Efficient Algorithms for Non-gaussian Single Index Models with Generative Priors

Junren Chen[1], **Zhaoqiang Liu**[2*]

[1] University of Hong Kong
[2] University of Electronic Science and Technology of China
chenjr58@connect.hku.hk, zqliu12@gmail.com

Abstract

In this work, we focus on high-dimensional single index models with non-Gaussian sensing vectors and generative priors. More specifically, our goal is to estimate the underlying signal from i.i.d. realizations of the semi-parameterized single index model, where the underlying signal is contained in (up to a constant scaling) the range of a Lipschitz continuous generative model with bounded low-dimensional inputs, the sensing vector follows a non-Gaussian distribution, the noise is a random variable that is independent of the sensing vector, and the unknown non-linear link function is differentiable. Using the first- and second-order Stein's identity, we introduce efficient algorithms to obtain estimated vectors that achieve the near-optimal statistical rate. Experimental results on image datasets are provided to support our theory.

Introduction

There has been a significant amount of research into the theoretical and computational aspects of high-dimensional linear inverse problems. The standard compressed sensing problem, which relies on the assumption of low-complexity structure through sparsity to achieve the accurate recovery of a high-dimensional signal using a small number of (noisy) linear measurements, has been extensively investigated since 2006 and is now well-understood (Candès, Romberg, and Tao 2006; Candes, Romberg, and Tao 2006; Donoho 2006; Candès and Wakin 2008; Foucart et al. 2013).

Although the linear measurement model is widely used in conventional compressed sensing and can be a useful tool for demonstrating conceptual phenomena, it may not be appropriate or even feasible in many real-world scenarios. For example, the phase retrieval problem is encountered in various fields of science and engineering, such as acoustics, astronomy, microscopy, optics, quantum information, wireless communications, and X-ray crystallography, where direct linear measurements are not possible, and only the magnitudes or intensities of the measurements can be recorded (Candes, Li, and Soltanolkotabi 2015). The limitations of the linear data model have led to the exploration of general non-linear measurement models, such as the semi-parametric single index model (SIM), which has been extensively studied for both conventional and high-dimensional settings (Han 1987; Sherman 1993; Hristache, Juditsky, and Spokoiny 2001; Plan and Vershynin 2016; Plan, Vershynin, and Yudovina 2017). The data of the SIM can be of the following form:

$$y_i = f\left(\mathbf{a}_i^T\mathbf{x}^*, w_i\right), \quad i = 1, 2, \ldots, m, \tag{1}$$

where $\{y_i\}_{i=1}^m$ are the observations, f is an unknown nonlinear link function, $\{\mathbf{a}_i\}_{i=1}^m$ are the sensing vectors, $\mathbf{x}^* \in \mathbb{R}^n$ is the target signal, and $\{w_i\}_{i=1}^m$ are the noises.

The semi-parametric single index model (SIM) is a flexible nonlinear measurement model that can accommodate various measurement models of interest, including those related to the first-order link functions defined after (4). However, to utilize the first-order Stein's identity or to transform the nonlinear model into an unconventional linear one, the model must satisfy an important assumption like the one presented in equation (4). Unfortunately, this assumption is not valid for the non-linear link functions associated with the phase retrieval problem. To address this issue, we propose to use second-order link functions (defined after (12)) based on the second-order Stein's identity.

Furthermore, most prior research on SIMs assumes that the sensing vectors are Gaussian-distributed and that the underlying signal is limited to a low-complexity set that conforms to traditional modeling assumptions, such as sparsity and low-rankness. In contrast, the present work focuses on the scenario where the sensing vectors are non-Gaussian and the signal falls within the range of a generative model. For both first- and second-order link functions, we present efficient algorithms that can find estimated vectors achieving the near-optimal statistical rate.

Related Work

In this subsection, we provide a brief overview of relevant works on SIMs.

Low-dimensional SIMs: There is a substantial body of research on the SIM in the conventional *low-dimensional* setting where the number of measurements m is greater than the data dimension n, see, e.g., (Han 1987; Li and Duan 1989; Sherman 1993; Hristache, Juditsky, and Spokoiny 2001). We do not aim to cover all of them since our attention is focused on the high-dimensional scenario where $m \ll n$.

*Corresponding author.

High-dimensional SIMs without generative priors (excluding phase retrieval models): The SIM has also garnered considerable research in the high-dimensional setting, utilizing Lasso (Tibshirani 1996) and related high-dimensional linear estimators (Bühlmann and Van De Geer 2011), with various studies focusing on variable selection, estimation, and inference under *traditional low-complexity structures such as sparsity and low-rankness* (Ganti et al. 2015; Radchenko 2015; Luo and Ghosal 2016; Neykov, Liu, and Cai 2016; Cheng, Zeng, and Zhu 2017; Oymak and Soltanolkotabi 2017; Pananjady and Foster 2021).

For instance, the authors of (Plan and Vershynin 2016) show that the generalized Lasso approach works for SIMs in high dimensions, and a tighter but asymptotic result is presented in (Thrampoulidis, Abbasi, and Hassibi 2015) under the same setting. In addition, the results of (Plan and Vershynin 2016) are extended to the case of general convex loss functions in (Genzel 2016), and it is shown in (Plan, Vershynin, and Yudovina 2017) that a simple projection-based approach leads to accurate reconstruction in high dimensions. All these works adopt the assumption of *Gaussian* sensing vectors. In the two recent works (Goldstein, Minsker, and Wei 2018; Wei 2018), high-dimensional SIMs with heavy-tailed elliptical symmetric sensing vectors are investigated and the authors introduce thresholded least square estimators that achieve similar performance guarantees as those for the Gaussian case. The assumption of Gaussian or elliptical symmetric sensing vectors has been relaxed in (Yang et al. 2015; Soltani and Hegde 2017; Zhang, Yang, and Wang 2018) to allow for general sensing vectors, but at the cost of assuming a known and monotonic link function. To obtain a consistent estimator for general sensing vectors with unknown (but differentiable) non-linear link functions, the authors of (Yang, Balasubramanian, and Liu 2017) propose thresholded score function estimators based on Stein's identity. However, these methods are restricted to the estimation of sparse or low-rank signals and thus their performance heavily depends on the chosen basis.

High-dimensional SIMs without generative priors (encompassing phase retrieval models): It is worth mentioning that all the above-discussed works for high-dimensional SIMs rely on the pivotal assumption similar to (4), for the purpose of making use of the first-order Stein's identity or converting the non-linear model into an unconventional linear model. Such an assumption fails to hold for non-linear link functions corresponding to the popular *phase retrieval* problem. SIMs that encompass phase retrieval models as special cases have been studied in (Yang et al. 2019; Neykov, Wang, and Liu 2020), under the assumptions that the sensing vectors are Gaussian and the target signal is sparse. The assumption of Gaussian sensing vectors has been relaxed in (Yang et al. 2017) by employing the second-order Stein's method, but this work shares the same limitation as (Yang, Balasubramanian, and Liu 2017) and is only applicable to the reconstruction of sparse or low-rank signals.

