuk. Regime change: Bit-depth versus measurement-rate in compressive sensing. *IEEE Trans. Sig. Proc.*, 60(7):3496–3505, 2012. 1

[30] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proc. IEEE*, 86(11):2278–2324, 1998. 5

[31] Ker-Chau Li and Naihua Duan. Regression analysis under link violation. *Ann. Stat.*, pages 1009–1052, 1989. 2

[32] Zhaoqiang Liu, Subhroshekhar Ghosh, Jun Han, and Jonathan Scarlett. Robust 1-bit compressive sensing with partial Gaussian circulant matrices and generative priors. *https://arxiv.org/2108.03570*, 2021. 2

[33] Zhaoqiang Liu, Subhroshekhar Ghosh, and Jonathan Scarlett. Towards sample-optimal compressive phase retrieval with sparse and generative priors. In *Conf. Neur. Inf. Proc. Sys. (NeurIPS)*, 2021. 2

[34] Zhaoqiang Liu, Selwyn Gomes, Avtansh Tiwari, and Jonathan Scarlett. Sample complexity bounds for 1-bit com-

pressive sensing and binary stable embeddings with generative priors. In *Int. Conf. Mach. Learn. (ICML)*, 2020. 2, 3, 5

[35] Zhaoqiang Liu, Jiulong Liu, Subhroshekhar Ghosh, Jun Han, and Jonathan Scarlett. Generative principal component analysis. In *Int. Conf. Learn. Repr. (ICLR)*, 2022. 4, 5

[36] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pages 3730–3738, 2015. 5

[37] Zhaoqiang Liu and Jonathan Scarlett. The generalized Lasso with nonlinear observations and generative priors. In *Conf. Neur. Inf. Proc. Sys. (NeurIPS)*, volume 33, 2020. 2, 3, 4, 7

[38] Zhaoqiang Liu and Jonathan Scarlett. Information-theoretic lower bounds for compressive sensing with generative models. *IEEE J. Sel. Areas Inf. Theory*, 1(1):292–303, 2020. 4

[39] Shikai Luo and Subhashis Ghosal. Forward selection and estimation in high dimensional single index models. *Stat. Methodol.*, 33:172–179, 2016. 2

[40] Sachit Menon, Alexandru Damian, Shijia Hu, Nikhil Ravi, and Cynthia Rudin. Pulse: Self-supervised photo upsampling via latent space exploration of generative models. In *IEEE Comput. Soc. Conf. Comput. Vis. Pattern. Recognit. (CVPR)*, pages 2437–2445, 2020. 2

[41] Matey Neykov, Jun S Liu, and Tianxi Cai. L_1-regularized least squares for support recovery of high dimensional single index models with Gaussian designs. *J. Mach. Learn. Res.*, 17(1):2976–3012, 2016. 2

[42] Thanh V Nguyen, Gauri Jagatap, and Chinmay Hegde. Provable compressed sensing with generative priors via Langevin dynamics. *https://arxiv.org/2102.12643*, 2021. 2

[43] Gregory Ongie, Ajil Jalal, Christopher A Metzler, et al. Deep learning techniques for inverse problems in imaging. *IEEE J. Sel. Areas Inf. Theory*, 1(1):39–56, 2020. 2

[44] Samet Oymak and Mahdi Soltanolkotabi. Fast and reliable parameter estimation from nonlinear observations. *SIAM J. Optim.*, 27(4):2276–2300, 2017. 2

[45] Ashwin Pananjady and Dean P Foster. Single-index models in the high signal regime. *IEEE Trans. Inf. Theory*, 67(6):4092–4124, 2021. 2

[46] Pei Peng, Shirin Jalali, and Xin Yuan. Solving inverse problems via auto-encoders. *IEEE J. Sel. Areas Inf. Theory*, 1(1):312–323, 2020. 4, 5, 6

[47] Yaniv Plan and Roman Vershynin. The generalized Lasso with non-linear observations. *IEEE Trans. Inf. Theory*, 62(3):1528–1537, 2016. 2, 4

[48] Yaniv Plan, Roman Vershynin, and Elena Yudovina. High-dimensional estimation with geometric constraints. *Inf. Inference*, 6(1):1–40, 2017. 2