High-dimensional SIMs with generative priors: In recent years, motivated by tremendous successful applications of deep generative models in an abundance of real-world applications, to solve high-dimensional inverse problems, there has been an increasing interest in replacing the commonly made sparsity assumption with the generative modeling assumption. That is, instead of being sparse, the underlying signal is assumed to be contained in the range of a generative model. The seminal work (Bora et al. 2017) studies linear compressed sensing with generative priors and demonstrates via numerical experiments on image datasets that using a pre-trained generative prior can significantly reduce the number of required measurements (compared to that of using the sparse prior) for accurate signal recovery. This has led to a significant volume of follow-up works in (Dhar, Grover, and Ermon 2018; Hand, Leong, and Voroninski 2018; Heckel and Hand 2018; Shah and Hegde 2018; Aubin et al. 2019; Jagatap and Hegde 2019; Latorre, Cevher et al. 2019; Asim et al. 2020; Jalal et al. 2020; Ongie et al. 2020; Daras et al. 2021; Jalal et al. 2021b,a; Joshi et al. 2021), which explore various aspects of high-dimensional inverse problems with generative priors. A recent literature review in this area can be found in (Scarlett et al. 2022).

In particular, under generative priors, there have been several studies on 1-bit compressed sensing (Qiu, Wei, and Yang 2020; Liu et al. 2020) and phase retrieval (Hand, Leong, and Voroninski 2018; Hyder et al. 2019; Jagatap and Hegde 2019; Aubin et al. 2020; Shamshad and Ahmed 2020; Liu, Ghosh, and Scarlett 2021; Killedar and Seelamantula 2022), among others. However, it is worth noting that the non-linear link function used in these studies is typically known and specific.

SIMs that account for unknown nonlinearity and generative priors have been explored in several studies, including (Wei, Yang, and Wang 2019; Liu and Scarlett 2020a; Liu and Liu 2022; Liu and Han 2022; Liu, Wang, and Liu 2022; Chen et al. 2023). More specifically, the works (Liu and Scarlett 2020a; Liu and Liu 2022; Liu and Han 2022; Chen et al. 2023) focus on SIMs that are unable to handle phase retrieval models, and provide recovery guarantees for the generalized Lasso approach and practical algorithms under the assumptions of Gaussian sensing vectors. SIMs that encompass phase retrieval models have been studied in (Liu, Wang, and Liu 2022) under a generative prior, but this work is also restricted to Gaussian sensing vectors. The work (Wei, Yang, and Wang 2019) is the most relevant to ours. In (Wei, Yang, and Wang 2019), SIMs with link functions of first- and second-order (including classical phase retrieval models as special cases) are studied through Stein's identity, allowing for general non-Gaussian sensing vectors. However, this work is primarily theoretical, and the recovery guarantees are only given with respect to globally optimal solutions to corresponding optimization problems, which are challenging to attain due to the typical non-convexity of these optimization problems.

Contributions

Throughout this work, we make the assumption that the underlying signal is within (up to a constant scaling) the range of a Lipschitz continuous generative model with bounded inputs. The main contributions of this paper are as follows:

- Based on the first-order Stein's identity, we provide near-optimal recovery guarantees for a practical projected gradient descent algorithm for SIMs with first-order link functions and non-Gaussian sensing vectors. We also show that for this case, a simple non-iterative approach with arbitrary initialization is effective for accurate reconstruction.

- Based on the second-order Stein's identity, we provide near-optimal recovery guarantees for a practical projected power method for SIMs with second-order link functions and non-Gaussian sensing vectors.
- We conduct experiments on image datasets to validate the effectiveness of our proposed approaches.

Preliminary

Throughout this paper, we will use the following notations.

Notations

We use upper and lower case boldface letters to denote matrices and vectors respectively. We write $[N] = \{1, 2, \cdots, N\}$ for a positive integer N. We use $\|\cdot\|$ to denote the ℓ_2 norm and use $a \wedge b$ to denote $\min\{a, b\}$. We define the ℓ_2-ball in $\mathbb{R}^k$ as $B^k(r) := \{\mathbf{z} \in \mathbb{R}^k : \|\mathbf{z}\| \le r\}$ and the unit sphere in $\mathbb{R}^n$ as $\mathcal{S}^{n-1} := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = 1\}$. A *generative model* is a function $G : \mathcal{D} \to \mathbb{R}^n$, with latent dimension k, ambient dimension n, and input domain $\mathcal{D} \subseteq \mathbb{R}^k$. We focus on the setting where $k \ll n$ and $\mathcal{D} = B^k(r)$. Additionally, we use $\mathcal{R}(G) = G(B^k(r)) = \{G(\mathbf{z}) : \mathbf{z} \in B^k(r)\}$ to denote the range of G. A random variable X is sub-Gaussian if $\sup_{p\ge1} p^{-1/2}(\mathbb{E}[|X|^p])^{1/p} < \infty$ and $\|X\|_{\psi_2} := \sup_{p\ge1} p^{-1/2}(\mathbb{E}[|X|^p])^{1/p}$ denotes its sub-Gaussian norm. A random vector $\mathbf{b} \in \mathbb{R}^n$ is sub-Gaussian if the one-dimensional marginals $\langle \mathbf{b}, \mathbf{s}\rangle$ are sub-Gaussian random variables for all $\mathbf{s} \in \mathbb{R}^n$, and the sub-Gaussian norm of $\mathbf{b}$ is defined as $\|\mathbf{b}\|_{\psi_2} := \sup_{\mathbf{s}\in\mathcal{S}^{n-1}} \|\langle \mathbf{b}, \mathbf{s}\rangle\|_{\psi_2}$. For a univariate function $g : \mathbb{R} \to \mathbb{R}$ and a vector $\mathbf{b}$, we denote $g \odot (\mathbf{b})$ as the output of applying g element-wisely to $\mathbf{b}$. We use standard Landau symbols for asymptotic notations.

The Assumption on the SIM

Except where stated otherwise, we make the following assumption on the SIM.

Assumption 1. *We have m i.i.d. realizations of the semi-parameterized SIM*

$$y = f(\mathbf{a}^T\mathbf{x}^*, w), \tag{2}$$

where similarly to (Wei, Yang, and Wang 2019), we assume that the following conditions are satisfied by the parameters of the SIM.

- *The underlying signal $\mathbf{x}^* \in \mathbb{R}^n$ is contained in (up to a constant scaling) the range of an L-Lipschitz generative model $G : B^k(r) \to \mathbb{R}^n$. In addition, since the norm of the signal is sacrificed in the SIM, we assume that $\mathbf{x}^*$ is a unit vector for brevity.*
- *The sensing vector $\mathbf{a} \in \mathbb{R}^n$ has joint density p with $p(\mathbf{a}) = \prod_{j=1}^n p_0(a_j)$ for some non-Gaussian 1-d density p_0. Let s_0 be the 1-d score function corresponding to p_0, i.e., $s_0(a) = -p_0'(a)/p_0(a)$. We assume that the score function $S_p(\mathbf{a}) := -\nabla p(\mathbf{a})/p(\mathbf{a}) = s_0 \odot (\mathbf{a})$ is sub-Gaussian.*
- *The random noise w is independent of $\mathbf{a}$.*
- *The non-linear link function $f : \mathbb{R}^2 \to \mathbb{R}$ is* unknown *and differentiable with respect to the first argument. We make no assumption on the specific form of the nonlinearity other than differentiability.*
- *Given the non-linear link function f is unknown and has arbitrary properties (except the differentiability with respect to the first argument), it is not reasonable to expect the observed random variable y to possess nice statistical properties like finite higher-order moments. Therefore, following (Wei, Yang, and Wang 2019), we assume that y is heavy-tailed with $\|y\|_{L_q} := \mathbb{E}\big[|y|^q\big]^{1/q} < \infty$ for some $q > 4$.*

The Algorithm and Theory for First-order Links

We have the following lemma for the first-order Stein's identity.

Lemma 1. (First-order Stein's identity (Stein et al. 2004)) *Let $g : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and $\mathbf{a} \in \mathbb{R}^n$ be a random vector with continuously differentiable density $p : \mathbb{R}^n \to \mathbb{R}$. Let $S_p(\mathbf{a}) = -\nabla \log p(\mathbf{a}) = -\nabla p(\mathbf{a})/p(\mathbf{a})$. Then, under the assumption that the expectations $\mathbb{E}[g(\mathbf{a})S_p(\mathbf{a})]$ and $\mathbb{E}[\nabla g(\mathbf{a})]$ are both well-defined, we have the generalized Stein's identity $\mathbb{E}[g(\mathbf{a})S_p(\mathbf{a})] = \mathbb{E}[\nabla g(\mathbf{a})]$.*

Based on Lemma 1, it is easy to calculate that for y generated from the SIM (2) (see, e.g., (Wei, Yang, and Wang 2019, Eq. (3))),

$$\mathbb{E}[yS_p(\mathbf{a})] = \mathbb{E}[f'(\mathbf{a}^T\mathbf{x}^*, w)] \cdot \mathbf{x}^*, \tag{3}$$

where we use $f'(x, w)$ to abbreviate $\partial f(x, w)/\partial x$. Then, if

$$\mu := \mathbb{E}\left[f'\left(\mathbf{a}^T\mathbf{x}^*, w\right)\right] \neq 0, \tag{4}$$

we can estimate $\mathbf{x}^*$ from $\mathbb{E}[yS_p(\mathbf{a})]$. A link functions f that satisfies (4) is called a first-order link. In particular, based on (3), if assuming $\mu\mathbf{x}^* \in \mathcal{R}(G)$, we can estimate $\mu\mathbf{x}^*$ via solving the following optimization problem:

$$\min_{\mathbf{x}\in\mathcal{R}(G)} \quad \|\mathbf{x}\|^2 - 2\mathbb{E}\left[yS_p(\mathbf{a})^T\mathbf{x}\right], \tag{5}$$

which yields $\mathbf{x} = \mu\mathbf{x}^*$ as a solution. Furthermore, the authors of (Wei, Yang, and Wang 2019) consider estimating $\mu\mathbf{x}^*$ by solving the empirical version of (5):

$$\min_{\mathbf{x}\in\mathcal{R}(G)} \|\mathbf{x}\|^2 - \frac{2}{m}\sum_{i=1}^m \tilde{y}_i S_p(\mathbf{a}_i)^T\mathbf{x}, \tag{6}$$

where $\tilde{y}_i = \mathrm{sign}(y_i) \cdot (|y_i| \wedge \tau)$ is the truncated version of y_i and is utilized to improve concentration with respect to heavy-tailed y_i, and $\frac{1}{m}\sum_{i=1}^m \tilde{y}_i S_p(\mathbf{a}_i)$ is an empirical approximation of $\mathbb{E}[yS_p(\mathbf{a})]$. The truncation parameter $\tau > 0$ is specified in the statement of Theorem 1 below. Due to the typical non-convexity of $\mathcal{R}(G)$, the optimization problem (6) is non-convex, and the optimal solution is difficult to attain. In this work, we propose to approximately solve (6) using the following projected gradient descent algorithm:

$$\mathbf{x}^{(\ell+1)} = \mathcal{P}_G\left(\mathbf{x}^{(\ell)} - 2\eta\left(\mathbf{x}^{(\ell)} - \frac{1}{m}\sum_{i=1}^{m}\tilde{y}_i S_p(\mathbf{a}_i)\right)\right), \tag{7}$$

where $\eta > 0$ is the step size, the initial vector $\mathbf{x}^{(0)}$ is arbitrarily chosen, and $\mathcal{P}_G(\cdot)$ denotes the projection onto the range of G, i.e., $\mathcal{P}_G(\mathbf{x}) = \arg\min_{\mathbf{v}\in\mathcal{R}(G)} \|\mathbf{v} - \mathbf{x}\|$ for any $\mathbf{x} \in \mathbb{R}^n$.[1] We have the following theorem for the projected gradient descent algorithm (7).

Theorem 1. *Suppose that Assumption 1 is satisfied with $\mu\mathbf{x}^* \in \mathcal{R}(G)$ and the non-linear link function f satisfies* (4)*. Let $\beta \geq 2$ be a positive constant and let the truncation parameter $\tau = m^{1/2(1+\kappa)}M_y$ with some $M_y \geq \|y\|_{L_q}$ and $\kappa \in \big(0, \frac{q}{4} - 1\big)$. Let $C_{\mathbf{a},y,\beta} = C\|S_p(\mathbf{a})\|_{\psi_2} M_y\sqrt{\beta(1+\kappa)/\kappa}$, where C is an absolute constant. Then, for any $\delta > 0$ satisfying $\delta \leq 4\eta C_{\mathbf{a},y,\beta}\sqrt{\frac{k\log\frac{4Lr}{\delta}}{m}}$ and any $\eta \in \big(\frac{1}{4}, \frac{1}{2}\big]$, with probability at least $1 - e^{-\Omega(k\log\frac{Lr}{\delta})} - e^{-\beta}$, we have for any integer $\ell \geq 0$, it holds that*

$$\left\|\mathbf{x}^{(\ell+1)} - \mu\mathbf{x}^*\right\| \leq 2(1-2\eta)\left\|\mathbf{x}^{(\ell)} - \mu\mathbf{x}^*\right\| + 4\eta C_{\mathbf{a},y,\beta}\sqrt{\frac{k\log\frac{4Lr}{\delta}}{m}}. \tag{8}$$