[49] Shuang Qiu, Xiaohan Wei, and Zhuoran Yang. Robust one-bit recovery via ReLU generative networks: Improved statistical rates and global landscape analysis. In *Int. Conf. Mach. Learn. (ICML)*, 2020. 2

[50] Peter Radchenko. High dimensional single index models. *J. Multivar. Anal.*, 139:266–282, 2015. 2

[51] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015. 5

[52] Ankit Raj, Yuqi Li, and Yoram Bresler. GAN-based projector for faster recovery with convergence guarantees in linear inverse problems. In *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2019. 4, 5

[53] JH Rick Chang, Chun-Liang Li, Barnabas Poczos, BVK Vijaya Kumar, and Aswin C Sankaranarayanan. One network to solve them all–solving linear inverse problems using deep projection models. In *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pages 5888–5897, 2017. 2

[54] Kevin Roth, Aurelien Lucchi, Sebastian Nowozin, and Thomas Hofmann. Stabilizing training of generative adversarial networks through regularization. *Conf. Neur. Inf. Proc. Sys. (NeurIPS)*, 30, 2017. 7

[55] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *https://arxiv.org/2104.07636*, 2021. 7

[56] Jonathan Scarlett and Volkan Cevher. Limits on support recovery with probabilistic models: An information-theoretic framework. *IEEE Trans. Inf. Theory*, 63(1):593–620, 2016. 1

[57] Jonathan Scarlett and Volkan Cevher. An introductory guide to Fano's inequality with applications in statistical estimation. *https://arxiv.org/1901.00555*, 2019. 1

[58] Viraj Shah and Chinmay Hegde. Solving linear inverse problems using GAN priors: An algorithm with provable guarantees. In *Int. Conf. Acoust. Sp. Sig. Proc. (ICASSP)*, pages 4609–4613. IEEE, 2018. 4, 5, 6

[59] Robert P Sherman. The limiting distribution of the maximum rank correlation estimator. *Econometrica*, pages 123–137, 1993. 2

[60] Robert Tibshirani. Regression shrinkage and selection via the Lasso. *J. Royal Stat. Soc. Series B*, pages 267–288, 1996. 5

[61] Dave Van Veen, Ajil Jalal, Mahdi Soltanolkotabi, et al. Compressed sensing with deep image prior and learned regularization. *https://arxiv.org/1806.06438*, 2018. 2

[62] Martin J Wainwright. Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting. *IEEE Trans. Inf. Theory*, 55(12):5728–5741, 2009. 1

[63] Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019. 4

[64] Xiaohan Wei. Structured recovery with heavy-tailed measurements: A thresholding procedure and optimal rates. *https://arxiv.org/abs/1804.05959*, 2018. 2

[65] Xiaohan Wei, Zhuoran Yang, and Zhaoran Wang. On the statistical rate of nonlinear recovery in generative models with heavy-tailed data. In *Int. Conf. Mach. Learn. (ICML)*, pages 6697–6706, 2019. 2, 3, 7

[66] Jinming Wen, Zhengchun Zhou, Jian Wang, Xiaohu Tang, and Qun Mo. A sharp condition for exact support recovery with orthogonal matching pursuit. *IEEE Trans. Sig. Proc.*, 65(6):1370–1382, 2016. 1

[67] Jay Whang, Qi Lei, and Alexandros G Dimakis. Compressed sensing with invertible generative models and dependent noise. *https://arxiv.org/2003.08089*, 2020. 2

[68] Yan Wu, Mihaela Rosca, and Timothy Lillicrap. Deep compressed sensing. In *Int. Conf. Mach. Learn. (ICML)*, pages 6850–6860. PMLR, 2019. 2

[69] Zhuoran Yang, Krishnakumar Balasubramanian, and Han Liu. High-dimensional non-Gaussian single index models via thresholded score function estimation. In *Int. Conf. Mach. Learn. (ICML)*, pages 3851–3860. PMLR, 2017. 2

[70] Huishuai Zhang, Yi Zhou, Yingbin Liang, and Yuejie Chi. A nonconvex approach for phase retrieval: Reshaped Wirtinger flow and incremental algorithms. *J. Mach. Learn. Res.*, 18, 2017. 2

[71] Lijun Zhang, Jinfeng Yi, and Rong Jin. Efficient algorithms for robust one-bit compressive sensing. In *Int. Conf. Mach. Learn. (ICML)*, pages 820–828, 2014. 6