Remark 1. *In (Wei, Yang, and Wang 2019), it is assumed that both $\mathbf{x}^*$ and $\mu\mathbf{x}^*$ are contained in the range of the generative model. When $\mu \neq 1$, this is a restrictive assumption, and is only naturally satisfied when the structured set $\mathcal{R}(G)$ is a cone (e.g., when G is a neural network with ReLU activations and no offsets). We relax this assumption to only require that $\mu\mathbf{x}^*$ is contained in the structured set, which coincides with the assumptions made in prior works such as (Plan and Vershynin 2016; Liu and Scarlett 2020a).*

Given that a d-layer fully-connected neural network typically has the Lipschitz constant $L = n^{\Theta(d)}$ (Bora et al. 2017), r can be set to of the same order as L, and δ can be set to be as small as $O(1/(Lr))$, the term $\sqrt{\frac{k\log\frac{4Lr}{\delta}}{m}}$ in (8) is roughly of the near-optimal order $\sqrt{\frac{k\log L}{m}}$ (Liu and Scarlett 2020b). Then, we observe from Theorem 1 that when $m = O\big(\frac{k}{\varepsilon^2}\log L\big)$ and $2(1-2\eta) \in (0,1)$ (or equivalently, $\frac{1}{4} < \eta < \frac{1}{2}$), the projected gradient descent algorithm (7) converges linearly to a point achieving ε-accurate recovery. Moreover, when $\eta = \frac{1}{2}$, our result indicates that one step is sufficient to obtain an estimated vector that achieves the near-optimal statistical rate. For this case, when setting the initial vector $\mathbf{x}^{(0)}$ to be a zero vector, the projected gradient descent algorithm (7) reduces to the non-iterative approach:

$$\mathbf{x}^{(1)} = \mathcal{P}_G\left(\frac{1}{m}\sum_{i=1}^{m}\tilde{y}_i S_p(\mathbf{a}_i)\right). \tag{9}$$

The Algorithm and Theory for Second-order Links

For SIM with first-order links, we impose the first-order link condition (4), which does not hold for non-linear link functions associated with popular phase retrieval models (Yang, Balasubramanian, and Liu 2017). In order to address phase retrieval models, we rely on the following lemma for second-order Stein's identity.

Lemma 2. (Second-order Stein's identity (Janzamin, Sedghi, and Anandkumar 2014)) *Let $\mathbf{a} \in \mathbb{R}^n$ be a random vector with twice differentiable density p. We define the second-order score function $T_p(\mathbf{a}) : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ as $T_p(\mathbf{a}) = \nabla^2 p(\mathbf{a})/p(\mathbf{a})$. Then, for any twice differentiable function $g : \mathbb{R}^n \to \mathbb{R}$ such that $\mathbb{E}[\nabla^2 g(\mathbf{a})]$ exists, we have*

$$\mathbb{E}\left[g(\mathbf{a})T_p(\mathbf{a})\right] = \mathbb{E}\left[\nabla^2 g(\mathbf{a})\right]. \tag{10}$$

Based on the second-order Stein's identity, for y generated from the SIM (2), it is easy to calculate that (see, e.g., (Wei, Yang, and Wang 2019, Eq. (4)))

$$\mathbb{E}[yT_p(\mathbf{a})] = \mathbb{E}\left[f''(\mathbf{a}^T\mathbf{x}^*, w)\right]\cdot\mathbf{x}^*(\mathbf{x}^*)^T, \tag{11}$$

where we use $f''(x,w)$ to abbreviate $\partial^2 f(x,w)/\partial x^2$. Therefore, when [2]

$$\nu := \mathbb{E}\left[f''\left(\mathbf{a}^T\mathbf{x}^*, w\right)\right] \neq 0, \tag{12}$$

we can estimate the underlying signal $\mathbf{x}^*$ by reconstructing the leading eigenvector of $\mathbb{E}\left[g(\mathbf{a})T_p(\mathbf{a})\right]$. A link function f that satisfies (12) is called a second-order link. Then, based on (11), if assuming $\mathbf{x}^* \in \mathcal{R}(G) \subseteq \mathcal{S}^{n-1}$,[3] we can estimate $\mathbf{x}^*$ by solving the following optimization problem:

$$\max_{\mathbf{x}\in\mathcal{R}(G)} \mathbf{x}^T\mathbb{E}[yT_p(\mathbf{a})]\mathbf{x}, \tag{13}$$

which has the solution $\mathbf{x} = \mathbf{x}^*$. Note that $T_p(\mathbf{a}) = S_p(\mathbf{a})S_p(\mathbf{a})^T - \mathrm{Diag}(s_0' \odot (\mathbf{a}))$ (*cf.* Assumption 1). The authors of (Wei, Yang, and Wang 2019) consider recovering $\mathbf{x}^*$ by solving the empirical version of (13), which gives the following optimization problem:

$$\max_{\mathbf{x}\in\mathcal{R}(G)} \frac{1}{m}\sum_{i=1}^{m}\tilde{y}_i\left(S_p(\mathbf{a}_i)^T\mathbf{x}\right)^2, \tag{14}$$

[1]We will follow prior works (Shah and Hegde 2018; Peng, Jalali, and Yuan 2020; Liu et al. 2022; Liu and Liu 2022) to implicitly assume that the projection step can be accurately performed, although approximate methods, such as gradient descent (Shah and Hegde 2018; Peng, Jalali, and Yuan 2020) or GAN-based projection methods (Raj, Li, and Bresler 2019), may need to be used in practice.

[2]We will focus on the case that $\nu > 0$. The case that $\nu < 0$ can be similarly handled by replacing y with $-y$.

[3]For convenience, here we follow (Liu et al. 2020, 2022) to assume that the range of the generative model is a subset of the unit sphere in $\mathbb{R}^n$. For a general unnormalized Lipschitz continuous generative model, we can essentially consider its normalized version. See (Liu et al. 2022, Remark 1) for more details.

where $\tilde{y}_i = \mathrm{sign}(y_i) \cdot (|y_i| \wedge \tau)$ is a truncated version of y_i and $\frac{1}{m}\sum_{i=1}^m \tilde{y}_i S_p(\mathbf{a}_i) S_p(\mathbf{a}_i)^T$ is an empirical approximation of $\mathbb{E}[y S_p(\mathbf{a}) S_p(\mathbf{a})^T]$. The truncation parameter $\tau > 0$ differs slightly from the one used in the section on first-order links and is specified in the statement of Theorem 2 below.

The optimization problem (14) is typically non-convex, making it difficult to find the optimal solution. To tackle this issue, we propose the following iterative approach, called the projected power method, which is a variation of the traditional power method with an additional projection operation in each iteration:

$$\mathbf{x}^{(\ell+1)} = \mathcal{P}_G\left(\frac{1}{m}\sum_{i=1}^m \tilde{y}_i \left(S_p(\mathbf{a}_i)^T \mathbf{x}^{(\ell)}\right) S_p(\mathbf{a}_i)\right). \quad (15)$$

We have the following theorem for the projected power method (15).

Theorem 2. *Suppose that Assumption 1 is satisfied with* $\mathbf{x}^* \in \mathcal{R}(G) \subseteq \mathcal{S}^{n-1}$ *and the non-linear link function* f *satisfies* (12). *Let* $\beta \geq 2$ *be a positive constant and let* $\tau = \left(\frac{k\log(Lr)}{m}\right)^{-\frac{1}{2(1+\kappa)}} \cdot M_y$ *with some* $M_y \geq \|y\|_{L_q}$ *and* $\kappa \in \left(0, \frac{q}{4}-1\right)$. *Let* $C'_{\mathbf{a},y,\beta} = \frac{C(1+\kappa)(\|S_p(\mathbf{a})\|_{\psi_2} + \|S_p(\mathbf{a})\|_{\psi_2}^2) M_y \sqrt{\beta}}{\nu\kappa}$, *where* C *is an absolute constant. Then, for any* $\delta > 0$ *satisfying* $\delta \leq 2C'_{\mathbf{a},y,\beta}\sqrt{\frac{k\log\frac{4Lr}{\delta}}{m}}$, *with probability at least* $1 - e^{-\Omega(k\log\frac{Lr}{\delta})} - e^{-\beta}$, *we have that if there exists an* $\ell_0 \geq 0$ *such that* $c_0 := \langle \mathbf{x}^{(\ell_0)}, \mathbf{x}^* \rangle > 0$, *then it holds for any* $\ell > \ell_0$ *that*

$$\left\|\mathbf{x}^{(\ell)} - \mathbf{x}^*\right\| \leq \frac{4C'_{\mathbf{a},y,\beta}}{\nu c_0} \cdot \sqrt{\frac{k\log\frac{4Lr}{\delta}}{m}}. \quad (16)$$

Remark 2. *In certain scenarios, we may make the assumption that the data only contains non-negative vectors, such as in the case of image datasets. In addition, during pre-training, we can set the activation function of the final layer of the neural network generative model to be a non-negative function, such as ReLU or sigmoid, which further restricts the range of the generative model to be contained in the non-negative orthant. Therefore, the assumption that* $c_0 := \langle \mathbf{x}^{(\ell_0)}, \mathbf{x}^* \rangle > 0$ *is mild and similar assumptions have been adopted in prior works including (Liu et al. 2022; Liu, Wang, and Liu 2022). As a result, we provide an upper bound on* $\left\|\mathbf{x}^{(\ell)} - \mathbf{x}^*\right\|$, *instead of the distance measure* $\min\left\{\left\|\mathbf{x}^{(\ell)} - \mathbf{x}^*\right\|, \left\|\mathbf{x}^{(\ell)} + \mathbf{x}^*\right\|\right\}$ *that is commonly adopted in relevant literature on real-valued phase retrieval problems.*

Remark 3. *The iterative algorithm* (15) *is similar to that proposed by (Liu et al. 2022). However, the authors of (Liu et al. 2022) do not employ Stein's identity to make use of the score functions of distributions in their algorithm, and their theoretical guarantees are only applicable to phase retrieval with Gaussian sensing vectors, as noted in Example 2 of their paper. For the case of handling phase retrieval with non-Gaussian sensing vectors as we explore in this paper, the proof technique is significantly different from that used in (Liu et al. 2022).*

Similarly to the discussion after Theorem 2, we know that the term $\sqrt{(k\log(4Lr/\delta))/m}$ in (16) is roughly of order $\sqrt{(k\log L)/m}$. Therefore, Theorem 2 essentially says that under appropriate initialization, approximately $m = O\left(\frac{k}{\varepsilon^2}\log L\right)$ measurements suffice to ensure that the projected power method (15) returns an estimated vector that achieves ε-accurate recovery.

Experiments

We present numerical results for the MNIST (LeCun et al. 1998) and CelebA (Liu et al. 2015) datasets to support our theoretical results, with the results for CelebA being presented in the supplementary material due to the page limit.

The MNIST dataset contains 60,000 images of handwritten digits, each measuring 28×28 pixels, resulting in an ambient dimension of $n = 784$. The generative model G for the MNIST dataset is chosen to be a pre-trained variational autoencoder (VAE) model with a latent dimension of $k = 20$. The encoder and decoder are both fully connected neural networks with two hidden layers, having an architecture of $20 - 500 - 500 - 784$. The VAE is trained using the Adam optimizer with a mini-batch size of 100 and a learning rate of 0.001 on the original MNIST training set. To approximately perform the projection step $\mathcal{P}_G(\cdot)$, we use a gradient descent method with the Adam optimizer, using step size of 100 and a learning rate of 0.1. This approximation method has been used in several works, including (Shah and Hegde 2018; Peng, Jalali, and Yuan 2020; Liu et al. 2020, 2022; Liu and Liu 2022). The reconstruction task is evaluated on a random subset of 10 images drawn from the testing set of the MNIST dataset.

We follow (Yang, Balasubramanian, and Liu 2017) to set the 1-d density p_0 to be that corresponds to the Gamma distribution with shape parameter 5 and scale parameter 1, i.e., $p_0(a) = a^4 e^{-a}/\Gamma(5)$, or the Rayleigh distribution with scale parameter 2, i.e., $p_0(a) = (a e^{-a^2/8})/4$. We conduct the experiments for the SIM (2) with first- and second-order link functions. Since the norm of the signal is absorbed in the SIM, we focus on the estimation of the direction of the signal and do not consider its norm. To compare the performance of different algorithms, we use the scale-invariant Cosine Similarity metric, which is defined as $CosSim(\mathbf{x}^*, \hat{\mathbf{x}}) = \hat{\mathbf{x}}^T\mathbf{x}^*$, with $\mathbf{x}^*$ being the underlying signal and $\hat{\mathbf{x}}$ referring to the normalized output vector of each algorithm. We use 10 random restarts to mitigate the impact of local minima and select the best result among these random restarts. The cosine similarity is averaged over the 10 test images and over these 10 restarts. All experiments are conducted using Python 3.10.6 and PyTorch 2.0.0 with an NVIDIA RTX 3060 Laptop 6GB GPU.

First-order Links

Similarly to (Yang, Balasubramanian, and Liu 2017), we set f to be either

$$f(x, y) = 3x + 10\sin(x) + y, \quad (17)$$

or

$$f(x, y) = \sqrt{2}x + 4\exp\left(-2x^2\right) + y. \quad (18)$$

Both non-linear link functions (17) and (18) are not monotonic with regards to x. We assign the random noise w (see (2)) to have zero-mean Gaussian distribution with a

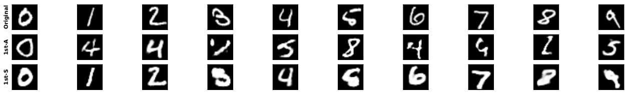


Figure 1: Reconstructed images of the MNIST dataset for SIM with link function (17) and Gamma sensing vectors. The number of measurements $m = 400$.

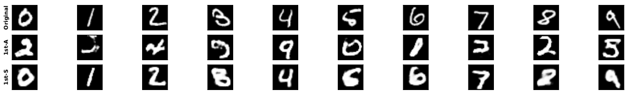
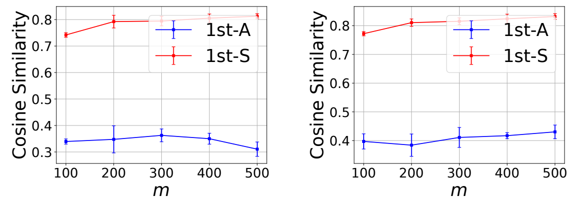


Figure 2: Reconstructed images of the MNIST dataset for SIM with link function (17) and Rayleigh sensing vectors. The number of measurements $m = 500$.



(a) Gamma distr.

(b) Rayleigh distr.

Figure 3: Quantitative results of the performance of `1st-S` and `1st-A` for SIM with link function (17) and Gamma or Rayleigh sensing vectors on the MNIST dataset.

standard deviation of 0.1. To demonstrate the effect of the number of measurements m, we vary m within the values of $\{100, 200, 300, 400, 500\}$. The truncation parameter τ is to be $\tau = 5\sqrt{m}$. Based on Theorem 1, we perform the experiments for the efficient non-iterative approach (9) (denoted by `1st-S`). We note that it has been shown in (Liu and Liu 2022) that a similar non-iterative approach with respect to $\mathbf{a}_i$ (instead of $S_p(\mathbf{a}_i)$; the corresponding estimated vector is $\mathcal{P}_G\big(\frac{1}{m}\sum_{i=1}^m \tilde{y}_i\mathbf{a}_i\big)$) is effective for standard Gaussian sensing vectors. We compare with such an approach, and denote it by `1st-A` to highlight that it is with respect to the sensing vectors $\mathbf{a}_i$ directly. It has been established in various previous studies (see, e.g. (Bora et al. 2017; Liu et al. 2020; Liu and Liu 2022)) that when the number of measurements is comparatively small in relation to the ambient dimension, generative model-based approaches yield significantly superior reconstructions compared to sparsity-based approaches. Therefore, in this work, we will not compare our proposed method with sparsity-based approaches.

The reconstructed images from the experiments are depicted in Figures 1, 2, 4, 5, and the corresponding quantitative results are demonstrated in Figures 3, 6. From the figures provided, it is evident that for non-Gaussian sensing vectors, the non-iterative method using score function-valued sensing vectors $S_p(\mathbf{a}_i)$ (i.e., (9)) yields significantly better reconstructions compared to the non-iterative method using sensing vectors $\mathbf{a}_i$ themselves. This is consistent with the theoretical findings in (Ai et al. 2014; Goldstein and Wei 2019) that for SIM with non-Gaussian sensing vectors, conventional estimators that directly apply to sensing vectors can lead to fixed bias terms, regardless of the number of measurements taken.

Second-order Links

Similarly to (Yang et al. 2017), we set f to be either

$$f(x, y) = |x| + y, \tag{19}$$

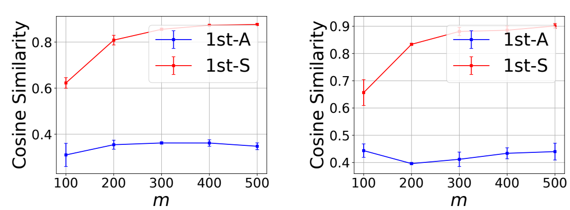
or

$$f(x, y) = 4x^2 + 3\sin(x) + y, \tag{20}$$

which corresponds to (misspecified) phase retrieval models with additive noise. The random noise w (see (2)) is also

Figure 4: Reconstructed images of the MNIST dataset for SIM with link function (18) and Gamma sensing vectors. The number of measurements $m = 300$.

Figure 5: Reconstructed images of the MNIST dataset for SIM with link function (18) and Rayleigh sensing vectors. The number of measurements $m = 400$.



(a) Gamma distr.

(b) Rayleigh distr.

Figure 6: Quantitative results of the performance of `1st-S` and `1st-A` for SIM with link function (18) and Gamma or Rayleigh sensing vectors on the MNIST dataset.

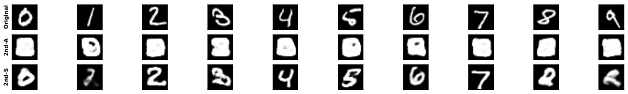


Figure 7: Reconstructed images of the MNIST dataset for SIM with link function (19) and Gamma sensing vectors. The number of measurements $m = 400$.

Figure 8: Reconstructed images of the MNIST dataset for SIM with link function (19) and Rayleigh sensing vectors. The number of measurements $m = 300$.

set to be zero-mean Gaussian with a standard deviation of 0.1. To illustrate the effect of the number of measureme m, we vary m in $\{100, 200, 300, 400, 500\}$. The truncat parameter τ is set to be $\tau = 5\sqrt{m/k}$ with $k = 20$. Ba on Theorem 2, we perform the experiments for the projec power method with respect to $S_p(\mathbf{a}_i)$ as presented in (15) (denoted by `2nd-S`). We compare with the projected power method that applies to the sensing vectors $\mathbf{a}_i$ directly (denoted by `2nd-A`; for such an approach, the estimated vector is $\mathcal{P}_G\big(\frac{1}{m}\sum_{i=1}^m \tilde{y}_i\big(\mathbf{a}_i^T\mathbf{x}^{(\ell)}\big)\mathbf{a}_i\big)$). For both methods, the number of iterations is set to 10. It has also been established in prior works, such as (Hyder et al. 2019; Jagatap and Hegde 2019; Liu et al. 2022), that for phase retrieval problems, when the number of measurements is relatively small cc pared to the ambient dimension, generative model-based proaches yield significantly better reconstructions compa to those of sparsity-based approaches. As such, we will compare with sparsity-based approaches.

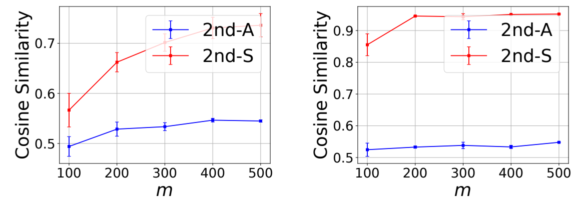
The reconstructed image results are displayed in Figures 7, 8, 10, 11, and the corresponding quantitative results are demonstrated in Figures 9, 12. These figures indicate that for SIM with non-Gaussian sensing vectors and second-order links (19) and (20), the projected power method with respect to $S_p(\mathbf{a}_i)$ (as shown in (15)) yields significantly better reconstructions compared to that of the projected power method with respect to sensing vectors $\mathbf{a}_i$.

Conclusion

This work offers recovery guarantees for efficient algorithms designed for high-dimensional SIMs with first- and second-order links, utilizing generative priors and non-Gaussian sensing vectors.

Acknowledgments

J. Chen was supported by a Hong Kong PhD Fellowship from the Hong Kong Research Grants Council (RGC). Z. Liu was supported by the National Natural Science Foundation of China (No. 62176047, 82121003), and Sichuan Science and Technology Program (No. 2021YFS0374, 2022YFS0600).

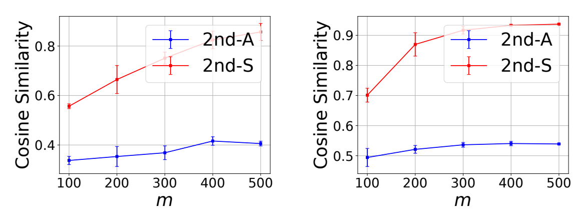


(a) Gamma distr. (b) Rayleigh distr.

Figure 9: Quantitative results of the performance of `1st-S` and `1st-A` for SIM with link function (19) and Gamma or Rayleigh sensing vectors on the MNIST dataset.

Figure 10: Reconstructed images of the MNIST dataset for SIM with link function (20) and Gamma sensing vectors. The number of measurements $m = 400$.

Figure 11: Reconstructed images of the MNIST dataset for SIM with link function (20) and Rayleigh sensing vectors. The number of measurements $m = 400$.



(a) Gamma distr. (b) Rayleigh distr.

Figure 12: Quantitative results of the performance of `2nd-S` and `2nd-A` for SIM with link function (20) and Gamma or Rayleigh sensing vectors on the MNIST dataset.

References

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

Asim, M.; Daniels, M.; Leong, O.; Ahmed, A.; and Hand, P. 2020. Invertible generative models for inverse problems: mitigating representation error and dataset bias. In *ICML*.

Aubin, B.; Loureiro, B.; Baker, A.; Krzakala, F.; and Zdeborová, L. 2020. Exact asymptotics for phase retrieval and compressed sensing with random generative priors. In *MSML*, 55–73. PMLR.

Aubin, B.; Loureiro, B.; Maillard, A.; Krzakala, F.; and Zdeborová, L. 2019. The spiked matrix model with generative priors. In *NeurIPS*, volume 32.

Bora, A.; Jalal, A.; Price, E.; and Dimakis, A. G. 2017. Compressed sensing using generative models. In *ICML*, 537–546. PMLR.

Bühlmann, P.; and Van De Geer, S. 2011. *Statistics for high-dimensional data: Methods, theory and applications*. Springer Science & Business Media.

Candes, E. J.; Li, X.; and Soltanolkotabi, M. 2015. Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4): 1985–2007.

Candès, E. J.; Romberg, J.; and Tao, T. 2006. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2): 489–509.

Candes, E. J.; Romberg, J. K.; and Tao, T. 2006. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8).

Candès, E. J.; and Wakin, M. B. 2008. An introduction to compressive sampling. *IEEE signal processing magazine*.

Chen, J.; Scarlett, J.; Ng, M.; and Liu, Z. 2023. A Unified Framework for Uniform Signal Recovery in Nonlinear Generative Compressed Sensing. In *NeurIPS*.

Cheng, L.; Zeng, P.; and Zhu, Y. 2017. BS-SIM: An effective variable selection method for high-dimensional single index model. *Electronic Journal of Statistics*, 11: 3522–3548.

Daras, G.; Dean, J.; Jalal, A.; and Dimakis, A. 2021. Intermediate layer optimization for inverse problems using deep generative models. In *ICML*, 2421–2432. PMLR.

Dhar, M.; Grover, A.; and Ermon, S. 2018. Modeling sparse deviations for compressed sensing using generative models. In *ICML*, 1214–1223. PMLR.

Donoho, D. L. 2006. Compressed sensing. *IEEE Transactions on information theory*, 52(4): 1289–1306.

Foucart, S.; Rauhut, H.; Foucart, S.; and Rauhut, H. 2013. *An invitation to compressive sensing*. Springer.

Ganti, R.; Rao, N.; Willett, R. M.; and Nowak, R. 2015. Learning single index models in high dimensions. *stat*, 1050: 30.

Genzel, M. 2016. High-dimensional estimation of structured signals from non-linear observations with general convex loss functions. *IEEE Transactions on Information Theory*, 63(3): 1601–1619.

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

Goldstein, L.; and Wei, X. 2019. Non-Gaussian observations in nonlinear compressed sensing via Stein discrepancies. *Information and Inference: A Journal of the IMA*, 8(1): 125–159.

Han, A. K. 1987. Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator. *Journal of Econometrics*, 35(2-3): 303–316.

Hand, P.; Leong, O.; and Voroninski, V. 2018. Phase retrieval under a generative prior. In *NeurIPS*, volume 31.

Heckel, R.; and Hand, P. 2018. Deep decoder: Concise image representations from untrained non-convolutional networks. In *ICLR*.

Hristache, M.; Juditsky, A.; and Spokoiny, V. 2001. Direct estimation of the index coefficient in a single-index model. *Annals of Statistics*, 595–623.

Hyder, R.; Shah, V.; Hegde, C.; and Asif, M. S. 2019. Alternating phase projected gradient descent with generative priors for solving compressive phase retrieval. In *ICASSP*, 7705–7709. IEEE.

Jagatap, G.; and Hegde, C. 2019. Algorithmic guarantees for inverse imaging with untrained network priors. In *NeurIPS*, volume 32.

Jalal, A.; Arvinte, M.; Daras, G.; Price, E.; Dimakis, A. G.; and Tamir, J. 2021a. Robust compressed sensing MRI with deep generative priors. In *NeurIPS*, volume 34, 14938–14954.

Jalal, A.; Karmalkar, S.; Dimakis, A.; and Price, E. 2021b. Instance-optimal compressed sensing via posterior sampling. In *ICML*.

Jalal, A.; Liu, L.; Dimakis, A. G.; and Caramanis, C. 2020. Robust compressed sensing using generative models. In *NeurIPS*, volume 33, 713–727.

Janzamin, M.; Sedghi, H.; and Anandkumar, A. 2014. Score function features for discriminative learning: Matrix and tensor framework. *arXiv preprint arXiv:1412.2863*.

Joshi, B.; Li, X.; Plan, Y.; and Yilmaz, O. 2021. PLUGIn: A simple algorithm for inverting generative models with recovery guarantees. In *NeurIPS*, volume 34, 24719–24729.

Killedar, V.; and Seelamantula, C. S. 2022. Compressive phase retrieval based on sparse latent generative priors. In *ICASSP*, 1596–1600. IEEE.

Latorre, F.; Cevher, V.; et al. 2019. Fast and provable ADMM for learning with generative priors. In *NeurIPS*, volume 32.

LeCun, Y.; Bottou, L.; Bengio, Y.; and Haffner, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11): 2278–2324.

Li, K.-C.; and Duan, N. 1989. Regression analysis under link violation. *The Annals of Statistics*, 17(3): 1009–1052.

Liu, J.; and Liu, Z. 2022. Non-iterative recovery from nonlinear observations using generative models. In *CVPR*.

Liu, Z.; Ghosh, S.; and Scarlett, J. 2021. Towards sample-optimal compressive phase retrieval with sparse and generative priors. In *NeurIPS*, volume 34, 17656–17668.

Liu, Z.; Gomes, S.; Tiwari, A.; and Scarlett, J. 2020. Sample complexity bounds for 1-bit compressive sensing and binary stable embeddings with generative priors. In *ICML*, 6216–6225. PMLR.

Liu, Z.; and Han, J. 2022. Projected gradient descent algorithms for solving nonlinear inverse problems with generative priors. *arXiv preprint arXiv:2209.10093*.

Liu, Z.; Liu, J.; Ghosh, S.; Han, J.; and Scarlett, J. 2022. Generative Principal Component Analysis. In *ICLR*.

Liu, Z.; Luo, P.; Wang, X.; and Tang, X. 2015. Deep learning face attributes in the wild. In *ICCV*, 3730–3738.

Liu, Z.; and Scarlett, J. 2020a. The generalized Lasso with nonlinear observations and generative priors. In *NeurIPS*, volume 33, 19125–19136.

Liu, Z.; and Scarlett, J. 2020b. Information-theoretic lower bounds for compressive sensing with generative models. *IEEE Journal on Selected Areas in Information Theory*, 1(1): 292–303.

Liu, Z.; Wang, X.; and Liu, J. 2022. Misspecified phase retrieval with generative priors. In *NeurIPS*, volume 35, 5109–5123.

Luo, S.; and Ghosal, S. 2016. Forward selection and estimation in high dimensional single index models. *Statistical Methodology*, 33: 172–179.

Neykov, M.; Liu, J. S.; and Cai, T. 2016. L_1-regularized least squares for support recovery of high dimensional single index models with gaussian designs. *The Journal of Machine Learning Research*, 17(1): 2976–3012.

Neykov, M.; Wang, Z.; and Liu, H. 2020. Agnostic estimation for misspecified phase retrieval models. *The Journal of Machine Learning Research*, 21(1): 4769–4807.

Ongie, G.; Jalal, A.; Metzler, C. A.; Baraniuk, R. G.; Dimakis, A. G.; and Willett, R. 2020. Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory*, 1(1): 39–56.

Oymak, S.; and Soltanolkotabi, M. 2017. Fast and reliable parameter estimation from nonlinear observations. *SIAM Journal on Optimization*, 27(4): 2276–2300.

Pananjady, A.; and Foster, D. P. 2021. Single-index models in the high signal regime. *IEEE Transactions on Information Theory*, 67(6): 4092–4124.

Peng, P.; Jalali, S.; and Yuan, X. 2020. Solving inverse problems via auto-encoders. *IEEE Journal on Selected Areas in Information Theory*, 1(1): 312–323.

Plan, Y.; and Vershynin, R. 2016. The generalized lasso with non-linear observations. *IEEE Transactions on information theory*, 62(3): 1528–1537.

Plan, Y.; Vershynin, R.; and Yudovina, E. 2017. High-dimensional estimation with geometric constraints. *Information and Inference: A Journal of the IMA*, 6(1): 1–40.

Qiu, S.; Wei, X.; and Yang, Z. 2020. Robust one-bit recovery via ReLU generative networks: Near-optimal statistical rate and global landscape analysis. In *ICML*, 7857–7866. PMLR.

Radchenko, P. 2015. High dimensional single index models. *Journal of Multivariate Analysis*, 139: 266–282.

Raj, A.; Li, Y.; and Bresler, Y. 2019. GAN-based projector for faster recovery with convergence guarantees in linear inverse problems. In *ICCV*, 5602–5611.

Scarlett, J.; Heckel, R.; Rodrigues, M. R.; Hand, P.; and Eldar, Y. C. 2022. Theoretical perspectives on deep learning methods in inverse problems. *IEEE Journal on Selected Areas in Information Theory*, 3(3): 433–453.

Shah, V.; and Hegde, C. 2018. Solving linear inverse problems using GAN priors: An algorithm with provable guarantees. In *ICASSP*, 4609–4613. IEEE.

Shamshad, F.; and Ahmed, A. 2020. Compressed sensing-based robust phase retrieval via deep generative priors. *IEEE Sensors Journal*, 21(2): 2286–2298.

Sherman, R. P. 1993. The limiting distribution of the maximum rank correlation estimator. *Econometrica: Journal of the Econometric Society*, 123–137.

Soltani, M.; and Hegde, C. 2017. Fast algorithms for demixing sparse signals from nonlinear observations. *IEEE Transactions on Signal Processing*, 65(16): 4209–4222.

Stein, C.; Diaconis, P.; Holmes, S.; and Reinert, G. 2004. Use of exchangeable pairs in the analysis of simulations. *Lecture Notes-Monograph Series*, 1–26.

Thrampoulidis, C.; Abbasi, E.; and Hassibi, B. 2015. Lasso with non-linear measurements is equivalent to one with linear measurements. In *NeurIPS*, volume 28.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1): 267–288.

Wei, X. 2018. Structured recovery with heavy-tailed measurements: A thresholding procedure and optimal rates. *arXiv preprint arXiv:1804.05959*.

Wei, X.; Yang, Z.; and Wang, Z. 2019. On the statistical rate of nonlinear recovery in generative models with heavy-tailed data. In *ICML*, 6697–6706. PMLR.

Yang, Z.; Balasubramanian, K.; and Liu, H. 2017. High-dimensional non-Gaussian single index models via thresholded score function estimation. In *ICML*.

Yang, Z.; Balasubramanian, K.; Wang, Z.; and Liu, H. 2017. Learning non-gaussian multi-index model via second-order stein's method. In *NeurIPS*, volume 30, 6097–6106.

Yang, Z.; Wang, Z.; Liu, H.; Eldar, Y. C.; and Zhang, T. 2015. Sparse nonlinear regression: Parameter estimation and asymptotic inference. *arXiv preprint arXiv:1511.04514*.

Yang, Z.; Yang, L. F.; Fang, E. X.; Zhao, T.; Wang, Z.; and Neykov, M. 2019. Misspecified nonconvex statistical optimization for sparse phase retrieval. *Mathematical Programming*, 176: 545–571.

Zhang, K.; Yang, Z.; and Wang, Z. 2018. Nonlinear structured signal estimation in high dimensions via iterative hard thresholding. In *AISTATS*, 258–268. PMLR.